UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

The Output Convergence Debate Revisited: Lessons from Recent Developments in the Analysis of Panel Data Models

M. Hashem Pesaran
University of Southern California and Trinity College, University of Cambridge

Ron P. Smith
Birkbeck, University of London

Abstract

This paper provides a critical examination of the empirical basis of the output convergence debate in the light of recent developments in the analysis of dynamic heterogeneous panels with interactive effects. It shows that popular tools such as Barro's cross-country regressions and two-way fixed effects (TWFE) estimators that assume parallel trends and homogeneous dynamics lead to substantial under-estimation of the speed of convergence and misleading inference. Instead, dynamic common correlated effects (DCCE) estimators due to Chudik and Pesaran (2015a) provide consistent estimates and valid inference that are robust to nonparallel trends and correlated heterogeneity and apply even if there are breaks, trends and/or unit roots in the latent technology factor. It also suggests a way to estimate the effect of slowly moving determinants of growth. The theoretical findings are augmented with empirical evidence using PennWorld Tables data, finding little evidence of per capita output convergence across countries, very slow evidence of cross-country growth convergence, and reasonably fast within country convergence. Capital accumulation is found to be the most important single determinant of cross-country differences in output while slow moving indicators such as potential for conflict and protection of property rights proved to be statistically significant determinants of the steady state levels of output per capita. We are also able to replicate a positive evidence of democratization on output, but we find that the statistical significance of this effect to fall as we allow for nonparallel trends and dynamic heterogeneity.

Reference Details

2607	Cambridge Working Papers in Economics
Published	03 February 2026
Keywords	Output growth, output convergence, Barro regression, panel data estimators, TWFE and DCCEP, heterogeneity bias, nonparallel trends.
JEL codes	C1, C33, E10, F43, O4
Website	https://www.econ.cam.ac.uk/publications/cwpe

The Output Convergence Debate Revisited: Lessons from recent developments in the analysis of panel data models*

M. Hashem Pesaran
University of Southern California and Trinity College, University of Cambridge

Ron P. Smith
Birkbeck, University of London

February 3, 2026

Abstract

This paper provides a critical examination of the empirical basis of the output convergence debate in the light of recent developments in the analysis of dynamic heterogeneous panels with interactive effects. It shows that popular tools such as Barro's cross-country regressions and two-way fixed effects (TWFE) estimators that assume parallel trends and homogeneous dynamics lead to substantial underestimation of the speed of convergence and misleading inference. Instead, dynamic common correlated effects (DCCE) estimators due to Chudik and Pesaran (2015a) provide consistent estimates and valid inference that are robust to nonparallel trends and correlated heterogeneity and apply even if there are breaks, trends and/or unit roots in the latent technology factor. It also suggests a way to estimate the effect of slowly moving determinants of growth. The theoretical findings are augmented with empirical evidence using Penn World Tables data, finding little evidence of per capita output convergence across countries, very slow evidence of cross country growth convergence, and reasonably fast within country convergence. Capital accumulation is found to be the most important single determinant of cross-country differences in output while slow moving indicators such as potential for conflict and protection of property rights proved to be statistically significant determinants of the steady state levels of output per capita. We are also able to replicate a positive evidence of democratization on output, but we find that the statistical significance of this effect to fall as we allow for nonparallel trends and dynamic heterogeneity.

Keywords: Output growth, output convergence, Barro regression, panel data estimators, TWFE and DCCEP, heterogeneity bias, nonparallel trends.
JEL Classification: C1, C33, E10, F43, O4

*This paper was presented as a keynote talk at 19^{th} International Joint Conference on Computational and Financial Econometrics (CFE) and Computational and Methodological Statistics (CMStatistics), at Birkbeck, University of London, 13-15 December 2025. We are grateful to the conference participants, Daron Acemoglu, Alexander Chudik, Jan Ditzen, and Oscar Jorda for helpful comments, and to Jingshu Chen, a PhD candidate at USC, for excellent research assistance.

1 Introduction

There is an extensive econometric literature on whether log output per capita (y_{it}) of different countries $i = 1, 2, ..., n$ converges over time $t = 1, 2, ..., T$ and what may determine the level and growth in y_{it}. This paper provides a critical examination of the empirical basis of the output convergence literature in the light of recent developments in the analysis of dynamic heterogeneous panels with interactive effects, which provides the methodology needed to analyse output determination across countries. The paper provides a taxonomy of models of the speed of convergence and potential determinants of growth, time-varying and time-invariant, when the parallel trend and dynamic homogeneity assumptions that are prevalent in this literature are relaxed. It revisits the asymptotic biases of the estimators obtained from Barro (1991) type cross-country regressions of growth on initial output and from two-way fixed effect (TWFE) panel models. The estimate of the speed of convergence obtained from Barro type regressions will be biased towards zero if there are any random differences between countries even if the parallel trend assumption is maintained. The TWFE estimators are robust to random or fixed effects, but continue to result in biased estimates if the parallel trends assumption is relaxed even if dynamic homogeneity is maintained.

By adhering to the parallel trends assumption, cross-country regressions as well as TWFE panel estimators require that all countries are affected similarly by the global factors, implying that all countries have the same steady state growth rate. To allow the average growth rate to differ between countries requires interactive time effects, where the loadings on the latent global factor(s) differ across countries and can depend on time-invariant or slowly moving variables like institutions that influence the degree of openness to new ideas.[1] Furthermore, to generate the observed pattern of global growth the latent technology factor must be trended. As well as considering time-varying covariates, we distinguish between time-invariant covariates that influence the intercepts, which can only have a one-off impact on growth, and those that influence the loadings on the global factor, which give rise to sustained growth.

Nonparallel trends arise naturally if countries are differently placed to benefit from technological progress, due to historical, political or institutional factors. In the presence of nonparallel trends, the TWFE estimator of the speed of convergence is biased downward and this bias does not vanish even if $T \to \infty$. More importantly, this paper is the first to show that the TWFE estimator of the speed of convergence tends to zero, irrespective of its true value, if the technology factor is trended, which is required for the world economy to exhibit non-zero output growth. Relaxing the assumption of dynamic homogeneity is another source of the bias when the TWFE estimator is employed; similar to the ones

[1] In this paper, we focus on a single factor that we label technology for exposition, though models with interactive time effects can accomodate multiple factors, including climate. We do not pursue the role of climate change because of the measurement issues involved.

documented by Pesaran and Smith (1995) in the context of fixed effects dynamic panel data models. The empirical importance of dynamic heterogeneity in output convergence regressions was demonstrated by Lee, Pesaran and Smith (1997, 1998) and more recently by, for instance, Ong (2024).

The focus of this paper is to provide a systematic methodological investigation of the implications of relaxing the assumptions of parallel trends and dynamic homogeneity that are commonly made in the empirical growth literature. To relax these assumptions we propose the use of correlated common effect (CCE) type estimators instead of the TWFE estimates. The CCE estimator was first advanced by Pesaran (2006) for panels with strictly exogenous regressors, and has since been extended by Kapetanios, Pesaran and Yamagata (2011) to allow for unit roots in the latent factors, and further by Westerlund (2018) to account for deterministic trends amongst the latent factors. The analysis of dynamic panels with latent factors has received less attention. Chudik and Pesaran (2015a) propose dynamic CCE estimators for panels with interactive effects and heterogeneous dynamics. A related paper by Moon and Weidner (2017) proposes least squares estimators that allow for predetermined regressors (including lagged dependent variables), assuming homogeneous slopes and stationary latent factors. Everaert and De Groote (2016) have also considered CCE type estimators for panels with homogeneous dynamics. However, all these papers that allow for dynamics assume the latent factors are stationary, which is restrictive for the analysis of convergence in the case of growing economies. Assuming dynamic homogeneity, we show that the CCE estimator for the speed of convergence is consistent irrespective of whether the global technological factor is stationary or trended. The trend could be stochastic, with or without a drift, or could even be subject to a number of breaks. We refer to this estimator as DCCEP, to highlight the dynamics and the pooled nature of the underlying panel data model. We also consider a mean group version of DCCE, which we denote by DCCEMG, that allows for heterogeneous slopes, which applies generally to dynamics and the effects of time-varying covariates such as capital stocks across countries.[2]

In comparing Barro, TWFE, and DCCE type estimates, it is important to be clear that while all these methods aim to estimate the same "cross-country average speed of convergence", they differ in their treatment of parameter heterogeneity across countries. Barro regressions provide an estimate of it under the assumption of full parameter homogeneity. Were it true that there was full parameter homogeneity, convergence would be to a single level of output for all countries. Were it not true, the estimate of the average speed of convergence will be biased towards zero. Similarly, TWFE regressions provide an estimate of it under the assumption that all countries have the same steady state growth rates.

[2]Chudik and Pesaran (2015b) provide a survey of CCE type estimators, more recently, in a special issue of *The Journal of Econometrics* on panel data analysis introduced by Sarafidis and Wansbeek (2021) there are a number of CCE related papers including Juodis, Karabiyik and Westerlund (2021) on the robustness of the pooled CCE estimator.

Were this true, convergence, would be to country specific parallel trends for output. Were this not true, the estimated average speed of convergence will again be biased towards zero. The Barro or TWFE biases are asymptotic, not reduced by more information. In contrast, DCCE type estimates, with or without the assumption of dynamic homogeneity, focus on the average speed of convergence to country-specific output paths, and do not impose that countries converge in either per capita outputs or growth rates. With this in mind, we quantify the size of the biases of Barro and TWFE estimates of the speed of convergence using Penn World Table data on log output per capita (and per employee) covering 157 countries over seventy years, 1950-2019. Both unconditional and conditional Barro regressions have speeds of convergence that range between 0.3 and 0.5 per cent implying mean lags that fall between 171 and 316 years, suggesting that there is little evidence of cross country output convergence. TWFE estimates of the speed of convergence, that relate to an assumed *common growth* path across countries, are around 3 per cent with implied mean lags of around 25 years. In contrast, the DCCEMG estimates of speed of convergence that relate to *country-specific* output paths are much higher, around 14 per cent with mean lags of around 6 years, corresponding to business cycle frequencies documented in the time series literature. For example, according to NBER (National Bureau of Economic Research) the average duration of business cycle in the U.S. over the period 1945-2020 has been around 75 months or 6.25 years.[3] Similar results are also found for G7 and OECD countries. See, for example, Chuavet and Yu (2006) who also show that there exists a significant degree of synchronization of business cycles across countries driven partly by globalization and common shocks like total factor productivity shifts.[4] The rise in the estimated average speed of convergence as we relax parallel trends and dynamic homogeneity is in line with our theoretical findings, and highlight the important differences across the models that underlie them.

For determinants of growth in the long run, we focus on capital stock as the primary time-varying covariate. We argue that the DCCEMG estimators of the long run effect of capital on output is reliable, despite endogeneity concerns, since CCE type estimators allow global latent factors to simultaneously drive both capital and output variables and thus filter out the common sources of dependence between capital and output. Focussing on DCCEMG, the long run estimate of capital coefficient is around 0.48, which is quite close to the average share of capital at 0.46, obtained from PWT data.

As an example of slowly moving variables we considered years of schooling as a proxy for human capital. We found that education variable was only statistically significant when the capital stock variable was excluded, and the panel was estimated by DCCEP. Trying other variations, such as adding lagged values of the education variable or interact-

[3]See https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions

[4]Chuavet and Yu (2006, Table 2) report average expansion and recession durations for OECD and G7 countries over the period 1960Q1-2000Q1. For OECD their estimate of average business cycle duration is 7 years.

ing it with global output did not made any difference. It is likely that years of schooling is not a satisfactory measure of human capital and further investigation is needed before a definite conclusion can be reached.

We also consider the effects of democracy on output using Acemoglu, Naidu, Restrepo, and Robinson, (2019) replication data and their specification. This is an important example where the democracy variable is time-varying only for half of the countries in the panel. In our empirical investigation of the effects of democracy we also exclude countries with less than 20 time series data points for estimation, otherwise it would not be possible to relax the parallel trends assumption. This gives a sample of 148 rather than 175 countries used by Acemoglu et al. (2019). We find the TWFE estimates of the long run effect of democracy on output to be very similar in size and significance to the ones they report. However, when the parallel trend assumption is relaxed, the DCCEP estimate of the long run effect of democracy is reduced in size and becomes statistically insignificant.

We apply our suggested procedure for time-invariant variables after removing the time-varying effects using either TWFE or DCCEP estimates. From the large number of suggested determinants of the level or growth of output we choose: an indicator of geography, absolute latitude; an indicator of the potential for conflict, namely ethnolinguistic fractionalization, and an indicator of property rights, *protection from expropriation.* Whether TWFE or DCCEP estimates are used to filter out the effects of the time-varying covariates, we find ethnolinguistic fractionalization and protection from expropriation to be statistically significant with the expected signs. Latitude was not statistically significant for both filtering procedures.

Related Literature: In the vast literature on economic growth, there are many strands. A quantitative literature initiated by Baumol (1986) and Barro (1991), and more recently Kremer, Willis, and You (2022) is concerned with modelling convergence in per-capita output across countries. Another quantitative strand associated with Acemoglu and coauthors has emphasised the role of institutions and methods of identifying causality. A third strand is concerned with understanding sustained economic growth propelled by technological progress implemented through a process of learning and innovation. This strand involved theoretical work by Aghion and Howitt (1992) on how the new technology surpasses the old through a process of creative destruction.

In terms of the econometric methods, the early contribution by Baumol (1986) opened the debate. It used a cross-country regression of country growth rates on initial income and interpreted the coefficient on initial income, beta, as a measure of "unconditional" or absolute convergence. Barro (1991), Mankiw Romer and Weil (1992) and Barro and Sala-i-Martin (1992) added initial covariates and interpreted the coefficient on initial income as a measure of "conditional" convergence to country specific steady states. Nerlove (1998, 1999) labelled such cross-country regressions Baumol-Barro regressions, but we will follow the common usage and call them Barro regressions.

Before allowing for convergence using a Barro type model, Mankiw, Romer and Weil (1991) estimated a cross-country regression that did not include initial income, which explained log GDP per working age person in 1985 by corresponding values of Solow variables like investment and population growth. Similarly, Acemoglu, Johnson and Robinson (2001) estimate a cross-country regression explaining log per capita output in 1995 as a function of a measure of institutions and time-invariant variables like latitude, but not initial income. Such regressions are not informative about convergence and it is not clear how one chooses a particular year. Maseland (2021) examines the movement over time in the coefficients of 43 such "deep" determinants used by Acemoglu et al. (2001) finding the effect of many of them, including latitude, increasing from around 1960 until about 2000, around when the paper was published, and then declining.

Barro regressions were criticised on a number of counts. It was argued that such cross section snapshots were not be informative about time series processes and that, so called, "beta convergence", a negative coefficient on initial income in this regression, does not imply "sigma convergence" a falling variance of incomes across countries. In Barro regressions any intercept heterogeneity is inherently correlated because it is inherited by the regressor, namely initial income. As a result, the estimated speed of convergence is downward biased unless the country specific intercepts are identical. This bias is present even if the intercepts are randomly distributed independently of all errors. To the extent that including covariates reduces the variance of the intercepts, making them more similar, the bias is reduced. Thus estimates of the speed of convergence in what are called conditional convergence regressions are less biased towards zero than in unconditional ones. Similar issues arise with dynamic heterogeneity in the coefficients of lagged dependent variables which are inherently correlated with the regressors.

The title of Nerlove (1998) posed a central question: "growth rate convergence, fact or artifact?" Like him we will argue that such regressions are badly biased. However, repeated criticisms of Barro regressions, including that by Friedman (1992) who asked "Do old fallacies ever die?", did not kill them and they have been recently resurrected by Kremer et al. (2022). But we hope the econometric analysis that we provide will put one last nail in the coffin.

Subsequent literature, such as Islam (1995), used panels with country and time effects. These widely used two-way fixed effects panel data models, also known as fixed-effect, time-effect, imposed a parallel trend assumption, which we will relax. After a rapid growth in the literature, interest waned because there seemed little evidence for "unconditional" convergence and while there was evidence for "conditional" convergence, specification of the appropriate covariates was problematic. An early survey is Temple (1999). Durlauf (2009) documents "The rise and fall of cross-country growth regressions".

Subsequent papers such as Barro (2015) and Acemoglu et al. (2019) have argued for the importance of institutions, such as democracy, in panel studies of conditional convergence.

After an extensive survey, Johnson and Papageorgiou (2020) concluded that "there is a broad consensus of no evidence supporting absolute convergence in cross-country per capita incomes — that is poor countries do not seem to be unconditionally catching up to rich ones." However, that consensus was challenged in papers by Kremer et al. (2022), and Patel, Sandefur and Subramanian (2021). Kremer et al. (2022) resurrected Barro regressions and claimed that these provided evidence for unconditional convergence since 2000, if not from earlier dates, and argued that while conditional convergence held throughout the period, absolute convergence did not hold initially, but as human capital and policies improved in poorer countries, the difference in institutions across countries has shrunk, leading to unconditional convergence, a conclusion we challenge. Smith (2024) provides a recent review of some of the econometric methods used in the convergence literature.

Outline of the paper: The rest of the paper is set out as follows: Section 2 provides an overview of the models of output convergence. Section 3 onwards considers estimation methods. Section 3 investigates the bias of the Barro cross-country regression when there is intercept heterogeneity with and without conditioning on time-invariant covariates. Section 4, examines the bias of the TWFE estimator when the assumption of parallel trends is relaxed. Section 5 considers estimation of models with nonparallel trends and homogeneous dynamics using the DCCEP estimator. Section 6 considers estimation of models with nonparallel trends and heterogeneous dynamics using the DCCEMG estimator. Section 7 considers time-varying covariates, like capital. Section 8 considers the identification and estimation of the coefficients of time-invariant variables, like geography, or slowly moving ones like climate or institutions. Section 9 examines the quantitative importance of the theoretical results using data from the Penn World Tables and the replication files of Acemoglu et al. (2019) and Kremer et al. (2022). Section 10 contains some concluding remarks.

2 An overview of models of output convergence

This section provides an overview of the models that will be examined to provide context for the subsequent discussion of estimation methods. To simplify the exposition here we assume balanced panels, but use both balanced and unbalanced panels in our empirical analysis, depending on the estimation procedure under consideration.

Much of the literature can be related to the widely used TWFE autoregressive (AR) panel data model. Assuming a linear specification with first-order dynamics, the baseline panel model is:

$$y_{it} = \alpha_i + f_t + \rho y_{i,t-1} + u_{it}, \tag{1}$$

where y_{it} is the logarithm of per capita output in countries $i = 1, 2, ..., n$, at time periods

$t = 1, 2, ..., T$, with $|\rho| < 1$. The α_i are heterogeneous country-specific intercepts. The time effect, f_t, can be viewed as a latent global factor, such as technology, assumed here to have the same effects on y_{it} across all i. Thus without loss of generality the TWFE specification (1) can be written equivalently in deviation form as

$$y_{it} - g_t = \alpha_i + \rho\left(y_{i,t-1} - g_{t-1}\right) + u_{it}, \tag{2}$$

where $f_t = g_t - \rho g_{t-1}$ and

$$g_t = \sum_{s=0}^{\infty} \rho^s f_{t-s} \tag{3}$$

In this representation, and given that $|\rho| < 1$, then for each $i = 1, 2, ..., n$

$$E\left(y_{it} - g_t\right) = \frac{\alpha_i}{1-\rho}, \text{ and } E\left(\Delta y_{it}\right) = E\left(\Delta g_t\right).$$

Thus under the TWFE specification, unless $\alpha_i = \alpha$ all i, there is no unconditional convergence. Log output paths will be parallel to one another, with all countries having the same mean output growth, the so called "parallel trend" assumption.

The parallel trend assumption can be relaxed by allowing heterogeneous loadings on the global factor giving a more general, interactive time effects, specification where f_t is replaced by $\gamma_i f_t$, with γ_i representing the factor loading for country i.[5] With the parallel trend assumption relaxed, mean output growth will no longer be equal across countries and we have

$$E\left(\Delta y_{it}\right) = \gamma_i \sum_{s=0}^{\infty} \rho^s E\left(\Delta f_{t-s}\right),$$

which simplifies to $E\left(\Delta y_{it}\right) = (1-\rho)^{-1}\gamma_i E\left(\Delta f_t\right)$, if Δf_t is stationary. Equivalently, relaxing the parallel trend assumption using the deviation form, (2), we have $E\left(\Delta y_{it}\right) = \gamma_i E\left(\Delta g_t\right)$. The two representations of the cross-country mean growth rates are equivalent given (3). Differences in factor loadings across countries could be due to differences in trade, capital and immigration policies across countries due to geopolitical and other considerations. We discuss the implications of differences in factor loadings in Section 4. In practice, these loadings could also vary over time, thus introducing another degree of complexity in the analysis of output convergence.

Assuming that ρ lies in the range $(-1, +1)$, the parameter of interest is $\phi = 1 - \rho > 0$, which measures the speed of convergence of $y_{it} - \gamma_i g_t$ to its steady state value of $E\left(y_{it} - \gamma_i g_t\right) = (1-\rho)^{-1}\alpha_i$. For estimation, or when there are higher order lags, it is more convenient to work with the error correction representations of (1) or (2) given by

$$\Delta y_{it} = \alpha_i + \gamma_i f_t - \phi y_{i,t-1} + u_{it}, \text{ or } \Delta\left(y_{it} - \gamma_i g_t\right) = \alpha_i - \phi\left(y_{i,t-1} - \gamma_i g_{t-1}\right) + u_{it}. \tag{4}$$

[5] As noted above, given the growth context for exposition we treat there as being a single global factor. But it easily generalises to a multi-factor model: $\boldsymbol{\gamma}_i'\mathbf{f}_t$, with $\mathbf{f}_t$ representing a vector of latent variables, and $\boldsymbol{\gamma}_i$ the associated vector of factor loadings.

In these specifications it is assumed that country-specific error terms, u_{it}, are serially uncorrelated and cross-sectionally independent with mean zero and variances σ^2_{iu}, for $i = 1, 2, ..., n$ that vary across countries but are fixed over time. Assuming u_{it} are cross-sectionally independent could be problematic even if interactive time effects are included in the panel, since there may be spatial or spill-over effects that cannot be eliminated by use of factors. See Pesaran and Xie (2026).

Higher order dynamics might be required to allow for possible serial correlation in u_{it}. In practice, this is achieved by adding lagged values of the output growth variables to the error correction representations, (4). Assuming a p^{th} order panel AR for log output

$$y_{it} = \alpha_i + \gamma_i f_t + \textstyle\sum_{\ell=1}^{p} \rho_\ell y_{i,t-\ell} + u_{it} \tag{5}$$

yields the following $p-1$ order panel error correction specification

$$\Delta y_{it} = \alpha_i + \gamma_i f_t - \phi y_{i,t-1} + \sum_{\ell=1}^{p-1} \psi_\ell \Delta y_{i,t-\ell} + u_{it}, \tag{6}$$

where $\phi = 1 - \sum_{\ell=1}^{p} \rho_\ell$.

As we shall see the choice of the lag order, p, plays an important role for the empirical validity of some of the estimation procedures. One prominent example arises when Barro regressions are used to estimate ϕ, which requires that there is only a single lag, in addition to a number of other assumptions that will be discussed below in some detail.

The TWFE specification, (1), and its extension in (6) imposes that all countries converge to their steady state per capita output levels at the same speed, ϕ, which is a rather strong assumption. Countries differ in their political institutions and climatic conditions which affect the speed with which they adapt to shocks. We discuss the implications of these assumptions in Section 6. Accordingly, we also consider a fully heterogeneous version of (6) that allows for ϕ and ψ_ℓ to differ across i, namely

$$\Delta y_{it} = \alpha_i + \gamma_i f_t - \phi_i y_{i,t-1} + \sum_{\ell=1}^{p-1} \psi_{i,\ell} \Delta y_{i,t-\ell} + u_{it}. \tag{7}$$

The parameter of interest is the average speed of convergence across countries, defined by $E(\phi_i)$. Adding a vector of time-varying covariates, $\mathbf{x}_{it}$, is straightforward giving an error correction model of the form:

$$\begin{aligned} \Delta y_{it} = {} & \alpha_i + \gamma_i f_t - \phi_{y,i} y_{i,t-1} + \sum_{\ell=1}^{p-1} \psi_{y,i,\ell} \Delta y_{i,t-\ell} \\ & + \boldsymbol{\phi}'_{x,i} \mathbf{x}_{it} + \sum_{\ell=1}^{p-1} \boldsymbol{\psi}'_{x,i,\ell} \Delta \mathbf{x}_{i,t-\ell} + u_{it}. \end{aligned} \tag{8}$$

In addition to the time-varying $\mathbf{x}_{it}$, the literature has emphasised the importance of covariates that do not vary over time or vary very slowly. These may influence the intercepts α_i, which under the parallel trends assumption are the main source of differences in steady state output levels. Once the parallel trend assumption is relaxed, it is the differences in output growth paths, $\gamma_i f_t$, that will matter, since countries that are on different growth paths will diverge and differences in α_i will be of secondary importance. It is the differences in γ_i across i that determine the changing dispersion of output per capita across countries. In our empirical analysis we consider the following models for the determination of α_i and γ_i :

$$\alpha_i = \alpha_\alpha + \boldsymbol{\theta}'_\alpha \mathbf{z}_{i\alpha} + \eta_{i\alpha}, \tag{9}$$

$$\gamma_i = \alpha_\gamma + \boldsymbol{\theta}'_\gamma \mathbf{z}_{i\gamma} + \eta_{i\gamma}, \tag{10}$$

where $\mathbf{z}_{i\alpha}$ and $\mathbf{z}_{i\gamma}$ are likely to include climatic, institutional or policy variables; $\eta_{i\alpha}$ and $\eta_{i\gamma}$ are random errors, representing the unexplained parts of α_i and γ_i; and α_a, α_γ, θ_α and θ_γ are fixed coefficients. There is an important difference between the two types of variables, $\mathbf{z}_{i\alpha}$ and $\mathbf{z}_{i\gamma}$. Changes in the latter could induce growth effects that are sustained, whilst the effects of changing $\mathbf{z}_{i\alpha}$, could at best result in a once and for all change in the level of output, without generating sustained growth effects. Thus policies that change the value of γ_i are likely to have more fundamental impacts on output.

Abstracting from time-varying effects, and for simplicity assuming that $\mathbf{z}_{i\alpha}$ and $\mathbf{z}_{i\gamma}$ have no variables in common, then using (9) and (10) in (7) and averaging over t we have

$$a_{iT} = \bar{y}_{i\circ} - \textstyle\sum_{\ell=1}^{p} \rho_{i,\ell} \bar{y}_{i,-\ell} = \alpha_T + \boldsymbol{\theta}'_T \mathbf{z}_i + \eta_{iT} + \bar{u}_{i\circ}, \tag{11}$$

where $\bar{y}_{i,-\ell} = T^{-1}\sum_{t=1}^{T} y_{i,t-\ell}$, $\alpha_T = \alpha_\alpha + \alpha_\gamma \bar{f}_T$, $\boldsymbol{\theta}_T = (\boldsymbol{\theta}'_\alpha, \boldsymbol{\theta}'_\gamma \bar{f}_T)'$, $\mathbf{z}_i = \left(\mathbf{z}'_{i\alpha}, \mathbf{z}'_{i\gamma}\right)'$, $\eta_{iT} = \eta_{i\alpha} + \eta_{i\gamma}\bar{f}_T$, and $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^{T} u_{it}$. Equation (8) could also be used to filter the dynamics. Therefore, in principle the time-invariant effects, $\boldsymbol{\theta}_T$, can be identified using cross-country regressions where the dynamics have been filtered out, and $\mathbf{z}_i$ are distributed independently of the composite errors, $\eta_{iT} + \bar{u}_{i\circ}$.

We shall return to the problem of identification and estimation of time-invariant effects in Section 8.

3 Estimation of the speed of convergence by Barro regressions

We begin with the Barro regression. This can be interpreted in terms of the panel AR(1) model given by (1) assuming u_{it} are serially uncorrelated and cross-sectionally independent, and the fixed effects, α_i, follow a random coefficient specification. The unconditional Barro regression involves running a cross-country regression of $y_{iT} - y_{i0}$ on an intercept and y_{i0}. The estimate of $\phi = 1 - \rho$ is then recovered from the least squares estimate of the coefficient of y_{i0}. Specifically, the unconditional Barro regression can be written as

$$y_{iT} - y_{i0} = a_T - (1-\rho^T)y_{i0} + v_{iT}, \tag{12}$$

where it is assumed that the panel is balanced in the sense that y_{i0} and y_{iT} are available for all i. In addition to assuming homogeneous dynamics, the Barro regression also requires y_{i0} and v_{iT} are independently distributed.

To investigate the relationship of Barro regressions to the TWFE panel data model, we assume[6]

$$\alpha_i = \alpha + \boldsymbol{\theta}'\mathbf{z}_i + \eta_i, \tag{13}$$

and eliminate the latent factor, f_t, from (1) by use of cross section averages. Averaging y_{it} given by (1), and augmented by (13) over i we have

$$\bar{y}_{\circ t} = \alpha + \boldsymbol{\theta}'\bar{\mathbf{z}} + \bar{\eta} + f_t + \rho\bar{y}_{\circ,t-1} + \bar{u}_{\circ t}, \tag{14}$$

where $\bar{y}_{\circ t} = n^{-1}\sum_{i=1}^n y_{it}$, $\bar{u}_{\circ t} = n^{-1}\sum_{i=1}^n u_{it}$, $\bar{\mathbf{z}} = n^{-1}\sum_{i=1}^n \mathbf{z}_i$, and $\bar{\eta} = n^{-1}\sum_{i=1}^n \eta_i$. Then f_t can be eliminated by subtracting (14) from (1) to obtain

$$y_{it} - \bar{y}_{\circ t} = \boldsymbol{\theta}'(\mathbf{z}_i - \bar{\mathbf{z}}) + \rho\left(y_{i,t-1} - \bar{y}_{\circ,t-1}\right) + u_{it} - \bar{u}_{\circ t} + (\eta_i - \bar{\eta}). \tag{15}$$

Now solving forward from the initial values $y_{i0} - \bar{y}_0$, where $\bar{y}_0 = n^{-1}\sum_{i=1}^n y_{i0}$, we have

$$y_{iT} - \bar{y}_{\circ T} = \rho^T\left(y_{i0} - \bar{y}_0\right) + \left(\frac{1-\rho^T}{1-\rho}\right)\left[\boldsymbol{\theta}'(\mathbf{z}_i - \bar{\mathbf{z}}) + (\eta_i - \bar{\eta})\right] + \sum_{s=0}^{T-1}\rho^s\left(u_{i,T-s} - \bar{u}_{\circ,T-s}\right), \tag{16}$$

After some rearrangements, (16) can be written equivalently as

$$y_{iT} - y_{i0} = a_T + b_T y_{i0} + \mathbf{c}_T'\mathbf{z}_i + \zeta_{iT}, \tag{17}$$

where

$$a_T = \bar{y}_{\circ T} - \rho^T\bar{y}_0 - \left(\frac{1-\rho^T}{1-\rho}\right)(\boldsymbol{\theta}'\bar{\mathbf{z}} + \bar{\eta}) - \sum_{s=0}^{T-1}\rho^s\bar{u}_{\circ,T-s}, \tag{18}$$

$$b_T = -(1-\rho^T), \quad \mathbf{c}_T = \left(\frac{1-\rho^T}{1-\rho}\right)\boldsymbol{\theta}, \tag{19}$$

and

$$\zeta_{iT} = \left(\frac{1-\rho^T}{1-\rho}\right)\eta_i + \sum_{s=0}^{T-1}\rho^s u_{i,T-s}. \tag{20}$$

Comparing this representation with the Barro regression given by (12), we note that (17) coincides with what is known as conditional Barro regression and reduces to unconditional

[6] Recall that under parallel trends, assumed when using Barro regressions, $\gamma_i = \gamma$ and only α_i is affected by time-invariant variables, $\mathbf{z}_i$

Barro regression when $\boldsymbol{\theta} = \mathbf{0}$. The parameters of Barro regressions are b_T and $\mathbf{c}_T$, and ρ and $\boldsymbol{\theta}$ are not directly estimated when one runs a standard conditional Barro regression. However, ϕ and $\boldsymbol{\theta}$ can be estimated using (recalling that $|\rho| < 1$, by assumption)

$$\phi = 1 - \rho = 1 - (1 + b_T)^{1/T}$$
$$\boldsymbol{\theta} = \left(\frac{1-\rho}{1-\rho^T}\right) \mathbf{c}_T = -\left[\frac{1-(1+b_T)^{1/T}}{b_T}\right] \mathbf{c}_T.$$

Clearly, a_T and $\mathbf{c}_T$ varies with T but not ϕ and θ. For comparability with pooled panel estimates of θ when reporting Barro's estimates one needs to report estimates of $\boldsymbol{\theta}$, in addition to the estimates of b_T and $\mathbf{c}_T$ that are routinely reported in the literature. It is also clear that estimates of ϕ and $\boldsymbol{\theta}$ based on Barro's regression estimates of b_T and $\mathbf{c}_T$ will be biased, due to the non-zero correlations between y_{i0} and ζ_{iT}, and between y_{i0} and $\mathbf{z}_i$, as established below.

Considering (16), there are a number of distinct possible sources of bias which come from the correlation between either of the regressors y_{i0} and $\mathbf{z}_i$ and any of the three elements of the composite error term of the Barro regression, namely $\left(\frac{1-\rho^T}{1-\rho}\right)\eta_i$, $\sum_{s=0}^{T-1}\rho^s u_{i,T-s}$ and $\sum_{s=0}^{T-1}\rho^s \bar{u}_{T-s}$. Below we show that as long as $\sigma_\eta^2 \neq 0$, then any heterogeneity in the α_i, even if it is random, will cause bias because of the inevitable correlation between initial output y_{i0} and η_i, the random component of α_i. In addition, serial correlation in the u_{it} will cause y_{i0} to become correlated with cross section averages of $u_{iT}, u_{i,T-1}, ...$ In the case of $\mathbf{z}_i$, while the shocks - pandemics, revolutions and the like - may be exogenous, many of the elements of $\mathbf{z}_i$ are likely to be adaptive to past shocks, causing them to be correlated with them.

3.1 Asymptotic bias of the Barro estimator

For the OLS estimator of b_T from (17) to be consistent it is required that $E\left(y_{i0}\zeta_{iT}\right) = 0$, for $i = 1, 2, ..., n$; irrespective of whether we consider unconditional or conditional Barro regressions. When conditioning variables are added to the regression it is also required that $E(\mathbf{z}_i\zeta_{iT}) = 0$. Both of these conditions involve restrictions on the way initial values, y_{i0}, are generated. Since the random component of α_i, namely η_i, is specific to country i it is likely that y_{i0} will also depend on η_i. Such a dependence arises irrespective of whether the $\{y_{it}\}$ processes have started from a finite past prior to the initial date $t = 0$, or from a very distant past. It also does not matter if the countries differ in their start dates, say at time $t = -M_i$, where M_i is a positive integer. In this general set up, supposing country i was established (started) with $y_{i,-M_i}$ at time $t = -M_i$, then iterating (1) forward from

$y_{i,-M_i}$, (similar to the way equation (16) was derived) we obtain

$$y_{i0} - \bar{y}_0 = \rho^{M_i}(y_{i,-M_i} - \bar{y}_{\circ,-M_i}) + \left(\frac{1-\rho^{M_i}}{1-\rho}\right)\left[\boldsymbol{\theta}'(\mathbf{z}_i - \bar{\mathbf{z}}) + (\eta_i - \bar{\eta})\right] \quad (21)$$
$$+ \sum_{s=0}^{M_i-1} \rho^s (u_{i,-s} - u_{\circ,-s}).$$

Under the panel AR(1) model (1) u_{it} are serially uncorrelated. As a first order approximation it is also plausible to assume that u_{it} is distributed independently of f_t, η_i and $\mathbf{z}_i$. Even under these assumptions, using (20), we have a non-zero correlation between y_{i0} and ζ_{iT}. Specifically,

$$E\left(y_{i0}\zeta_{iT}\right) = \left(\frac{1-\rho^T}{1-\rho}\right)\left(\frac{1-\rho^{M_i}}{1-\rho}\right)\sigma_\eta^2, \quad (22)$$

where $\sigma_\eta^2 = var(\eta_i)$, and renders Barro regression estimators of ϕ and $\boldsymbol{\theta}$ inconsistent if $\sigma_\eta^2 > 0$, irrespective of whether the regression is estimated unconditionally or conditional on $\mathbf{z}_i$. The magnitude of the bias could depend on the conditioning variables used and whether they are uncorrelated with u_{it}.[7]

We now derive the asymptotic bias of the unconditional Barro estimator of ρ_0 the true value of ρ, which is obtained as the least squares estimator of $b_T = -(1-\rho_0^T)$ from the OLS regression of $y_{iT} - y_{i0}$ on $y_{i0} - \bar{y}_0$, namely

$$\hat{b}_T = \frac{\sum_{i=1}^n \left(y_{i0} - \bar{y}_0\right)\left(y_{iT} - y_{i0}\right)}{\sum_{i=1}^n \left(y_{i0} - \bar{y}_0\right)^2},$$

recalling that $\bar{y}_{\circ 0} = n^{-1}\sum_{i=1}^n y_{i0}$. Then using (17)

$$\hat{b}_T - b_T = \frac{n^{-1}\sum_{i=1}^n \left(y_{i0} - \bar{y}_0\right)\left(\mathbf{c}_T'\mathbf{z}_i + \zeta_{iT}\right)}{n^{-1}\sum_{i=1}^n \left(y_{i0} - \bar{y}_0\right)^2}$$
$$\to_p \frac{\lim_{n\to\infty} n^{-1}\sum_{i=1}^n E\left[\left(y_{i0} - \bar{y}_0\right)\left(\mathbf{c}_T'\mathbf{z}_i + \zeta_{iT}\right)\right]}{\lim_{n\to\infty} n^{-1}\sum_{i=1}^n E\left(y_{i0} - \bar{y}_0\right)^2}.$$

To simplify the derivations we assume that $M_i \to \infty$ for all i (all processes have started in a distant past with $|\rho| < 1$) and note that in this case, using (21), we have

$$y_{i0} - \bar{y}_0 = \left(\frac{1}{1-\rho_0}\right)\left[\boldsymbol{\theta}'\left(\mathbf{z}_i - \bar{\mathbf{z}}\right) + \eta_i - \bar{\eta}\right] + \sum_{s=0}^{\infty} \rho_0^s \left(u_{i,-s} - \bar{u}_{\circ,-s}\right).$$

Using this result together with (20) and noting that u_{it} and η_i are assumed to be cross-sectionally independent we obtain

$$E\left[\left(y_{i0} - \bar{y}_0\right)\left(\mathbf{c}_T'\mathbf{z}_i + \zeta_{iT}\right)\right] = \frac{\left(1-\rho_0^T\right)}{\left(1-\rho_0\right)^2}\left[\boldsymbol{\theta}'Var\left(\mathbf{z}_i\right)\boldsymbol{\theta} + \sigma_\eta^2 - E\left(\eta_i\bar{\eta}\right)\right]. \quad (23)$$

[7] The endogeneity of the conditioning variables, which has been widely discussed in the literature, is not essential for the bias which arises if $\sigma_\eta^2 > 0$, which can be allowed for by using TWFE in a model with parallel trends.

But since η_i are independently distributed then $E\left(\eta_i\bar{\eta}\right) = n^{-1}\sigma_\eta^2$. Similarly

$$E\left(y_{i0}-\bar{y}_0\right)^2 = \frac{1-\rho_0^T}{\left(1-\rho_0\right)^2}\left[\boldsymbol{\theta}' Var\left(\mathbf{z}_i\right)\boldsymbol{\theta}+\sigma_\eta^2\right] + \frac{\sigma_{iu}^2}{1-\rho_0^2} + O(n^{-1}).$$

Hence, as $n \to \infty$

$$\hat{b}_T - b_T = \hat{\rho}^T - \rho_0^T \to_p \frac{\frac{\left(1-\rho_0^T\right)(1+\rho_0)}{1-\rho}\left(\frac{\boldsymbol{\theta}'\boldsymbol{\Omega}_z\boldsymbol{\theta}+\sigma_\eta^2}{\bar{\sigma}_u^2}\right)}{1+\left(\frac{1+\rho_0}{1-\rho_0}\right)\left(\frac{\boldsymbol{\theta}'\boldsymbol{\Omega}_z\boldsymbol{\theta}+\sigma_\eta^2}{\bar{\sigma}_u^2}\right)} > 0, \tag{24}$$

where

$$\boldsymbol{\Omega}_z = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n Var\left(\mathbf{z}_i\right), \text{ and } \bar{\sigma}_u^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sigma_{iu}^2$$

The magnitude of the bias, $\hat{\rho}^T - \rho_0^T$, depends on ρ_0, T and the ratio $\left(\boldsymbol{\theta}'\boldsymbol{\Omega}_z\boldsymbol{\theta}+\sigma_\eta^2\right)/\bar{\sigma}_u^2$, which represent the size of cross section dispersions relative to the average of the time series dispersions. If we set this ratio to unity, the bias of the Barro estimator is $(1-\rho_0^T)(1+\rho_0)/2$, which could be substantial considering that ρ_0 is likely to be close to 1. For instance, if $\rho_0 = 0.8$, and $T = 10$, then the asymptotic bias of estimating ρ_0^T is given by $(1-\rho_0^T)(1+\rho_0)/2 = 0.803$,which implies $\hat{\rho} \to_p 0.99$, namely a convergence rate of 1% instead of the actual value of 20%.[8] Also the bias in $\hat{b}_T$ is upward for any T, and rises with T since $|\rho| < 1$. The longer the time series data the more pronounced the bias of the Barro estimator will become.[9]

By conditioning on the observed characteristics, it is possible to reduce the heterogeneity bias of the Barro estimator if the conditioning variables, $\mathbf{z}_i$, are independently distributed of the shocks, u_{it}. The larger the size of $\boldsymbol{\theta}'\boldsymbol{\Omega}_z\boldsymbol{\theta}/\bar{\sigma}_u^2$ the greater the scope for bias reduction by means of conditioning.

In short, the Barro regression can only deliver consistent estimates of the speed of convergence of output across countries, if α_i is a *known deterministic* function of $\mathbf{z}_i$, the errors of the underlying panel AR(1) model are serially uncorrelated, and the idiosyncratic errors are unrelated to all time-invariant regressors such as climate, latitude, demographic and institutional factors, for example. These are very strong assumptions and are unlikely to hold.

Even if it is argued that α_i is a deterministic function of a known vector of time-invariant variables $\mathbf{z}_i$ and consequently $\sigma_\eta^2 = 0$, the Barro estimator could still be inconsistent due possible correlations between $\mathbf{z}_i$ and current and past values of the idiosyncratic errors, $\{u_{i,T+1}, u_{i,T}, ..., u_{i0}, ...\}$, as well as error serial correlation. Derivation of the Barro cross-country regression requires the underlying panel data model to follow an AR(1)

[8]Since $\hat{\rho}^T - \rho^T \to_p 0.803$, $\hat{\rho}^T \to_p 0.803 + \rho^{10} = 0.803 + 0.1074 = 0.9103$. $\hat{\rho}^T = 0.9103$ implies $\hat{\rho} \to_p 0.9883$.

[9]Kremer et al. (2022, p338) using $T = 10$ found divergence in income per capita at a rate of 0.5% annually for 1985 to 1995 then convergence at rate 0.7% from 2005-2015.

specification and cannot accommodate higher order dynamics. As a result if such higher order dynamics are ignored, the errors in (1), u_{it}, will become serially correlated and the assumption that y_{i0} and ζ_{iT} in (17) are uncorrelated will not hold any longer. The non-zero correlation between y_{i0} and ζ_{iT} also renders the least squares estimates of the time-invariant effects, $\boldsymbol{\theta}$ inconsistent if $\sigma_{\eta}^2 > 0$ and/or higher order dynamics are required.

Whilst random coefficient models are good at dealing with parameter heterogeneity in the context of static panels, assuming random effects (intercepts and/or slopes) in dynamic panels invariably lead to inconsistent estimates when heterogeneity is ignored. The Barro regression basically fails because it cannot allow for random differences across countries. All randomness must be taken into account by assuming an exact relationship between α_i and $\mathbf{z}_i$.

4 Asymptotic bias of TWFE estimator in presence of nonparallel trends

To avoid the bias of Barro regressions we need to turn to panel data techniques, the most popular of which is TWFE specification, given by (1). But, as is widely acknowledged, an important limitation of the TWFE estimator is the parallel trends assumption that requires the time effects, f_t, have the same impact on all countries.[10] This imposes identical steady state growth rates across all countries, irrespective of their differences in geography, endowments, and climatic conditions. Countries also differ in degree of openness to flows of goods, capital, people and ideas and these factors will influence the extent to which technology diffuses across countries. The obstacles to such flows may come from domestic institutions or from external constraints. Sanctions for instance are designed to cut countries off. Any similarities in the effects of latent factors across countries should be empirically investigated rather than assumed, *a priori*.

Starting from (1), we relax the parallel trends assumption but for now retain the slope homogeneity assumption, $\rho_i = \rho_0$, where ρ_0 is the true value of ρ. Specifically, we suppose that

$$y_{it} = \alpha_i + \gamma_i f_t + \rho_0 y_{i,t-1} + u_{it}, \tag{25}$$

where $|\rho_0| < 1$. We first consider the implications of assuming parallel trends when in reality the effects of the global factor, f_t, on country outputs are heterogeneous, namely $Var(\gamma_i) = \sigma_{\gamma}^2 > 0$. We also assume that f_t is strong in the sense that its effects is pervasive across countries and $E(\gamma_i) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \gamma_i > 0$. This condition is necessary for the global output to have a non-zero mean growth rate defined by $\bar{g}_t = n^{-1}\sum_{i=1}^{n} \Delta y_{it}$.

[10] The parallel trends is also a core assumption in causal inference, particularly in the Difference-in-Difference (DiD) analysis where it is assumed that the trend components of the outcomes for treated and control groups are the same. In the case of current application, the analysis is further complicated due to the dynamics, often abstracted from in DiD analysis.

The TWFE estimator of ρ_0 is given by

$$\hat{\rho}_{TWFE} = \frac{(\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1})'\mathbf{M}_T(\mathbf{y}_{i\circ} - \bar{\mathbf{y}})}{\sum_{i=1}^{n}(\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1})'\mathbf{M}_T(\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1})}, \tag{26}$$

where $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\tau_T\tau_T'$, τ_T is a $T \times 1$ vector of ones, $\mathbf{y}_{i\circ} = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, y_{i1}, ..., y_{i,T-1})'$, $\bar{\mathbf{y}} = (\bar{y}_{\circ 1}, \bar{y}_{\circ 2}, ..., \bar{y}_{\circ T})'$, $\bar{\mathbf{y}}_{-1} = (\bar{y}_{\circ 0}, \bar{y}_{\circ 1}, ..., \bar{y}_{\circ, T-1})'$, and $\bar{y}_{\circ t} = n^{-1}\sum_{i=1}^{n} y_{it}$. To derive an expression for the bias of $\hat{\rho}_{TWFE}$ we assume that f_t that u_{it} are independently distributed, u_{it} are cross-sectionally independent.

Under (25)

$$\mathbf{y}_{i\circ} - \bar{\mathbf{y}} = (\alpha_i - \bar{\alpha})\,\tau_T + (\gamma_i - \bar{\gamma})\,\mathbf{f} + \rho_0(\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1}) + \mathbf{u}_{i\circ} - \bar{\mathbf{u}} \tag{27}$$

where $\mathbf{f} = (f_1, f_2, ..., f_T)'$, $\bar{\mathbf{u}} = (\bar{u}_{\circ 1}, \bar{u}_{\circ 2}, ..., \bar{u}_{\circ T})'$, with $\bar{u}_{\circ t} = n^{-1}\sum_{i=1}^{n} u_{it}$, Using it in (26)

$$\hat{\rho}_{TWFE} - \rho_0 = \frac{T^{-1}n^{-1}\sum_{i=1}^{n}(\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1})'\mathbf{M}_T\left[(\gamma_i - \bar{\gamma})\,\mathbf{f} + \mathbf{u}_{i\circ} - \bar{\mathbf{u}}\right]}{T^{-1}n^{-1}\sum_{i=1}^{n}(\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1})'\mathbf{M}_T(\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1})}. \tag{28}$$

Also solving (25) forward, from an initial value in a distant past, we have

$$y_{it} = \alpha_i/(1-\rho) + \gamma_i g_t + v_{it}, \tag{29}$$

where

$$g_t = \sum_{j=0}^{\infty}\rho^j f_{t-j}, \text{ and } v_{it} = \sum_{j=0}^{\infty}\rho^j u_{i,t-j}, \tag{30}$$

which in turn gives

$$\mathbf{y}_{i,-1} - \bar{\mathbf{y}}_{-1} = (\gamma_i - \bar{\gamma}_n)\,\mathbf{g}_{-1} + \mathbf{v}_{i,-1} - \bar{\mathbf{v}}_{-1}, \tag{31}$$

where $\bar{\gamma}_n = n^{-1}\sum_{i=1}^{n}\gamma_i$, $\mathbf{g}_{-1} = (g_0, g_1, ..., g_{T-1})'$, and $\mathbf{v}_{i,-1}$ and $\bar{\mathbf{v}}_{-1}$ are defined analogously to $\mathbf{y}_{i,-1}$ and $\bar{\mathbf{y}}_{-1}$. Using this result, the numerator of (28) becomes

$$\begin{aligned} Num_{TWFE} = \left(n^{-1}\sum_{i=1}^{n}(\gamma_i - \bar{\gamma}_n)^2\right)\left(T^{-1}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{f}\right) + T^{-1}n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{v}}_{i,-1}'\mathbf{M}_T\left(\mathbf{u}_{i\circ} - \bar{\mathbf{u}}\right) \\ + T^{-1}n^{-1}\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})\,\mathbf{g}_{-1}'\mathbf{M}_T\left(\mathbf{u}_{i\circ} - \bar{\mathbf{u}}\right). \end{aligned}$$

where $\tilde{\mathbf{v}}_{i,-1} = \mathbf{v}_{i,-1} - \bar{\mathbf{v}}_{-1}$. Since by assumption u_{it} and f_t (and hence g_t) are independently distributed, and u_{it} are cross-sectionally independent, then for a fixed T we have

$$Num_{TWFE} = \sigma_\gamma^2\left(T^{-1}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{f}\right) + T^{-1}\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left[\tilde{\mathbf{v}}_{i,-1}'\mathbf{M}_T\left(\mathbf{u}_{i\circ} - \bar{\mathbf{u}}\right)\right] + O_p\left(n^{-1/2}\right), \tag{32}$$

where $\sigma_\gamma^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(\gamma_i - \bar{\gamma}_n\right)^2$.

Consider now the denominator of (28) and using (31) note that

$$\begin{aligned} Den_{TWFE} &= T^{-1}n^{-1}\sum_{i=1}^{n}(\mathbf{y}_{i,-1}-\bar{\mathbf{y}}_{-1})'\mathbf{M}_T(\mathbf{y}_{i,-1}-\bar{\mathbf{y}}_{-1}) \\ &= \sigma_\gamma^2\left(T^{-1}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{g}_{-1}\right)+T^{-1}\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}E\left(\tilde{\mathbf{v}}_{i,-1}'\mathbf{M}_T\tilde{\mathbf{v}}_{i,-1}'\right)+O_p\left(n^{-1/2}\right) \end{aligned} \tag{33}$$

Using (32) and (33) in (28) now yields (for a fixed T)

$$p\lim_{n\to\infty}\left(\hat{\rho}_{TWFE}-\rho_0\right)=\frac{\sigma_\gamma^2\left(T^{-1}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{f}\right)+\lim_{n\to\infty}(nT)^{-1}\sum_{i=1}^{n}E\left[\tilde{\mathbf{v}}_{i,-1}'\mathbf{M}_T\left(\mathbf{u}_{i\circ}-\bar{\mathbf{u}}\right)\right]}{\sigma_\gamma^2\left(T^{-1}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{g}_{-1}\right)+\lim_{n\to\infty}(nT)^{-1}\sum_{i=1}^{n}E\left(\tilde{\mathbf{v}}_{i,-1}'\mathbf{M}_T\tilde{\mathbf{v}}_{i,-1}'\right)} \tag{34}$$

When the parallel trend assumption does not hold $\sigma_\gamma^2>0$ and in general the TWFE estimator of ρ_0 will be biased.

The second term in the numerator of (34) is similar to the Nickell (1981) bias and is of order T^{-1}, and vanishes as $T\to\infty$. It is present whether there is a factor or not. The first term arises from the presence of the factor and the magnitude of the asymptotic bias of the TWFE estimator critically depends on the time series properties of f_t. To highlight this point we consider two cases: (a) when f_t is covariance stationary, (b) when f_t is trended or follows a random walk with a non-zero drift, namely $f_t=\mu t+s_t$ ($\mu\neq 0$). The latter specification is particularly relevant to the growth convergence literature where it is important to allow for the possibility of a non-zero steady state output growth for the global economy.

4.1 The bias of the TWFE estimator under nonparallel trends with a stationary latent factor

Under the assumption of a stationary factor, without loss of generality, we assume that f_t has mean zero with the autocovariance function, $\gamma_f(|s-s'|)=E(f_{t-s}f_{t-s'})$. In this case, using (30) we have

$$\begin{aligned} T^{-1}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{f} &= T^{-1}\sum_{t=1}^{T}f_tg_{t-1}-\bar{f}\bar{g}_{-1}\to_p\sum_{s=0}^{\infty}\rho^s\gamma_f(s+1), \\ T^{-1}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{g}_{-1} &= T^{-1}\sum_{t=1}^{T}g_{t-1}^2-\bar{g}_{-1}^2\to_p\sum_{s=0}^{\infty}\sum_{s'=0}^{\infty}\rho^{s+s'}\gamma_f(|s-s'|)>0. \end{aligned}$$

Therefore, when the parallel trends assumption does not hold ($\bar{\sigma}_\gamma^2>0$) and f_t is covariance stationary, the asymptotic bias of $\hat{\rho}_{TWFE}$ for *both* n and T large is given

$$p\lim_{n,T\to\infty}\left(\hat{\rho}_{TWFE}-\rho_0\right)=\frac{\left(\frac{\sigma_\gamma^2}{\omega^2}\right)\sum_{s=0}^{\infty}\rho_0^s\gamma_f(s+1)}{1+\left(\frac{\sigma_\gamma^2}{\omega^2}\right)\sum_{s=0}^{\infty}\sum_{s'=0}^{\infty}\rho_0^{s+s'}\gamma_f(|s-s'|)}, \tag{35}$$

where

$$\omega^2 = \lim_{n,T\to\infty} T^{-1}n^{-1}\sum_{i=1}^{n} E\left[(\mathbf{v}_{i,-1} - \bar{\mathbf{v}}_{-1})' \mathbf{M}_T (\mathbf{v}_{i,-1} - \bar{\mathbf{v}}_{-1})\right] > 0.$$

The sign of this bias depends on the sign of $\sum_{s=0}^{\infty} \rho_0^s \gamma_f(s+1)$, which is likely to be positive since $\rho_0 > 0$, and one would expect f_t to be positively autocorrelated. The magnitude of the bias depends on the relative importance of the heterogeneity of γ_i, given by $\kappa^2 = \sigma_\gamma^2/\omega^2$, and the degree of persistence of the time effects, given by $\sum_{s=0}^{\infty} \rho_0^s \gamma_f(s+1)$. The TWFE estimator is asymptotically unbiased (as n and $T \to \infty$) only if f_t is serially independent. Everaert and De Groote (2016) derive expressions for the inconsistency of TWFE estimators under (25) assuming that f_t follows a stationary AR(1) process.

4.2 The bias of the TWFE estimator under non parallel trends with a trended latent factor

Consider now the case where f_t is trended, and suppose that $f_t = \mu t + s_t$, which is decomposed into a linear trend, μt, and a stochastic component s_t. The latter could be stationary or a unit root process. In either case s_t will be dominated by the trend component, so long as the drift term $\mu \neq 0$. Then it is easily seen that

$$g_{t-1} - \bar{g}_{-1} = \frac{\mu}{1-\rho_0}\left(t - \frac{T+1}{2}\right) + (q_{t-1} - \bar{q}_{-1}),$$

where $q_t = \rho_0 q_{t-1} + s_t$. Since $|\rho_0| < 1$, then q_t will have the same order of integration as s_t, namely q_t could be stationary or a unit root process. In either case the linear trend in $g_{t-1} - \bar{g}_{-1}$ will dominate $(q_{t-1} - \bar{q}_{-1})$ and we have

$$T^{-3}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{f} = \frac{\mu^2}{12(1-\rho_0)} + O_p(T^{-1}),$$
$$T^{-3}\mathbf{g}_{-1}'\mathbf{M}_T\mathbf{g}_{-1} = \frac{\mu^2}{12(1-\rho_0)^2} + O_p(T^{-1}).$$

Hence, scaling down the terms in (34) by T^{-3}, and using the above results it follows that $p\lim_{n,T\to\infty}(\hat{\rho}_{TWFE} - \rho_0) = 1 - \rho_0$, or

$$p\lim_{n,T\to\infty}(\hat{\rho}_{TWFE}) = 1. \tag{36}$$

This is a striking result that follows from the deterministic trend component of f_t, and holds irrespective of whether the stochastic component, s_t, is stationary or a unit root process.

In short, the TWFE estimator of ρ_0 is biased upward under nonparallel trends and the bias becomes highly pronounced when f_t is trended. In the latter case even a small degree of heterogeneity in γ_i can lead to substantial upward bias in estimation of ρ_0. In the context of growth regressions, the random walk model with drift is likely to be the most relevant case, and largely explains the very small estimates obtained in the literature for the speed of convergence, $\phi_0 = 1 - \rho_0$, when TWFE estimators are used.

4.3 Panel data models with group-specific time effects

The restrictive nature of the parallel trend assumption is becoming increasing recognized in the empirical growth literature. For example in addition to using TWFE, Acemoglu et al. (2019) let the time effect vary over groups. For instance, in column 8 of Table 4, they interact the time effects with the product of 7 regional dummies and the initial regime (democratic or non democratic). Having 14 groups in each year leaves very small numbers in some groups.

In general, the panel data model with group-specific interactive effects can be written as

$$y_{it} = \alpha_i + \gamma_g f_t \left(\sum_{g=1}^{G} d_{ig} \right) + \rho_0 y_{i,t-1} + u_{it},$$

where $d_{ig} = 1$ if country i belongs to group (region) $g = 1, 2, ..., G$, and *zero* otherwise. The number of groups, G, and the group membership is typically assumed known.[11] As before, f_t is the common latent factor. This model can be written more compactly as

$$y_{ig,t} = \alpha_{i,g} + \gamma_g f_t + \rho_0 y_{i,g,t-1} + u_{i,g,t} \tag{37}$$

for country (i, g), that denotes country i in group g, with $i = 1, 2, ..., n_g$, $g = 1, 2, ..., G$, and $n = \sum_{g=1}^{G} n_g$. The fixed effects, $\alpha_{i,g}$, and the group-specific time effects can be eliminated using the group-specific filtering, with ρ_0 estimated by

$$\hat{\rho}_{FE-GTE} = \frac{\sum_{t=1}^{T} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \tilde{y}_{ig,t-1} \tilde{y}_{ig,t}}{\sum_{t=1}^{T} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \tilde{y}_{ig,t-1}^2} \tag{38}$$

where $\tilde{y}_{ig,t} = (y_{ig,t} - \bar{y}_{ig,\circ}) - (\bar{y}_{\circ,gt} - \bar{y}_{\circ g \circ})$ and $\tilde{y}_{ig,t-1} = (y_{ig,t-1} - \bar{y}_{ig,-1}) - (\bar{y}_{\circ,gt} - \bar{y}_{\circ g,-1})$, where $\bar{y}_{ig,-s} = T^{-1} \sum_{t=1}^{T} y_{ig,t-s}$, $\bar{y}_{\circ,gt} = n_g^{-1} \sum_{i=1}^{n_g} y_{ig,t}$, and $\bar{y}_{\circ g,-s} = n_g^{-1} \sum_{i=1}^{n_g} y_{ig,t-s}$ for $s = 0$ and 1.

As an estimator of ρ_0, $\hat{\rho}_{FE-GTE}$ is subject to the Nickell bias as well as the group-size bias. For a fixed G, $\hat{\rho}_{FE-GTE}$ is a consistent estimator of ρ_0 if $(T, n_1, n_2, ..., n_G) \to \infty$. Under this condition group-specific time effects, γ_g, can be consistently estimated and hence eliminated form the analysis. The key condition on n_g, namely $n_g \to \infty$, for $g = 1, 2, ..., G$ is likely to be restrictive in the case of cross country analysis and arises due to the dynamic nature of the panel data model and will not be required in the case of static panel data models with strictly exogenous regressors.

[11] For large n and T panels, machine learning techniques, such as cluster analysis, can be used to estimate the number of groups and the group membership. Such grouping is particularly relevant if the aim is to identify convergence clubs that are based on geographical proximity and political and historical similarities.

5 The DCCEP estimator: nonparallel trends with homogeneous dynamics

The use of group-specific time effects can be avoided by following the common correlated effects (CCE) approach introduced in Pesaran (2006) and later implemented for dynamic panels by Chudik and Pesaran (2015a), referred to as dynamic CCE (DCCE). The basic idea behind DCCE is to proxy the latent factor, f_t, by current and lagged cross country averages of log output, namely $\bar{y}_{\circ t}$ and $\bar{y}_{\circ,t-1}$. The case with a homogeneous, ρ is referred to as DCCEP. Averaging (25) over i, we have $\bar{y}_{\circ t} = \bar{\alpha} + \bar{\gamma} f_t + \rho\bar{y}_{\circ,t-1} + \bar{u}_{\circ t}$, and assuming $\bar{\gamma} \neq 0$, then $f_t = \bar{\gamma}^{-1}\left(\bar{y}_{\circ t} - \rho\bar{y}_{\circ,t-1} - \bar{\alpha} - \bar{u}_{\circ t}\right)$. Using this result back in (25) now yields

$$y_{it} = (\alpha_i - \delta_{i0}\bar{\alpha}) + \delta_{i0}\bar{y}_{\circ t} + \delta_{i1}\bar{y}_{\circ,t-1} + \rho y_{i,t-1} + u_{it} - \delta_{i0}\bar{u}_{\circ t}, \tag{39}$$

where $\delta_{i0} = \bar{\gamma}^{-1}\gamma_i$, $\delta_{i1} = -\delta_{i0}\rho$, and $\bar{u}_{\circ t} = n^{-1}\sum_{i=1}^{n} u_{it}$. Running the above panel AR(1) regression in y_{it}, augmented with the cross section averages, $\bar{y}_{\circ t}$ and $\bar{y}_{\circ,t-1}$, we obtain the following pooled DCCE estimator

$$\hat{\rho}_{DCCEP} = \frac{n^{-1}T^{-1}\sum_{i=1}^{n}\mathbf{y}_{i,-1}'\mathbf{M}_{\bar{h}}\mathbf{y}_{i\circ}}{n^{-1}T^{-1}\sum_{i=1}^{n}\mathbf{y}_{i,-1}'\mathbf{M}_{\bar{h}}\mathbf{y}_{i,-1}}, \tag{40}$$

where $\mathbf{M}_{\bar{h}} = \mathbf{I}_T - \bar{\mathbf{H}}\left(\bar{\mathbf{H}}'\bar{\mathbf{H}}\right)^{-}\bar{\mathbf{H}}'$, $\bar{\mathbf{H}} = \left(\tau_T, \bar{\mathbf{y}}, \bar{\mathbf{y}}_{-1}\right)$, and $\mathbf{A}^{-}$ denotes the generalized inverse of the square matrix $\mathbf{A}$. Note that $\mathbf{M}_{\bar{h}}\mathbf{y}_{i\circ}$ and $\mathbf{M}_{\bar{h}}\mathbf{y}_{i,-1}$ are residuals from the least squares regressions of $\mathbf{y}_{i\circ}$ and $\mathbf{y}_{i,-1}$ on $\bar{\mathbf{H}}$ and are invariant to the choice of the generalized inverse. The use of generalized inverse allows for possible perfect collinearity that arise between $\bar{y}_{\circ t}$ and $\bar{y}_{\circ,t-1}$ when f_t is trended, as discussed below.

To establish the consistency of $\hat{\rho}_{DCCEP}$ for ρ_0, writing (39) in vector notations gives

$$\mathbf{y}_{i\circ} = \bar{\mathbf{H}}\boldsymbol{\pi}_i + \rho_0\mathbf{y}_{i,-1} + \mathbf{u}_{i\circ} - \delta_{i0}\bar{\mathbf{u}},$$

where $\boldsymbol{\pi}_i = (\alpha_i - \delta_{i0}\bar{\alpha}, \delta_{i0}, \delta_{i1})'$. Then

$$\hat{\rho}_{DCCEP} - \rho_0 = \frac{n^{-1}\sum_{i=1}^{n}\mathbf{y}_{i,-1}'\mathbf{M}_{\bar{h}}\left(\mathbf{u}_{i\circ} - \delta_{i0}\bar{\mathbf{u}}\right)}{n^{-1}\sum_{i=1}^{n}\mathbf{y}_{i,-1}'\mathbf{M}_{\bar{h}}\mathbf{y}_{i,-1}}.$$

Also, from (29) we obtain $\mathbf{y}_{i,-1} = \alpha_i/(1-\rho)\tau_T + \gamma_i\mathbf{g}_{-1} + \boldsymbol{v}_{i,-1}$, and by cross section averaging, $\bar{\mathbf{y}}_{-1} = \bar{\alpha}/(1-\rho)\tau_T + \bar{\gamma}\mathbf{g}_{-1} + \bar{\mathbf{v}}_{-1}$, which in turn yields

$$\mathbf{y}_{i,-1} = \left(\frac{\alpha_i - \delta_{i0}\bar{\alpha}}{1-\rho_0}\right)\tau_T + \delta_{i0}\bar{\mathbf{y}}_{-1} + \boldsymbol{v}_{i,-1} - \delta_{i0}\bar{\mathbf{v}}_{-1}.$$

Recall that $\boldsymbol{v}_{i,-1} = (v_{i0}, v_{i1}, ..., v_{i,T-1})$, $\bar{\mathbf{v}}_{-1} = n^{-1}\sum_{i=1}^{n}\boldsymbol{v}_{i,-1}$, where $v_{it} = \sum_{j=0}^{\infty}\rho^j u_{i,t-j}$.

Hence

$$\hat{\rho}_{DCCEP} - \rho_0 = \frac{\sum_{i=1}^{n}\left(\boldsymbol{v}_{i,-1} - \delta_{i0}\bar{\mathbf{v}}_{-1}\right)' \mathbf{M}_{\bar{h}}\left(\mathbf{u}_{i\circ} - \delta_{i0}\bar{\mathbf{u}}\right)}{\sum_{i=1}^{n}\left(\boldsymbol{v}_{i,-1} - \delta_{i0}\bar{\mathbf{v}}_{-1}\right)' \mathbf{M}_{\bar{h}}\left(\boldsymbol{v}_{i,-1} - \delta_{i0}\bar{\mathbf{v}}_{-1}\right)}.$$

Let $\mathbf{M}_g = \mathbf{I}_T - \mathbf{P}_g$, $\mathbf{P}_g = \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-}\mathbf{G}'$, and $\mathbf{G} = \begin{pmatrix} \tau_T & \mathbf{g} & \mathbf{g}_{-1} \end{pmatrix}$. Also

$$\bar{\mathbf{H}} = \left(\tau_T, \bar{\mathbf{y}}, \bar{\mathbf{y}}_{-1}\right) = \left(\tau_T,\ \bar{\alpha}/(1-\rho)\tau_T + \bar{\gamma}\mathbf{g} + \bar{\mathbf{v}},\ \bar{\alpha}/(1-\rho)\tau_T + \bar{\gamma}\mathbf{g}_{-1} + \bar{\mathbf{v}}_{-1}\right).$$

Then $\bar{\mathbf{H}} = \mathbf{G}\bar{\mathbf{Q}} + \bar{\mathbf{V}}$, where $\bar{\mathbf{Q}}$ is a non-singular matrix (by assumption $\bar{\gamma} \neq 0$) given by

$$\bar{\mathbf{Q}} = \begin{pmatrix} 1 & \bar{\alpha}/(1-\rho_0) & \bar{\alpha}/(1-\rho_0) \\ 0 & \bar{\gamma} & 0 \\ 0 & 0 & \bar{\gamma} \end{pmatrix},$$

and $\bar{\mathbf{V}} = \left(\mathbf{0}, \bar{\mathbf{v}}, \bar{\mathbf{v}}_{-1}\right)$. Since u_{it} (and hence v_{it}) are cross-sectionally independent, then $\bar{\mathbf{u}} = O_p(n^{-1/2})$, $\bar{\mathbf{v}} = O_p(n^{-1/2})$, and $\bar{\mathbf{v}}_{-1} = O_p(n^{-1/2})$. It is then easily established that for any fixed $T > 3$, and as $n \to \infty$, we have

$$\mathbf{M}_{\bar{h}} - \mathbf{M}_g = \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-}\mathbf{G}' - \left(\mathbf{G}\bar{\mathbf{Q}} + \bar{\mathbf{V}}\right)\left[\left(\mathbf{G}\bar{\mathbf{Q}} + \bar{\mathbf{V}}\right)'\left(\mathbf{G}\bar{\mathbf{Q}} + \bar{\mathbf{V}}\right)\right]^{-}\left(\mathbf{G}\bar{\mathbf{Q}} + \bar{\mathbf{V}}\right)' = O_p(n^{-1/2})$$

This follows since $\bar{\mathbf{V}} = O_p\left(n^{-1/2}\right)$ and $\mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-}\mathbf{G}' = \mathbf{G}\bar{\mathbf{Q}}\left[\left(\mathbf{G}\bar{\mathbf{Q}}\right)'\left(\mathbf{G}\bar{\mathbf{Q}}\right)\right]^{-}\bar{\mathbf{Q}}'\mathbf{G}'$. Using this result, for a fixed T and conditional on $\mathbf{f}$, or $\mathbf{G}$, we now have

$$\hat{\rho}_{DCCEP} - \rho_0 \to_p \frac{lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right)}{lim_{n\to\infty}\ n^{-1}\sum_{i=1}^{n} E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\boldsymbol{v}_{i,-1}\left|\mathbf{f}\right.\right)}. \quad (41)$$

For the denominator of (41) we have $E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\boldsymbol{v}_{i,-1}\left|\mathbf{f}\right.\right) = Tr\left[\mathbf{M}_g E\left(\boldsymbol{v}_{i,-1}\boldsymbol{v}_{i,-1}'\right)\right]$, and

$$E\left(\boldsymbol{v}_{i,-1}\boldsymbol{v}_{i,-1}'\right) = \frac{\sigma_i^2}{1-\rho_0^2}\begin{pmatrix} 1 & \rho_0 & \cdots & \rho_0^{T-1} \\ \rho_0 & 1 & \cdots & \rho_0^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_0^{T-1} & \rho_0^{T-2} & \cdots & 1 \end{pmatrix} = \sigma_i^2\mathbf{V}_0,$$

where $\lambda_{\max}\left(\mathbf{V}_0\right)$ is bounded noting that $|\rho_0| < 1$. Hence, $lim_{n\to\infty}\ n^{-1}\sum_{i=1}^{n} E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\boldsymbol{v}_{i,-1}\left|\mathbf{f}\right.\right) = \bar{\sigma}^2 Tr\left(T^{-1}\mathbf{M}_g\mathbf{V}_0\right)$, where $\bar{\sigma}^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_i^2$. Also, $Tr\left(T^{-1}\mathbf{M}_g\mathbf{V}_0\right) = Tr\left(T^{-1}\mathbf{M}_g\mathbf{V}_0\mathbf{M}_g\right) \leq T^{-1}\lambda_{\max}\left(\mathbf{V}_0\right) Tr\left(\mathbf{M}_g\right) = \left[(T-3)/T\right]\lambda_{\max}\left(\mathbf{V}_0\right)$ which is bounded in T, irrespective of whether f_t is trended or not.

Now consider the numerator of (41) and note that

$$E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right) = T^{-1}\sum_{t=1}^{T} E\left(v_{i,t-1}u_{it}\right) - T^{-1}E\left(\boldsymbol{v}_{i,-1}'\mathbf{P}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right). \quad (42)$$

Since $v_{it} = \sum_{s=0}^{\infty}\rho^s u_{i,t-s}$ and u_{it} is serially uncorrelated, then $E\left(v_{i,t-1}u_{it}\right) = 0$, and $E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right) = -T^{-1}E\left(\boldsymbol{v}_{i,-1}'\mathbf{P}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right)$. Further, since f_t (and g_t) is independently distributed of u_{it} (and

v_{it}) then conditional on $\mathbf{f}$, we have $T^{-1}E\left(\boldsymbol{v}_{i,-1}'\mathbf{P}_g\mathbf{u}_{i\circ}\right) = Tr\left[\mathbf{P}_g E\left(T^{-1}\mathbf{u}_{i\circ}\boldsymbol{v}_{i,-1}'\right)\right]$. It is also easily seen that $E\left(\mathbf{u}_{i\circ}\boldsymbol{v}_{i,-1}'\right) = \sigma_i^2\mathbf{W}_\rho$, where

$$\mathbf{W}_\rho = \begin{pmatrix} 0 & 1 & \rho_0 & \rho_0^2 & \cdots & \rho_0^{T-2} \\ 0 & 0 & 1 & \rho_0 & \cdots & \rho_0^{T-3} \\ \vdots & \vdots & \vdots & & \cdots & \vdots \\ 0 & 0 & \ddots & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix} = (w_{tt'}).$$

Then

$$\begin{aligned} E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right) &= -T^{-1}E\left(\boldsymbol{v}_{i,-1}'\mathbf{P}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right) \\ &= T^{-2}\sigma_i^2 Tr\left(\mathbf{A}\mathbf{W}_\rho\right) = T^{-2}\sigma_i^2\sum_{t=1}^{T}\sum_{t=1}^{T} a_{tt'}w_{tt'} \end{aligned}$$

where $a_{tt'}$ is the (t,t') element of $\mathbf{A}$ $=T\mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'$. Hence

$$\begin{aligned} \left|E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right)\right| &\leq T^{-2}\sigma_i^2\sup_{t,t'}\left|a_{tt'}\right|\sum_{t=1}^{T}\sum_{t=1}^{T}\left|w_{tt'}\right| \\ &\leq T^{-1}\sup_{t,t'}\left|a_{tt'}\right|\sigma_i^2\sum_{s=1}^{T-1}\left(1-\frac{s}{T}\right)\left|\rho\right|^{s-1}, \end{aligned}$$

and

$$lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}E\left(T^{-1}\boldsymbol{v}_{i,-1}'\mathbf{M}_g\mathbf{u}_{i\circ}\left|\mathbf{f}\right.\right) \leq T^{-1}\sup_{t,t'}\left|a_{tt'}\right|\bar{\sigma}^2\sum_{s=1}^{T-1}\left(1-\frac{s}{T}\right)\left|\rho\right|^{s-1}.$$

Since $\bar{\sigma}^2\sum_{s=1}^{T-1}\left(1-\frac{s}{T}\right)\left|\rho_0\right|^{s-1}$ is bounded in T, then conditional on $\mathbf{G}$, the order of the numerator of (41) is determined by the order of $T^{-1}\sup_{t,t'}\left|a_{tt'}\right|$. It is interesting that the probability order of this term does depend on whether f_t is stationary or trended, which is in contrast to the order of the bias of the TWFE estimator that crucially depended on whether f_t is stationary or trended.

When f_t is stationary then $g_t = O_p(1)$ and $T^{-1}\mathbf{G}'\mathbf{G}$ $=O_p(1)$ and since $\mathbf{A} = \mathbf{G}\left(T^{-1}\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}$, then $\sup_{t,t'}\left|a_{tt'}\right| = O_p(1)$. In the case where f_t is trended we first write $\mathbf{A}$ as

$$\mathbf{A} = \left(\sqrt{T}\mathbf{G}\mathbf{D}_T\right)\left(\mathbf{D}_T\mathbf{G}'\mathbf{G}\mathbf{D}_T\right)^{-1}\left(\sqrt{T}\mathbf{D}_T\mathbf{G}'\right),$$

where $\mathbf{D}_T = diag(T^{-1/2}, T^{-3/2}, T^{-3/2})$. Then it is easily established that $\mathbf{D}_T\mathbf{G}'\mathbf{G}\mathbf{D}_T = O_p(1)$ and $\sqrt{T}\mathbf{G}\mathbf{D}_T = O_p(1)$. Hence, irrespective of whether f_t is stationary or trended, $\sup_{t,t'}\left|a_{tt'}\right| = O_p(1)$ and numerator of (41) is $O_p(T^{-1})$, and overall we have

$$\hat{\rho}_{DCCEP} - \rho_0 \to_p O_p\left(T^{-1}\right), \text{ as } n \to \infty. \tag{43}$$

Namely, the DCCEP estimator of ρ_0 is consistent when n and T are both large, and applies irrespective of whether f_t is stationary or trended. The trend in f_t could be stochastic and subject to breaks. However, not surprisingly, DCCEP suffers from the small T bias analogous to the Nickell bias of the TWFE estimator. De Vos and Everaert (2021) derive the fixed T bias under the assumption of a stationary factor. This small T bias will be present irrespective of whether the panel data model includes a latent factor or not. It is simply due to dynamic nature of the panel regression under consideration.

To avoid the small T bias of DCCEP estimator, Hayakawa, Pesaran and Smith (2023) allow for interactive time effects through a multi-factor error structure $\boldsymbol{\gamma}_i'\mathbf{f}$ in dynamic fixed effects panel data models in addition to the standard fixed and time effects, but assume that $\boldsymbol{\gamma}_i$ are cross-sectionally independent. They apply maximum likelihood estimation after first-differencing the data to remove the fixed effects. This transformed quasi maximum likelihood estimator is shown to be robust to the heterogeneity of the initial values and common unobserved effects, and is applicable to both stationary and unit root cases. They also propose a procedure for selection of the number of latent factors. They apply the estimator to the Acemoglu et al. (2019) growth data using averages taken over five yearly intervals.

6 The DCCEMG estimator: nonparallel trends with heterogeneous dynamics

In this case the output equations with heterogeneous dynamics as well as nonparallel trends are given by

$$y_{it} = \alpha_i + \gamma_i f_t + \rho_i y_{i,t-1} + u_{it}, \text{ for } i = 1, 2, ..., n. \tag{44}$$

This model is investigated in detail by Chudik and Pesaran (2015a), who show that once we allow for ρ_i to differ across i, the identification and estimation of f_t become much more complicated due to the heterogeneity in the dependence of y_{it} on f_t, and its lagged values. To identify f_t it is assumed that ρ_i lies in the range $(-1, 1)$, such that $E\left|\rho_i\right|^s \leq \bar{\rho}^s$, where $\bar{\rho} < 1$. This ensures that $E\left(\frac{1}{1-\rho_i^2}\right)$ exists which is required for identification of f_t from the macro outcomes, $\bar{y}_{\circ t}$. This assumption rules out the possibility of long memory processes.[12]

6.1 DCCEMG with stationary latent factors

Chudik and Pesaran (2015a) further assume that f_t is covariance stationary which is rather restrictive as noted above. Here we provide a sketch of the proof to highlight the

[12] On this see Robinson (1978) and Granger (1980). A more general treatment is provided in Pesaran and Chudik (2014).

way nonparallel trends and dynamic heterogeneity interact. Solving (44) from a distant past we have

$$y_{it} = \mu_i + \gamma_i \sum_{s=0}^{\infty} \rho_i^s f_{t-s} + \sum_{s=0}^{\infty} \rho_i^s u_{i,t-s},$$

with $\mu_i = \alpha_i/(1-\rho_i)$, $\theta_{is} = \gamma_i \rho_i^s$ and $v_{it} = \sum_{s=0}^{\infty} \rho_i^s u_{i,t-s}$, which if averaged over i yields

$$\bar{y}_{\circ t} = \bar{\mu}_n + \sum_{s=0}^{\infty} \bar{\theta}_{ns} f_{t-s} + \bar{v}_{\circ t}, \tag{45}$$

where $\bar{\mu}_n = n^{-1}\sum_{i=1}^{n} \mu_i$, $\bar{\theta}_{ns} = n^{-1}\sum_{i=1}^{n} \theta_{is}$, and $\bar{v}_{\circ t} = n^{-1}\sum_{i=1}^{n} v_{it}$.It is clear that conditional on ρ_i, $v_{it}{}'s$ continue to be cross-sectionally independent, and $Var\left(\bar{v}_{\circ t}\right) = n^{-2}\sum_{i=1}^{n} Var\left(v_{it}\right)$. Also

$$Var\left(v_{it}\right) = Var\left[E\left(v_{it}\left|\rho_i\right.\right)\right] + E\left[Var\left(v_{it}\left|\rho_i\right.\right)\right] = E\left(\frac{\sigma_i^2}{1-\rho_i^2}\right),$$

and it is easily seen that $E\left(\frac{\sigma_i^2}{1-\rho_i^2}\right) < C$. Hence, $Var\left(\bar{v}_{\circ t}\right) = O(n^{-1})$. Using these results in (45) we have

$$\bar{\theta}_n(L) f_t = \bar{y}_{\circ t} - \bar{\mu}_n + O_p(n^{-1/2}). \tag{46}$$

where $\bar{\theta}_n(L) = \sum_{s=0}^{\infty} \bar{\theta}_{ns} L^s$, and L is a lag operator. In the case of homogeneous dynamics, $\bar{\theta}_n(L) = \bar{\gamma}_n(1-\rho L)^{-1}$ and f_t can be identified as an affine function of $\bar{y}_{\circ t}$ and $\bar{y}_{\circ,t-1}$. When ρ_i is heterogeneous such a simple inversion is not possible and additional restrictions are required. Chudik and Pesaran (2015a), building on an earlier paper by Pesaran and Chudik (2014), show that when f_t is covariance stationary then f_t can be approximated by a distributed lag function of $y_{\circ t}, y_{\circ,t-1}, ..., y_{\circ .t-p_T}$, where the lag order, p_T, rises with T but at the slower rate of $T^{1/3}$. Specifically, they show that

$$y_{it} = \alpha_{in} - \gamma_i \psi(1)\bar{\mu}_n + \gamma_i \sum_{s=0}^{p_T} \psi_s \bar{y}_{\circ,t-s} + \rho_i y_{i,t-1} + \zeta_{it,n}, \text{ for } i = 1, 2, ..., n. \tag{47}$$

where

$$\tilde{\alpha}_{in} = \alpha_i - \gamma_i \psi(1)\bar{\mu}_n, \text{ and } \zeta_{it,n} = u_{it} + \gamma_i \left(\sum_{s=q_T+1}^{\infty} \psi_s \bar{y}_{\circ,t-s}\right) + O_p(n^{-1/2}).$$

6.2 DCCEMG with trended latent factors

The above approach can be readily extended to panel data models with trended latent factors and higher order dynamics:

$$\Delta y_{it} = \alpha_i + \gamma_i f_t - \phi_i y_{i,t-1} + \sum_{s=1}^{p-1} \delta_{is} \Delta y_{i,t-s} + u_{it}. \tag{48}$$

Despite the higher order dynamics, f_t can still be well approximated by $\bar{y}_{\circ,t-s}$, for $s = 0, 1, ..., q_T$, with ϕ_i estimated as

$$\hat{\phi}_{i,DCCE} = \left(\mathbf{y}'_{i,-1}\bar{\mathbf{M}}_i\mathbf{y}_{i,-1}\right)^{-1}\mathbf{y}'_{i,-1}\bar{\mathbf{M}}_i\Delta\mathbf{y}_i, \quad (49)$$

where $\bar{\mathbf{M}}_i = \mathbf{I}_T - \bar{\mathbf{W}}_i\left(\bar{\mathbf{W}}'_i\bar{\mathbf{W}}_i\right)^{-}\bar{\mathbf{W}}'_i$, $\bar{\mathbf{W}}_i = (\tau_T, \bar{\mathbf{y}}_\circ, \bar{\mathbf{y}}_{\circ,-1}, ..., \bar{\mathbf{y}}_{\circ,-q_T}, \Delta\mathbf{y}_{i,-1}, ..., \Delta\mathbf{y}_{i,-p+1})$, and $\left(\bar{\mathbf{W}}'_i\bar{\mathbf{W}}_i\right)^{-}$ denotes the generalized inverse of $\left(\bar{\mathbf{W}}'_i\bar{\mathbf{W}}_i\right)$. Note that the estimates, $\hat{\phi}_{i,DCCE}$, are invariant to the choice of the generalized inverse. The use of generalized inverse is particularly important if f_t is trended, since in that case $\bar{\mathbf{y}}_\circ, \bar{\mathbf{y}}_{\circ,-1}, ..., \bar{\mathbf{y}}_{\circ,-q_T}$ become highly multicollinear with their pairwise correlations tending to unity as $n \to \infty$. This result is also relevant to the choice of q_T, suggesting that small values of q_T should be sufficient for cross section averages to be a good proxy for f_t when y_{it} are trended

A consistent estimator of $E(\phi_i)$ is now given by

$$\hat{\phi}_{DCCEMG} = n^{-1}\sum_{i=1}^{n}\hat{\phi}_{i,DCCE}, \quad (50)$$

and

$$Var\left(\widehat{\hat{\phi}_{DCCEMG}}\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{\phi}_{i,DCCE} - \hat{\phi}_{DCCEMG}\right)^2. \quad (51)$$

7 Allowing for time-varying covariates

An important issue addressed in the literature has been how to allow both for observed factors (covariates) that affect log per-capita output, which vary over time, like capital and demography, and for covariates that are time-invariant like geography, or vary over time very slowly like climate or institutional factors. This section considers time-varying covariates, which raise no new methodological issues, and the next section considers slowly moving covariates.

To simplify the exposition we abstract from higher order dynamics and augment the panel data model (44) with the $k_x \times 1$ vector of time-varying covariates $\mathbf{x}_{it}$

$$y_{it} = \alpha_i + \boldsymbol{\gamma}'_i\boldsymbol{f}_t + \rho_i y_{i,t-1} + \boldsymbol{\beta}'_i\mathbf{x}_{it} + u_{it}, \quad (52)$$

where $\boldsymbol{\beta}_i$ is the $k_x \times 1$ vector of fixed coefficients that could vary over i. In this section we also allow for a finite number of multiple latent factors, which we denote by the $m \times 1$ vector $\boldsymbol{f}_t$, with the associated loading vector, $\boldsymbol{\gamma}_i$. For DCCEP and DCEEMG to be applicable to this more general set up it is required that $m \leq k_x + 1$. We also consider a random coefficient model and assume $\boldsymbol{\psi}_i = (\rho_i, \boldsymbol{\beta}'_i)' = \boldsymbol{\psi} + \boldsymbol{\eta}_i$, where $\boldsymbol{\eta}_i$ are distributed independently over i with mean zero and a finite variance, ensuring that $|\rho_i| < 1$ for all i. The focus is on estimation of the mean effects, $E(\boldsymbol{\psi}_i)$, which we denote by $\boldsymbol{\psi}_0$. Following Chudik and Pesaran (2015a), $\boldsymbol{\psi}_0$ can be estimated consistently by augmenting the panel regression with the cross-sectional averages $(\bar{y}_{\circ t}, \bar{y}_{\circ,t-1}, ..., \bar{y}_{\circ,t-p_T}, \bar{\mathbf{x}}_{\circ t}, \bar{\mathbf{x}}_{\circ,t-1}, ..., \bar{\mathbf{x}}_{\circ,t-p_T})$ where $\bar{\mathbf{x}}_{\circ t} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_{it}$, as proxies for the latent factors. As discussed already, the required number of lagged values, p_T, depends on whether ρ_i is homogenous or not. In the homogeneous case (with $\boldsymbol{\psi}_i = \boldsymbol{\psi}$) consistent estimation of $\boldsymbol{\psi}_0$ can be obtained by pooled regression of y_{it} on $(y_{i,t-1}, \mathbf{x}'_{it})'$ augmented with $(\bar{y}_{\circ t}, \bar{y}_{\circ,t-1}, \bar{\mathbf{x}}_{\circ t})$. This estimator is known as pooled dynamic CCE estimator and denoted by DCCEP. Specifically (See also (40))

$$\begin{pmatrix} \hat{\rho}_{DCCEP} \\ \hat{\boldsymbol{\beta}}_{DCCEP} \end{pmatrix} = \hat{\boldsymbol{\psi}}_{DCCEP} = \left(\sum_{i=1}^{n}\mathbf{Q}'_i\mathbf{M}_{\bar{h}}\mathbf{Q}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{Q}'_i\mathbf{M}_{\bar{h}}\mathbf{y}_{i\circ}, \tag{53}$$

where $\mathbf{Q}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$, $\mathbf{X}_i$ is the $T \times k_x$ matrix of observations on $\mathbf{x}_{it}$, $\mathbf{M}_{\bar{h}} = \mathbf{I}_T - \bar{\mathbf{H}}\left(\bar{\mathbf{H}}'\bar{\mathbf{H}}\right)^{-}\bar{\mathbf{H}}'$, $\bar{\mathbf{H}} = \left(\tau_T, \bar{\mathbf{y}}_{\circ}, \bar{\mathbf{y}}_{\circ,-1}, \bar{\mathbf{X}}\right)$, and $\bar{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$. Assuming $\mathbf{x}_{it}$ are weakly exogenous, following Chudik and Pesaran (2015a), it is easily established that $\hat{\boldsymbol{\psi}}_{DCCEP} \to_p \boldsymbol{\psi}_0 = (\rho_0, \boldsymbol{\beta}'_0)'$ for sufficiently large n and T. Large T is required to remove the Nickel type bias which continues to apply. The small T bias of the $DCCEP$ estimator can be reduced using half-Jackknife procedure proposed in Chudik et al. (2018).

In the fully heterogeneous case where $\boldsymbol{\psi}_i$ are allowed to vary across i, the mean effects, $\boldsymbol{\psi}_0$, can be estimated using the mean group approach discussed above but with the important difference that the country-specific regressions must now be augmented with the cross section averages of $\mathbf{x}_{it}$ and y_{it} as well as p_T lagged values of y_{it}. Namely,

$$\hat{\boldsymbol{\psi}}_{DCCEMG} = n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\psi}}_{i,DCCE},\ \hat{\boldsymbol{\psi}}_{i,DCCE} = \left(\mathbf{Q}'_i\tilde{\mathbf{M}}_{\bar{h}}\mathbf{Q}_i\right)^{-1}\mathbf{Q}'_i\tilde{\mathbf{M}}_{\bar{h}}\mathbf{y}_{i\circ}, \tag{54}$$

where $\tilde{\mathbf{M}}_{\bar{h}} = \mathbf{I}_T - \tilde{\mathbf{H}}\left(\tilde{\mathbf{H}}'\tilde{\mathbf{H}}\right)^{-}\tilde{\mathbf{H}}'$, and $\tilde{\mathbf{H}} = (\tau_T, \bar{\mathbf{y}}, \bar{\mathbf{y}}_{-1}, ..., \bar{\mathbf{y}}_{-p_T}, \bar{\mathbf{X}}, \bar{\mathbf{X}}_{-1}, ..., \bar{\mathbf{X}}_{-p_T})$.The asymptotic variance of $\hat{\boldsymbol{\psi}}_{DCCEMG}$ can be consistently estimated by

$$\widehat{Var\left(\hat{\boldsymbol{\psi}}_{DCCEMG}\right)} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\psi}}_{i,DCCE} - \hat{\boldsymbol{\psi}}_{DCCEMG}\right)\left(\hat{\boldsymbol{\psi}}_{i,DCCE} - \hat{\boldsymbol{\psi}}_{DCCEMG}\right)'. \tag{55}$$

The theoretical analysis suggests setting p_T to the integer part of $T^{1/3}$. However, as already discussed, in the case of trended series only few lags are required to ensure the factors $\boldsymbol{f}_t$ are adequately proxied by the cross country averages.

The DCCE approach can also accommodate models with a mix of homogeneous and heterogeneous coefficients. This arises in our empirical application where we consider the effects of democracy on output with no time variations in the democracy indicator for half of the countries in the panel.

8 Time-invariant or slowly moving covariates

One important advantage of DCCE estimators of $\boldsymbol{\psi}_0$ are their robustness to the processes that generate fixed effects, α_i, and the factor loadings, γ_i. However, when it comes to estimating the possible determinants of α_i and γ_i, this feature seems to be a disadvantage, since (like TWFE) the DCCE approach eliminates the time-invariant or slowly moving determinants and seemingly throws the baby out with the bathwater, so to speak. It is for this reason that many researchers do not favour FE estimation and prefer pooled least squares with α_i and γ_i replaced by their determinants. Barro (2015) argues against including fixed effects, saying "there is insufficient within country variation in the measured institutional quality to isolate a statistically significant effect on economic growth." Barro regression discussed in Section 3 imposes the parallel trends assumption and in effect assumes that α_i is a *deterministic* function of the time-invariant or slowly moving regressors. But, as already argued, pooling can lead to inconsistent estimators unless very strong assumption are made about determinants of α_i and γ_i.

Here we follow an alternative approach, whereby we first estimate the time-varying coefficients, $\boldsymbol{\psi}_i$, using DCCEP or DCCEMG type procedures that are invariant with respect to the values of α_i and γ_i, and then follow Pesaran and Zhou (2018) and estimate the effects of time-invariant regressors using a cross-country regression after filtering out the effects of time-varying regressors. For slowly moving variables, country-specific time averages can be used. This approach originates in the pioneering contributions of Hausman and Taylor (1981) in static panel data models with random coefficients. The present application is further complicated due to the dynamic nature of the output convergence process and heterogeneity, particularly when the parallel trends assumption is relaxed. Pesaran and Zhou (2018) focus on static short T panels and addresses the implications of estimation uncertainty of fixed effects estimation of coefficients of time-varying variables for inference on the time-invariant effects. Here we assume both n and T are large which considerably simplifies inference regarding the time-invariant effects.

Starting with (9) and (10), and denoting the vector of time-invariant variables by $\mathbf{z}_i$, after filtering out the dynamics and the effects of time-varying covariates, $\mathbf{x}_{it}$, we obtain

$$a_{iT} = \bar{y}_{i\circ} - \bar{\mathbf{q}}_{i\circ}'\boldsymbol{\psi}_i = \alpha_T + \boldsymbol{\theta}'\mathbf{z}_i + \eta_{iT} + \bar{u}_{i\circ}, \tag{56}$$

where $\bar{\mathbf{q}}_{i\circ} = (\bar{y}_{i,-1}, \bar{\mathbf{x}}_{i\circ}')'$, $\boldsymbol{\psi}_i = (\rho_i, \boldsymbol{\beta}_i')'$, $\mathbf{z}_i = \mathbf{z}_{i\alpha} \cup \mathbf{z}_{i\gamma}$, $\boldsymbol{\theta}$ is the $k_z \times 1$ vector of the time-invariant coefficients that can be identified, $\alpha_T = \alpha_\alpha + \alpha_\gamma \bar{f}_T$, and $\eta_{iT} = \eta_{i\alpha} + \eta_{i\gamma}\bar{f}_T$. Conditional on $\boldsymbol{\psi} = (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2', ..., \boldsymbol{\psi}_n')'$, the time-invariant effects are identified if $\mathbf{z}_i$ is distributed independently of the composite error term, $\xi_{iT} = \eta_{iT} + \bar{u}_{i\circ}$, or if there are instrumental variables that are uncorrelated with the composite errors, and are sufficiently correlated with $\mathbf{z}_i$. To simplify the exposition here we assume $\mathbf{z}_i$ is distributed independently of ξ_{iT}, and assume that $\mathbf{S}_{zz,n} = n^{-1}\sum_{i=1}^{n}(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}}_\circ)' \to_p \mathbf{S}_{zz}$, a non-stochastic positive definite matrix. Under these assumptions, and conditional on $\boldsymbol{\psi}$, $\boldsymbol{\theta}$ can be estimated consistently by the least squares regression of a_{iT} on an intercept and $\mathbf{z}_i$, namely

$$\hat{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right)=\left[n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)'\right]^{-1}n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)a_{iT}. \tag{57}$$

Denoting the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$, and using a_{iT} from (56) we have

$$\hat{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right)-\boldsymbol{\theta}_0=\left[n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)'\right]^{-1}n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)\left(\eta_{iT}+\bar{u}_{i\circ}\right),$$

and it also follows that $\sqrt{n}\left[\hat{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right)-\boldsymbol{\theta}_0\right]\rightarrow_d N(0,\mathbf{V}_\theta)$, where $\mathbf{V}_\theta=\mathbf{S}_{zz}^{-1}\boldsymbol{\Omega}_{zz}\mathbf{S}_{zz}^{-1}$, $\boldsymbol{\Omega}_{zz}=p\lim_n n^{-1}\sum_{i=1}^{n}\omega_{iT}^2\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)'$, and $\omega_{iT}^2=\sigma_\eta^2+T^{-1}\sigma_{iu}^2$.

In practice, the unknown coefficients $\boldsymbol{\psi}$ in (57) can be replaced by TWFE, DCCEP or DCCEMG estimators, depending on whether $\gamma_i=\gamma$, and/or $\boldsymbol{\psi}_i=\boldsymbol{\psi}$. Under slope homogeneity and parallel trends the TWFE estimator, $\hat{\boldsymbol{\psi}}_{TWFE}$, can be used to obtain the following feasible estimator $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{TWFE}=\left[n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)'\right]^{-1}n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)\left(\bar{y}_{i\circ}-\bar{\mathbf{q}}_{i\circ}'\hat{\boldsymbol{\psi}}_{TWFE}\right). \tag{58}$$

Substituting for $\bar{y}_{i\circ}$ using (56), we now have

$$\hat{\boldsymbol{\theta}}_{TWFE}-\boldsymbol{\theta}_0=\hat{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right)-\boldsymbol{\theta}_0-\mathbf{S}_{zz,n}^{-1}\left[n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)\bar{\mathbf{q}}_{i\circ}'\right]\left(\hat{\boldsymbol{\psi}}_{TWFE}-\boldsymbol{\psi}_0\right). \tag{59}$$

It is easily seen that $n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i-\bar{\mathbf{z}}_{\circ}\right)\bar{\mathbf{q}}_{i\circ}'$ tends to a fixed matrix, and since under parallel trends and homogeneous slopes assumption $\hat{\boldsymbol{\psi}}_{TWFE}-\boldsymbol{\psi}_0\rightarrow_p\mathbf{0}$, then it also follows that $\hat{\boldsymbol{\theta}}_{TWFE}-\boldsymbol{\theta}_0\rightarrow_p\mathbf{0}$, since $\hat{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right)-\boldsymbol{\theta}_0\rightarrow_p\mathbf{0}$, as already established. In short, $\hat{\boldsymbol{\theta}}_{TWFE}$ is a consistent estimator of $\boldsymbol{\theta}_0$ when $\hat{\boldsymbol{\psi}}_{TWFE}$ is a consistent estimator of $\boldsymbol{\psi}_0$. For carrying out inference on elements of $\boldsymbol{\theta}_0$, we suggest using the half-panel jackknife estimator, $\hat{\boldsymbol{\theta}}_{HJK}$, proposed for TWFE panel data models by Chudik, Pesaran and Yang (2018, CPY) instead of $\hat{\boldsymbol{\psi}}_{TWFE}$.[13] Using the jackknife estimator reduces the small T bias of the TWFE estimator and ensures the asymptotic equivalence of $\hat{\boldsymbol{\theta}}_{HJK}=\hat{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\psi}}_{HJK}\right)$ and $\hat{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right)$, which allows us to ignore the second term of (59), otherwise the contribution of the second term of (59) to the variance of $\hat{\boldsymbol{\theta}}_{TWFE}$ become non-negligible unless $n/T\rightarrow 0$. This condition

[13] Consider a balanced panel data with an even number of time series observations, $T=2T_h$. Denote the TWFE estimates on the first and the second half of the observations by $\hat{\boldsymbol{\psi}}_{a,TWFE}$, and $\hat{\boldsymbol{\psi}}_{b,TWFE}$, and the TWFE estimates based on all the time series observations by $\hat{\boldsymbol{\psi}}_{TWFE}$. Then the half-panel jackknife estimator is computed as

$$\hat{\boldsymbol{\psi}}_{HJK}=2\hat{\boldsymbol{\psi}}_{TWFE}-\frac{1}{2}\left(\hat{\boldsymbol{\psi}}_{a,TWFE}+\hat{\boldsymbol{\psi}}_{b,TWFE}\right).$$

is unlikely to be satisfied in the empirical growth literature. But as shown by CPY (Proposition 4), $\hat{\boldsymbol{\psi}}_{HJK} - \boldsymbol{\psi}_0 = O_p(n^{-1/2}T^{-1/2})$ if $N/T^3 \to 0$ as n and $T \to \infty$. Using this result it now follows that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{HJK} - \boldsymbol{\theta}_0\right) = \sqrt{n}\left(\hat{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right) - \boldsymbol{\theta}_0\right) + O_p(T^{-1/2}),$$

and valid inference can be made using $\hat{\boldsymbol{\psi}}_{HJK}$ to filter out the dynamics and the effects of time-varying covariates, even in the case of panels where n is large relative to T.

To allow for nonparallel time effects, but maintaining the slope homogeneity, we could use the jackknife version of the DCCEP estimator discussed above.

$$\hat{\boldsymbol{\theta}}_{DCCEP-HJK} = \left[n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i - \bar{\mathbf{z}}_\circ\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}_\circ\right)'\right]^{-1} n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i - \bar{\mathbf{z}}_\circ\right)\left(\bar{y}_{i\circ} - \bar{\mathbf{q}}_{i\circ}'\hat{\boldsymbol{\psi}}_{DCCEP,HJK}\right).$$

Once again using $\bar{y}_{i\circ}$ in (56) we have

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{DCCEP-HJK} - \boldsymbol{\theta}_0\right) &= \sqrt{n}\left(\hat{\boldsymbol{\theta}}\left(\psi\right) - \boldsymbol{\theta}_0\right) \\ &\quad - \mathbf{S}_{zz,n}^{-1}\left[n^{-1}\sum_{i=1}^{n}\left(\mathbf{z}_i - \bar{\mathbf{z}}_\circ\right)\bar{\mathbf{q}}_{i\circ}'\right]\left[\sqrt{n}\left(\hat{\boldsymbol{\psi}}_{DCCEP,HJK} - \boldsymbol{\psi}_0\right)\right].\end{aligned}$$

In this case the distributions of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{DCCE-HJK} - \boldsymbol{\theta}_0\right)$ and $\sqrt{n}\left(\hat{\boldsymbol{\theta}}\left(\psi\right) - \boldsymbol{\theta}_0\right)$ are asymptotically equivalent since $\sqrt{n}\left(\hat{\boldsymbol{\psi}}_{DCCEP,HJK} - \boldsymbol{\psi}_0\right) \to_p \mathbf{0}$, if $N/T^3 \to 0$ as n and $T \to \infty$.

In this paper we do not explore the use of country-specific estimates of ψ_i for filtering out the effects of the dynamics and the time-varying covariates, since the condition for the validity of this approach, namely $\sup_i \sqrt{n}\sup_i \left\|\hat{\boldsymbol{\psi}}_{i,HJK} - \boldsymbol{\psi}_i\right\| \to_p 0$, is unlikely to be met in practice. Using results in Theorem 1 of Chudik and Pesaran (2015a) it is possible to show that $\left(\hat{\boldsymbol{\psi}}_{i,HJK} - \boldsymbol{\psi}_i\right) \to_p 0$. However, this result on its own is not sufficient to establish the asymptotic equivalence of the distributions of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{DCCEMG-HJK} - \boldsymbol{\theta}_0\right)$ and its infeasible counterpart. This is a topic for further investigation.

9 Empirical results

To illustrate the methodological issues discussed above, we use the Penn World Tables, PWT, version 10.01, which provides data for up to 70 years over the period 1950-2019, together with the replication data of Acemoglu et al. (2019) and Kremer et al. (2022). The theory suggests that ignoring intercept heterogeneity, as is done in Barro regressions, biases the estimates of the speed of convergence towards zero. Similarly, wrongly imposing parallel trends or homogeneity on either the speeds of convergence or the other coefficients of the panel data model will also cause the resultant estimates to have biases that do not vanish with more data. It is further shown that when the parallel trends assumption does not hold then the popular TWFE estimator of the speed of convergence tends to zero if the steady state output growth in the global economy is non-zero. See the discussion before equation (36). Thus it is of interest to see how robust are the Barro and TWFE estimates to nonparallel trends and dynamic heterogeneity assumptions.

For output we use the national accounts measure of real GDP at constant 2017 national prices, expressed in million 2017US$, labelled *rgdpna* in PWT, which we denote by Y_{it}. Countries with fewer than 30 observations are removed, giving a baseline sample of 157 countries. It is not clear what scale measure should be used for cross-country comparisons of outputs. We consider population (POP_{it}) and employment (EMP_{it}), and work with log output per capita $y_{it} = \log(Y_{it}/POP_{it})$, as well as log output per employee, $\tilde{y}_{it} = \log(Y_{it}/EMP_{it})$. There are slightly shorter time series for employment in some countries, but always at least 30 years of data. Since we use up to 4 lags, the effective minimum number of time series observations used in panel regressions is 26. As we include covariates such as capital, education, and democracy, the number of countries and the number of time periods can vary across different panel regressions we report. For brevity we sometimes use output to refer to either of the two output variables.

For a visual picture of the time profiles of outputs across countries in Figure 1 we plot y_{it} for 99 countries covering all countries with data from 1960 onward except for a few countries with particularly erratic output profiles.[14] Despite the variability it is clear that a common (global) growth factor is driving output in this sample of countries, but it is unclear from this figure if the parallel trends assumption is likely to be met. But due to the upward trend in the global factor, we would expect that any violation of the parallel trends assumption would cause a substantial bias in Barro and TWFE estimates of the speed of convergence.

For a more formal analysis of the trend in the global factor, we estimated global output as a simple average of log output per capita of Asian, European, and North American. The output indices for these regions are computed as PPP-weighted averages, and their respective weights in the world economy are close to 1/3. Such a measure is more satisfactory than using simple averages of log per capita output based on a balanced panel of countries as depicted in Figure 1. U.S. log per capita output, in dark red, appears at the top of Figure 1, while our estimate of global output (GLO) is shown in dark blue in the middle of Figure 1. The GLO variable has a mean growth rate of 2.4 per cent per annum. Equations (29) and (30) show that the properties of f_t are transmitted to g_t and hence to y_{it},unless γ_i happens to be zero. It was shown in Section 5 that the DCCEP estimator of ρ_0 is consistent for n and T large, irrespective of the nature of the process for f_t. In contrast, in section 4 it was shown that when the parallel trends did not hold and f_t was trended, then $\hat{\rho}_{TWFE} \to 1$, irrespective of the value of ρ_0.

[14]The excluded countries are, namely Burundi, Côte d Ivoire , Gabon, Ghana, Guinea, Equatorial Guinea, Madagascar, Uganda Tanzania, Venezuela, Zambia and Zimbabwe

Figure 1: Log per capita output for selected countries over the period 1960-2019

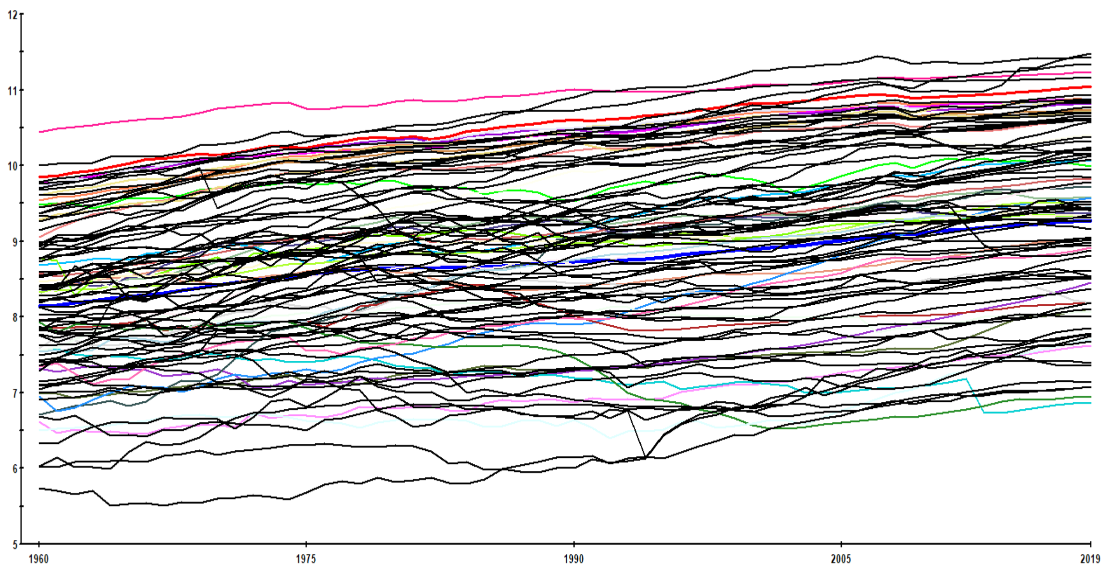

Table 1 gives some summary statistics for the variables we consider: the number of observations, overall mean, minimum, median, and maximum. Since we are interested in both time-varying and time-invariant variables, we give the summary statistics for the country means and the country standard deviations. For time-invariant variables we report their cross-country dispersions, noting that their time series dispersions is zero by construction. The summary statistics for output and capital per employee, not reported, are very similar to those reported for output and capital per capita.

Our empirical investigations closely follow the theoretical account set out above. We begin with Barro cross-country regressions in subsection 9.1; go on to panel autoregressions in subsection 9.2; consider panel regressions with time-varying covariates, capital, education and democracy, in subsection 9.3; and finish with cross-country regressions using time-invariant covariates in subsection 9.4. In the case of panel regressions we consider: TWFE, homogeneous slope models with parallel trends; DCCEP, homogeneous slope models with nonparallel trends; and DCCEMG heterogeneous slope models with nonparallel trends. The DCCE estimates are obtained with the Stata xtdcce2 package, Ditzen (2021).

Table 1: Summary statistics reporting dispersions of means and standard deviations of output and its possible drivers for the number of countries (n) for which there is data.

		Dispersion of country means, $\bar{y}_{i\circ}$				
Selected variables	n	Mean $\bar{y}_{\circ\circ}$	S.D. $sd_{\bar{y}}$	Min $\bar{y}_{\min}$	Median $\bar{y}_{med}$	Max $\bar{y}_{\max}$
per capita output growth	157	0.018	0.015	-0.023	0.019	0.067
per capita capital growth	157	0.026	0.019	-0.037	0.026	0.095
years of education	128	6.223	2.953	0.829	5.932	12.346
absolute latitude	144	0.300	0.195	0	0.261	0.722
ethno-linguistic fragmentation	121	0.364	0.308	0	0.294	1
protection against expropriation	107	7.203	1.729	0	7.045	10.00
		Dispersion of country standard deviations, $\overline{sd}_{i\circ}$				
	n	Mean $\overline{sd}_{\circ\circ}$	S.D. $sd_{\overline{sd}}$	Min $\overline{sd}_{\min}$	Median $\overline{sd}_{med}$	Max $\overline{sd}_{\max}$
per capita output growth	157	0.056	0.034	0.018	0.047	0.220
per capita capital growth	157	0.027	0.013	0.004	0.023	0.068
years of education	128	1.789	0.576	0.231	1.802	3.286
absolute latitude	144	0	0	0	0	0
ethno-linguistic fragmentation	121	0	0	0	0	0
protection against expropriation	107	0	0	0	0	0

Notes. The number of countries is n. The first panel reports summary statistics for the n country means, $\bar{y}_{i\circ} = T_i^{-1}\sum_{t=1}^{T_i} y_{it}$, $i = 1, 2, ..., n$, with different time series spans, T_i, for $i = 1, 2, ..., n$, covering the period 1950-2019, with means $\bar{y}_{\circ\circ} = \sum_{i=1}^{n} \bar{y}_{i\circ}/n$, and standard deviations, $sd_{\bar{y}} = \sqrt{\sum_{i=1}^{n}(\bar{y}_{i\circ} - \bar{y}_{\circ\circ})^2/(n-1)}$. The second panel is based on n country standard deviations, $sd_{i\circ} = \sqrt{\sum_{t=1}^{T_i}(y_{it} - \bar{y}_{i\circ})^2/(T_i - 1)}$, their mean $\overline{sd}_{\circ\circ} = n^{-1}\sum_{i=1}^{n} \overline{sd}_{i\circ}$, and their standard deviations $sd_{sd} = \sqrt{\sum_{i=1}^{n}(\overline{sd}_{i\circ} - \overline{sd}_{\circ\circ})^2/(n-1)}$. Both panels also report minimum, median and maximum of $\bar{y}_{i\circ}$ and $sd_{i\circ}$, respectively.

9.1 Estimates of the speed of convergence using Barro regressions

Consider the unconditional versions of the Barro regression given by setting $\boldsymbol{\theta} = \mathbf{0}$, in (17) and (16). namely

$$y_{iT} - y_{i0} = a_T + b_T y_{i0} + \zeta_{iT}, \tag{60}$$

where a_T and ζ_{iT} are defined by (18) and (20), respectively and $b_T = -(1 - \rho^T)$. The parameter of interest is the speed of convergence, $\phi = 1 - \rho$, and the mean number of years to convergence, given by $(1 - \phi)/\phi$.

While for equations that include a country specific intercept the base year does not matter, for Barro regressions which assume intercept homogeneity, $\alpha_i = \alpha$, the base year does matter. In estimating Barro regressions, Kremer et al. (2022, footnote 5) say "Specifically, for growth rates we use the variable "rdgpna," real GDP at constant 2017 national prices (2017 USD), and for growth levels we use "rdgpo," output-side real GDP at chained PPPs (2017 USD), as recommended by the PWT user guide." While we agree that output series based on national accounts are appropriate for constructing the dependent variable, $y_{iT} - y_{i0}$, it is difficult to rationalise having a different measure of output (for example the PPP measure) on the right hand side. Equation (12) is obtained by iterating (1) forward so that the same series must be used on the right hand side as the left hand side. Using the base year PPP measure, say y_{i0}^{PPP}, could be rationalised by assuming α_i is a function of the particular basket of goods produced in a country. Accordingly, we include both y_{i0} and y_{i0}^{PPP} as initial values in (60), or equivalently by adding $(y_{i0} - y_{i0}^{PPP})$ as an additional covariate with coefficient c_T. Namely, we estimate the following the cross-country regression

$$y_{iT} - y_{i0} = a_T + b_T y_{i0} + c_T \left(y_{i0} - y_{i0}^{PPP}\right) + \zeta_{iT}. \tag{61}$$

This equation reduces to the one estimated by Kremer et al. (2022), only if $c_T = -b_T$, which can be tested.

While regressions of this sort can be estimated on unbalanced samples, using $b_{T_i} = -(1-\rho^{T_i})$, in (60), we follow the literature and use a balanced sample. The choice of years $T = 2019$ and $0 = 1990$ gave a large value for n and the inclusion of the post-communist countries provide a wide range of initial outputs. Table 2 reports both the linear and non-linear estimates of equation (12) (with heteroskedasticity robust standard errors) using as dependent variable the change in the logarithm of output per capita between 1990 and 2019: $y_{i,2019} - y_{i,1990}$. In the first column the only regressor is initial output per capita, $y_{i,1990}$. The second column adds the difference between the national accounts and PPP measures $\left(y_{i,1990} - y_{i,19900}^{PPP}\right)$. It also provides comparable estimates using the logarithm of output per employee in 2019 and 1990: $\tilde{y}_{i,2019}$, $\tilde{y}_{i,1990}$.

In both 1990 and 2019 the two measures, output per capita and output per employee, are very similar with a cross-country correlation of 0.98. Thus it is not surprising that the two sets of Barro regressions give quite similar results. The PPP adjustment term $y_{i0} - y_{i0}^{PPP}$, has t ratios just over two and including it has little effect on the estimated speed of convergence, which range from 0.32 per cent per annum to 0.58 per cent. The restriction that $c_T = -b_T$ is strongly rejected, with t statistics of -3.10, for the per capita equation, and -3.96 for the per employee one. The table also gives the mean lag, estimated by $(1-\hat{\phi})/\hat{\phi}$, which ranges from 316 to 171 years, suggesting a very slow speed of convergence indeed. As we shall see, the very low estimates of ϕ obtained using Barro regressions is reflective of the bias towards zero of such estimates established in subsection

3.1.

Table 2: Estimates of Barro regressions and the implied speed of convergence using log per capita and per employee output using a balanced panel covering 157 countries over 1990-2019

	Output per-capita		Output per-employee	
	(1)	(2)	(1)	(2)
Regressors (estimates)				
$y_{i,1990}\ \left(\hat{b}_T\right)$	-0.0876 (0.0363)	-0.1025 (0.0346)	-0.1330 (0.0408)	-0.1559 (0.0365)
$y_{i,1990} - y_{i,1990}^{PPP}\ (\hat{c}_T)$	-0.2108 (0.1062)	....	-0.2420 (0.1117)	....
$\hat{\phi} = 1 - (1 + \hat{b}_T)^{1/T}$	0.0032 (0.0011)	0.0037 (0.0013)	0.0049 (0.0013)	0.0058 (0.0015)
Mean lag in years	316 (153)	268 (103)	203 (70)	171 (45)
Sample sizes				
Number of countries (n)	157	157	157	157
Years (T)	29	29	29	29

Notes. The dependent variable is $y_{i,2019} - y_{i,1990}$, where y_{it} is the log output variable at time t = 1990,2019. Estimates based on output per capita and output per employee use logarithms of "rgdpna/pop" and "rgdpna/emp" from *PWT 10.01*, respectively. b_T and c_T are given by the coefficients of the initial output variable, $y_{i,1990}$ and the coefficient of the PPP adjustment variable, $y_{i,1990} - y_{i,1990}^{PPP}$ where $y_{i,1990}^{PPP}$ is the log per-capita output in 1990 using the "rgdpo" purchasing power parity measure from *PWT 10.01.* Heteroskedasticity robust standard errors in parentheses. The estimates are computed by running non-linear cross-country regressions, the mean lag is estimated as $(1 - \hat{\phi})/\hat{\phi}$.

9.2 Estimates of speed of convergence based on dynamic panel data models

The solutions to stochastic growth models like that in Binder and Pesaran (1999) generate dynamic equations in output, which we represent by panel autoregressions. Unlike the Barro regressions, these allow for intercept heterogeneity, and are likely to be less biased. Table 3 reports estimates of the average speed of convergence, $\phi_0 = E(\phi_i)$ based on panel AR(3) regressions in log output per capita (y_{it}, on the left panel) and log output per employee ($\tilde{y}_{it}$, on the right panel). We report three estimates, TWFE estimates that assume parallel trends and homogeneous dynamics, (as in (1)), DCCEP estimates that allow for nonparallel trends but assume homogeneous slopes (as in (25)), and DCCEMG estimates that allow for both nonparallel trends and heterogeneous slopes, corresponding to equation (44). The DCCEMG estimates are computed using the error correction regressions in y_{it} (or $\tilde{y}_{it}$) augmented with cross section averages, $\bar{y}_{\circ t}$, and their lagged values:

$$\Delta y_{it} = \alpha_i - \phi_i y_{i,t-1} + \sum_{s=1}^{p_T-1} \delta_{is} \Delta y_{i,t-s} + \psi_{i0} \bar{y}_{\circ t} + \sum_{s=0}^{q_T-1} \psi_{is} \Delta \bar{y}_{\circ,t-s} + u_{it}, \tag{62}$$

The panel regressions are estimated using unbalanced panels, with $\bar{y}_{\circ t}$ computed as a simple average of output of countries with data in year t. The TWFE and DCCEP regressions are restricted version of the above specification.

The lag orders, p_T and q_T are set close to $T_{ave}^{1/3}$, as recommended by the theory with $q_T = p_T = 3$ under homogeneous dynamics, and $q_T = 4 > p_T = 3$, in the case of heterogeneous dynamics.[15] The use of TWFE estimators in the case of panels with heterogeneous dynamics results in terms like $(\phi_i - \phi_0) y_{i,t-1}$ to be included in the error term, which in turn could result in error serial correlation leading to more lagged terms being erroneously statistically significant. One might expect heterogeneous models to require fewer lags. This is in fact the case, with the statistical significance of $\Delta y_{i,t-2}$ falling as one moves from TWFE to the CCE type estimators. Furthermore, since f_t is likely to be trended in the present application, and as explained earlier, fewer lags of cross section averages are needed to adequately proxy the latent factor.

The estimates of the speed of convergence reported in Table 3 are in accordance with the theory which shows that relaxing the homogeneity restrictions reduces the downward bias in the estimated speed of convergence. For the case of output per capita, the estimates of the speed of convergence increases from 3 per cent per annum when TWFE is used, to 5 per cent if nonparallel trends are allowed by using DCCEP, and to 11 per cent when both the parallel trends and slope homogeneity restrictions are relaxed if estimation is carried out using the DCCEMG procedure. For output per employee the speeds are slightly higher. The mean number of years to convergence falls from 24-27 estimated using TWFE, to 14-18 using DCCEP, and to 5-6 years when using DCCEMG estimates (Table 3). Only the DCCEMG estimates correspond to the business cycle frequencies.

[15] T_{ave} is the average time dimension of the panel computed as $n^{-1} \sum_{i=1}^{n} T_i$, where T_i is the number of years of available data for country i and n is the total number of countries in the panel.

Table 3: Alternative estimates of speed of convergence based on unbalanced panel regressions of order three using PWT data covering 157 countries over the period 1950-2019

	Output per capita			Output per employee		
	TWFE	DCCEP	DCCEMG	TWFE	DCEEP	CEEMG
Regressors (estimates)						
$y_{i,t-1}$ $\left(-\hat{\phi}\right)$	-0.027 (0.004)	-0.046 (0.012)	-0.113 (0.012)	-0.031 (0.005)	-0.058 (0.014)	-0.144 (0.015)
$\Delta y_{i,t-1}$ $(\hat{\psi}_1)$	0.206 (0.030)	0.141 (0.029)	0.210 (0.018)	0.185 (0.034)	0.122 (0.0323)	0.130 (0.016)
$\Delta y_{i,t-2}$ $(\hat{\psi}_2)$	0.034 (0.027)	-0.015 (0.030)	-0.012 (0.014)	0.035 (0.029)	-0.019 (0.031)	0.001 (0.014)
Mean lag in years	27.1 (3.82)	18.2 (4.87)	6.1 (0.83)	23.81 (3.98)	14.4 (3.70)	5.1 (0.69)
Number of countries (n)	157	157	157	157	157	157
$\text{T}_{\min}$ (years)	27	27	26	27	27	26
T_{ave}	55.2	55.2	54.2	52	52	51
$\text{T}_{\max}$	67	67	66	67	67	66

Notes 1. The dependent variable is Δy_{it}, the change in log output per capita and log output per employee measured respectively using logarithms of "rgdpna/pop" and "rgdpna/emp" from *PWT 10.01*. Average speed of convergence, measured by $\phi = 1 - \rho_1 - \rho_2 - \rho_3$, is estimated as the coefficient of the lagged output variable. The mean lag is computed as $(1 - \hat{\phi} - \hat{\psi}_1 - \hat{\psi}_2)/\hat{\phi}$, where $\psi_1 = -(\rho_2 + \rho_3)$, $\psi_2 = -\rho_3$ and ρ_s, $s = 1, 2, 3$ are the (average) autoregressive coefficients. See also equation (62).
Notes 2. TWFE is the two way fixed effect estimator, DCCEP is the dynamic correlated common effect pooled estimator, and DCCEMG is the dynamic correlated common effect mean group estimator. See Sections (5) and (6) of the paper for further details. Heteroskedasticity robust standard errors are in parentheses, Huber-White for TWFE, the DCCE standard errors are heteroskedasticity robust. To filter out the nonparallel trends, DCCEP augments the panel regressions with the cross country averages $\bar{y}_{\circ,t-s}$, $s = 0, 1, 2, 3$, and the DCCEMG augments the panel regressions with $\bar{y}_{\circ,t-s}$, $s = 0, 1, 2, 3, 4$; the additional lagged cross section average, $\bar{y}_{\circ,t-4}$, is added to account for dynamic heterogeneity. Stata routine xtdcce2 was used for estimation.

9.3 Panel regressions with time-varying covariates

We now consider panel regressions that include time-varying covariates which we denote by the $k_x \times 1$ vector $\mathbf{x}_{it}$. Augmenting (62) with time-varying covariates and their lagged values we have

$$\begin{aligned}\Delta y_{it} = \alpha_i - \phi_i y_{i,t-1} + \textstyle\sum_{s=1}^{p_T-1} \delta_{is} \Delta y_{i,t-s} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \textstyle\sum_{s=0}^{p_T-1} \boldsymbol{\lambda}_{is}' \Delta \mathbf{x}_{i,t-s} \\ + \psi_{i0}' \bar{\mathbf{z}}_t + \textstyle\sum_{s=1}^{q_T-1} \psi_{is}' \Delta \bar{\mathbf{z}}_{t-s} + u_{it},\end{aligned} \quad (63)$$

where $\bar{\mathbf{z}}_t = (\bar{y}_{\circ t}, \bar{\mathbf{x}}_{\circ t}')'$, $\bar{\mathbf{x}}_{\circ t}$ is the cross section average of $\mathbf{x}_{it}$, and $\boldsymbol{\pi}_i = -\boldsymbol{\beta}_i / \phi_i$ is the long run elasticity of output with respect to $\mathbf{x}_{it}$. A large number of covariates have been considered in the literature as possible elements of $\mathbf{x}_{it}$. Here we focus on three of such variables only, namely physical capital, education as a proxy for human capital, and, a prominent institutional measure, democracy. The problem of how best to select a sub-set from a large number of potential covariates poses new methodological challenges and involves application of penalized regression techniques to dynamic panels with latent factors which is beyond the scope of the present application, primarily intended to illustrate the methodological issues discussed in this paper.

Our choice of capital and education variables is based on production function approaches to growth theories that postulate output in the long run to be a function of inputs of capital and labour as in a constant returns Cobb-Douglas production function, with the technology regarded as a latent factor.[16] For the capital variable we use log capital per-capita, $k_{it} = \log(K_{it}/POP_{it})$ to explain $y_{it} = \log(Y_{it}/POP_{it})$, and use log capital per-employee, $\tilde{k}_{it} = \log(K_{it}/EMP_{it})$ to explain $\tilde{y}_{it} = \log(Y_{it}/EMP_{it})$.[17] The impact effect of the capital input on growth is likely to be subject to simultaneity bias since cyclical demand and supply factors will complicate the identification of the short-run effects of capital accumulation. However, the long run effects (that we focus on) are more likely to be consistently estimated, particularly when we use CCE type estimators that allow capital and output to be jointly determined by the same set of latent factors representing tastes, technology, and demography.[18]

The panel data used for estimation is unbalanced and includes countries with at least 30 years of data, rendering a panel with 157 countries and a minimum of 26 data points, noting that the maximum lag order of the panel is set to $p_T = q_T = 4$. The estimation results are reported in Table 4 for the same set of specifications as in Table 3. As before, allowing for heterogeneity and nonparallel trends results in larger estimates for the speed of convergence, while at the same time reducing the required lag orders. We also see further increases in the estimated speed of convergence (irrespective of the estimation method or the choice of output and capital variables) once we condition on the capital variable. This is in line with the literature that finds conditional convergence to be faster than unconditional convergence.

We also report estimates of $\pi_k = E(\pi_{ik})$, the average long run elasticity of output with respect to capital, across the three estimation methods and the two measures of output and capital variables. Under some assumptions π_k should also measure the share of capital. Pesaran and Yang (2021) discuss the various estimates of the US share of capital used in the literature, which tend to be in the range 0.3-0.4. The estimates we obtain for π_k are somewhat higher, ranging from 0.48 to 0.64, with the DCCEMG estimates at 0.484

[16]Constant returns was not rejected using the DCCE estimators but was using TWFE.

[17]The capital measure K_{it} labelled rnna in PWT is capital stock at constant 2017 national prices (in millions 2017US$).

[18]Here we use a regression explaining output to estimate the long run coefficients. An alternative approach which does not involve specifying the dynamics or a direction of causation is described in Chudik, Pesaran and Smith (2025).

(0.122) (using output per capita) and 0.481 (0.084) (using output per employee) falling at the lower end. However, our estimates refer to an average across a wide range of countries. Over the 126 countries with available data, the average labour share in the PWT data is 0.54, that would suggest a coefficient on capital of around 0.46. It is interesting that the DCCEMG estimates, that fully allow for heterogeneity across countries, are close to this. As discussed above allowing for a latent factor that simultaneously drives capital accumulation and the output process reduces endogeneity bias (if any) for estimation of long run coefficients.

Our second time-varying covariate is the Barro-Lee measure of years of education. Table 5 reports the estimates with average years of education added to the panel regressions on its own and jointly with the capital variable. The number of countries in the panel is now reduced from 157 to 114, since not all countries in our baseline panel have time series data on the education variable, but allows us to increase the minimum number years of data available per country from 27 to 43 and 33, for per capita and per employee measures, respectively. Also, since in Table 4 the second lagged changes were not statistically significant, the estimates in Table 5 are based only on one lagged changes. For each of the three estimators (TWFE, DCCEP and DCCEMG), we report estimates for models with just capital, just education, and both. The six estimates of the long run coefficient of capital are very close to each other, between 0.53 and 0.57 for per capita measures, and between 0.47 and 0.52 for per employee measures. These are within the range of the estimates in Table 4. Education is only statistically significant when the capital stock variable is excluded and the panel is estimated by DCCEP using per capita measures. In this case the education variable has a t ratio of 2.12 for the impact effect and 2.24 for the long run effect. The TWFE estimates of the effect of education are negative, but statistically insignificant. Two robustness checks, not reported, were carried out. Firstly, lagged education was added and proved statistically significant only in 3 out of the 12 cases, and in no case was the long run effect of education significant. Second, we experimented with interacting the education variable with $\bar{y}_{\circ t}$, to see if, γ_i, the loading of the global factor, varies with education. Again the outcome was not significant.

Table 4: Alternative estimates of regressions explaining the change in log output by log capital based on unbalanced panel ARDL regressions of order three using PWT data covering 157 countries over the period 1950-2019.

	Output per capita			Output per employee		
	TWFE	DCCEP	DCCEMG	TWFE	DCCEP	DCCEMG
AR Coefficients						
$y_{i,t-1}$ (output)	-0.051 (0.007)	-0.131 (0.031)	-0.349 (0.025)	-0.057 (0.009)	-0.183 (0.030)	-0.375 (0.029)
$\Delta y_{i,t-1}$	0.163 (0.030)	0.061 (0.023)	0.118 (0.020)	0.178 (0.031)	0.118 (0.029)	0.132 (0.021)
$\Delta y_{i,t-2}$	0.028 (0.027)	-0.050 (0.028)	0.005 (0.016)	0.028 (0.030)	0.013 (0.044)	0.027 (0.018)
DL Coefficients						
k_{it} (Capital)	0.027 (0.005)	0.084 (0.026)	0.169 (0.045)	0.027 (0.005)	0.115 (0.025)	0.180 (0.034)
Δk_{it}	1.146 (0.110)	1.141 (0.187)	1.757 (0.156)	0.794 (0.043)	0.769 (0.058)	0.724 (0.156)
$\Delta k_{i,t-1}$	-0.705 (0.112)	-0.516 (0.165)	-0.815 (0.137)	-0.352 (0.050)	-0.279 (0.043)	-0.285 (0.044)
$\Delta k_{i,t-2}$	-0.069 (0.066)	-0.074 (0.085)	0.043 (0.121)	-0.010 (0.038)	-0.050 (0.050)	-0.029 (0.043)
Long run effect of capital	0.524 (0.043)	0.640 (0.105)	0.484 (0.122)	0.470 (0.044)	0.627 (0.068)	0.481 (0.084)
Sample sizes						
Number of Countries (n)	157	157	157	157	157	157
Year T_{min}	27	27	26	27	27	26
T_{ave}	55.2	55.2	54.2	52.0	52.0	51.0
T_{max}	67	67	66	67	67	66

Notes. k_{it} is either the logarithm of capital per capita "rnna/pop" or per employee "rnna/emp" from *PWT 10.01*, matching the way the output variable is scaled. To filter out the nonparallel trends, DCCEP augments the panel regressions with the cross country averages $\bar{\mathbf{z}}_{t-s}$, $s = 0, 1, 2, 3,$ where $\bar{\mathbf{z}}_t = (\bar{y}_{\circ t}, \bar{k}_{\circ t})'$, and the DCCEMG augments the panel regressions with the additional lagged cross section average, $\bar{\mathbf{z}}_{t-4}$ to account for dynamic heterogeneity. Speed of convergence, $\phi = 1 - \rho_1 - \rho_2 - \rho_3$, is estimated as the coefficient of the lagged output variable, $y_{i,t-1}$. See also Note 2 to Table 3

Table 5: Alternative estimates of regressions explaining the change in log output per capita or per employee by log capital per capita or per employee and education based on unbalanced dynamic panel regressions of order two.

Panel A:	Output per capita								
	TWFE			DCCEP			DCCEMG		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$y_{i,t-1}$	−0.049 (0.009)	−0.022 (0.004)	−0.049 (0.009)	−0.151 (0.025)	−0.094 (0.013)	−0.176 (0.029)	−0.263 (0.018)	−0.140 (0.014)	−0.329 (0.022)
$\Delta y_{i,t-1}$	0.141 (0.022)	0.204 (0.026)	0.141 (0.022)	0.117 (0.025)	0.169 (0.026)	0.117 (0.026)	0.124 (0.018)	0.207 (0.018)	0.125 (0.017)
k_{it}	0.027 (0.006)	...	0.027 (0.006)	0.085 (0.023)	...	0.094 (0.029)	0.151 (0.019)	...	0.174 (0.022)
Δk_{it}	1.172 (0.104)	...	1.171 (0.104)	1.176 (0.151)	...	1.157 (0.145)	1.735 (0.110)	...	1.701 (0.115)
$\Delta k_{i,t-1}$	−0.731 (0.099)	...	−0.731 (0.100)	−0.630 (0.128)	...	−0.605 (0.118)	−1.008 (0.106)	...	−0.892 (0.106)
$ed_{it}/100$	...	−0.172 (0.153)	−0.032 (0.134)	...	0.719 (0.338)	0.715 (0.424)	...	0.868 (0.657)	0.283 (0.902)
$L.R.\ k_{it}$	0.542 (0.052)	...	0.544 (0.054)	0.564 (0.088)	...	0.535 (0.103)	0.574 (0.064)	...	0.530 (0.064)
$L.R.\ ed_{it}$	...	−7.630 (7.292)	−0.648 (2.773)	...	7.610 (3.395)	4.062 (2.417)	...	6.216 (4.691)	0.861 (2.752)

Panel B:	Output per employee								
	TWFE			DCCEP			DCCEMG		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$y_{i,t-1}$	−0.052 (0.009)	−0.027 (0.005)	−0.052 (0.009)	−0.155 (0.039)	−0.108 (0.015)	−0.198 (0.030)	−0.292 (0.022)	−0.154 (0.016)	−0.343 (0.024)
$\Delta y_{i,t-1}$	0.146 (0.022)	0.160 (0.027)	0.145 (0.022)	0.112 (0.028)	0.116 (0.027)	0.120 (0.025)	0.119 (0.018)	0.106 (0.019)	0.120 (0.019)
k_{it}	0.025 (0.006)	...	0.025 (0.006)	0.079 (0.030)	...	0.096 (0.027)	0.149 (0.023)	...	0.180 (0.026)
Δk_{it}	0.840 (0.041)	...	0.838 (0.041)	0.826 (0.048)	...	0.797 (0.046)	0.776 (0.077)	...	0.741 (0.078)
$\Delta k_{i,t-1}$	−0.384 (0.042)	...	−0.385 (0.042)	−0.288 (0.046)	...	−0.278 (0.047)	−0.243 (0.049)	...	−0.237 (0.049)
$ed_{it}/100$	...	−0.329 (0.165)	−0.158 (0.142)	...	0.380 (0.356)	0.449 (0.408)	...	0.978 (0.796)	1.162 (1.060)
$L.R.\ k_{it}$	0.474 (0.059)	...	0.481 (0.061)	0.511 (0.094)	...	0.488 (0.082)	0.510 (0.071)	...	0.524 (0.067)
$L.R.\ ed_{it}$	...	−12.332 (7.220)	−3.062 (2.870)	...	3.510 (3.128)	2.268 (2.019)	...	6.333 (5.203)	3.387 (3.104)

Notes. ed_{it} is the Barro Lee 20-60 measure of average years of education taken from the Kremer et al. (2022) replication file. $n = 114$, minimum T_i is 43 for per capita, 33 for per employee, the maximum T_i is 55 for both, the average is 52.7 for per capita and 50 for per employee. L.R. indicates the long run coefficient of the variable. See also Notes 2 to Table 3.

There has been extensive research on the economic impact of democratization, most prominently by Acemoglu et al. (2019), and it is of interest to investigate the extent to which Acemoglu et al.'s estimates are robust to relaxing the assumptions of dynamic homogeneity and parallel trends. To focus on this problem we use their replication data set which is publicly available. Acemoglu et al. (2019, p.50) argue that existing measures of democracy suffer from measurement error and they develop a dichotomous democracy indicator, zero for non-democracy one for democracy, which combines several indices to purge spurious changes in each. They estimate a dynamic TWFE panel data model using four lags of log output per capita. This corresponds to our error correction specification with three lagged output changes. Their estimates are based on $n = 175$ countries for the years 1960-2010, but some countries only have six years of data, which is too few to relax the parallel trends assumption using CCE estimators. Accordingly, we consider a subset of $n = 148$ countries, that had at least 24 years of data, leaving a minimum of 20 time series observations per country for estimation, after allowing for lags.

Their democracy indicator shows time series variation in 73 of the 148 countries; in 75 countries it does not change. This lack of time series variation does not cause any problems when using pooled estimators such as TWFE and DCCEP, since these estimation methods assume the democracy variable to have the same effect across all countries, irrespective of whether their democracy indicator varies over time or not. This is a strong assumption which is difficult to test. We report estimates for two versions of DCCEMG that differ depending on whether homogeneity of the coefficient of the democracy variable is imposed or not. The first version, denoted by DCCEMG*, imposes the same coefficient on the democracy variable across all countries, thus including all the 148 countries, but differs from TWFE or DCCEP since it allows for dynamic heterogeneity. The second version, denoted by DCCEMG, allows for full parameter heterogeneity and uses the smaller sample of 73 countries whose democracy indicator varies over time.

The estimation results are reported in Table 6. The first column gives the Acemoglu et al.'s TWFE estimates based on their full sample of 175 countries and match their estimate of the long run effect of democracy, namely 21.240 (7.215), given in column 3 of their Table 2.[19] The subsequent columns give our TWFE estimates using the $n = 148$ sample, which is also used to produce the estimates in the columns under DCCEP and DCCEMG* that impose a homogeneous democracy coefficient. The final column gives the DCCEMG estimates using the $n = 73$ sample of countries with a time-varying democracy indicator, but using $\bar{y}_{\circ t}$ and $\bar{d}_{\circ t}$ computed as cross section averages of y_{it} and d_{it} corresponding to the $n = 148$ sample. While the point estimates of the long run effects for the three DCCE estimators are similar, the $n = 73$ estimate in the final column is not comparable to the other two DCCE estimates. It is an average treatment effect on the treated, where the treatment is a change in the democracy indicator.

[19] The bracketed figures give standard errors clustered by country.

Table 6: Alternative estimates of equations explaining the change in log output per capita using the dichotomous democracy indicator in unbalanced panel regressions using Acemoglu et al. (2019) data 1960-2010.

	Output per capita				
	TWFE		Dynamic CCE		
	Acemoglu et al.	This paper	DCCEP	DCCEMG*	DCCEMG
$y_{i,t-1}$ (Lagged log output)	-0.037 (0.005)	-0.036 (0.005)	-0.164 (0.020)	-0.218 (0.022)	-0.348 (0.035)
$\Delta y_{i,t-1}$	0.275 (0.036)	0.287 (0.035)	0.210 (0.043)	0.186 (0.025)	0.207 (0.034)
$\Delta y_{i,t-2}$	*0.069* (0.023)	0.058 (0.025)	0.018 (0.034)	-0.054 (0.018)	-0.016 (0.027)
$\Delta y_{i,t-3}$	0.042 (0.017)	0.047 (0.020)	0.057 (0.028)	0.030 (0.015)	0.086 (0.021)
d_{it} (Democracy indicator)	0.787 (0.226)	0.769 (0.226)	0.947 (0.574)	1.285 (0.654)	1.499 (1.034)
Long run effect of Democracy	21.24 (7.21)	21.53 (7.47)	5.78 (3.59)	5.90 (3.06)	4.30 (3.01)
Sample Sizes					
Number of countries n	175	148	148	148	73
Years					
$T_{\min}$	6	20	20	20	20
T_{ave}	36.2	40	40	40	41.3
$T_{\max}$	47	47	47	47	47

Notes. The dependent variable, Δy_{it}, the change in log output per capita, and d_{it}, the dichotomous democracy indicator, are both from the replication data set provided by Acemoglu et al. (2019). The first column, headed Acemoglu et al., replicates their results, subsequent columns exclude countries where the minimum T_i used for estimation was less that 20. To filter out the nonparallel trends, DCCEP and DCCEMG use cross section averages $\bar{y}_{\circ,t-s}$, $s = 0, 1, 2, 3$ and $\bar{d}_{\circ,t}$. DCCEMG* imposes that the coefficient of democracy is the same across all 148countries, whilst DCCEMG allows for full heterogeneity across the 73 countries in the panel with time-varying democracy indicators. See also the notes to Table 3.

A comparison of the first two columns of Table 6 shows that dropping countries with less than 24 years of data makes virtually no difference to the TWFE estimates. Looking across the first row of the table shows that, as in the preceding tables, relaxing parallel trends and dynamic homogeneity result in increased estimates of the average speed of convergence across countries, as predicted by the theory. Also, whereas all three lagged output changes are statistically significant when using TWFE, the lagged changes are less significant in the case of the other estimates that relax the homogeneity restrictions. Turning to the impact effect of democracy, the point estimates are very similar to those obtained by Acemoglu et al. (2019), though their statistical significance falls as we allow for a greater degree of heterogeneity, and become statistically insignificant if we condition on countries where the democracy indicator varies. A similar pattern can be observed if we consider the long run estimates. The estimates of the long run coefficient of the democracy variable fall as we relax the parallel trends and dynamic homogeneity assumptions. The statistically significant estimate reported by Acemoglu et al. (2019) become either insignificant or only marginally significant, although all the estimates have the expected positive sign. This loss of statistical significance is not due to multicollinearity between the country-specific democracy indicator, d_{it}, and the cross section average, $\bar{d}_{\circ t}$.[20] When the latter is dropped, democracy is less significant. For instance the DCCEP estimate falls from 0.947 with a t-ratio of 1.65 when $\bar{d}_{\circ t}$ is included, to 0.748 with a t-ratio of 1.27 when $\bar{d}_{\circ t}$ is excluded from the panel regression.

9.4 Regressions with time-invariant covariates

Finally, we turn to the effects of time invariant covariates on α_i and γ_i, following the procedure set out in subsection 8. From the large number of time-invariant variables considered in the literature and recently documented by Kremer et al. (2022), we consider an indicator of geography, *absolute lattitude*; an indicator that has often been associated with conflict, *ethnolinguistic fractionalization*; and an indicator of property rights, *protection from expropriation.* To these we add our PPP indicator, $y_{i,1990} - y_{i,1990}^{PPP}$ to control for possible systematic differences in the national accounts and PPP measures of output in the initial year of the sample. There are also many other variables that one could consider, such as alternative measures of climate change, technological adaptation, governance approaches, and institutional features. However, due to the challenging new methodological issues that surround the problem of variable selection in the context of high dimensional heterogeneous dynamic panel data models, we confine our empirical analysis to a few variables highlighted as potentially important in the literature. An extension of our empirical analysis to selection from a large set of covariates is beyond the scope of the present paper and must be the subject of a separate investigation.

The estimation of the time-invariant effects follows two steps: in the first step we estimate the coefficients of the time-varying effects used to estimate a_{iT} in (56). To avoid the small T bias of these estimates we only consider their pooled estimates, namely TWFE and DCCEP, denoted by $\hat{\boldsymbol{\psi}}_{TWFE}$ and $\hat{\boldsymbol{\psi}}_{DCCEP}$, estimated using panel ARDL(2,2) regressions in log output and log capital (either per capita or per employee). We also increase the minimum number of observations used for these first stage panel regressions, so that the average number of time periods used is over 60 years. In the second stage, we run cross-country regressions of $\hat{a}_{iT} = \bar{y}_{i\circ} - \bar{\mathbf{q}}_{i\circ}'\hat{\boldsymbol{\psi}}$, with $\hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\psi}}_{TWFE}$ or $\hat{\boldsymbol{\psi}}_{DCCEP}$, on

[20] We are grateful to Daron Acemoglu for suggesting this possibility.

an intercept and $\mathbf{z}_i$ which is an 4×1 vector of time-invariant regressors, with $n = 100$ countries.

The estimation results are summarized in Table 7. The estimated coefficient of *absolute latitude* is not statistically significant. In contrast, the variables *ethnolinguistic fractionalization* and *protection from expropriation* are both statistically significant with the expected signs: a higher probability of conflict or weaker protection of property rights both have negative output effects. The PPP correction coefficient is positive and significant. These outcomes are robust to the choice of the first stage estimates $\boldsymbol{\psi}$.

Table 7: Cross-country regressions estimating the coefficients (standard errors) of time-invariant regressors using different procedures to filter the time-varying effects.

	Output per capita		Output per employee	
	TWFE	DCCEP	TWFE	DCCEP
Dependent variable, $\hat{a}_i$				
absolute latitude (z_{i1})	0.383 (1.289)	0.832 (2.568)	0.747 (1.283)	1.353 (3.390)
ethnolinguistic fractionalization (z_{i2})	−1.71 (0.608)	−3.062 (1.341)	−1.643 (0.658)	−4.554 (1.831)
protection from expropriation (z_{i3})	0.699 (0.122)	0.952 (0.236)	0.627 (0.135)	1.219 (0.336)
$y_{i,1990} - y^{PPP}_{i,1990}$ (z_{i4})	0.704 (0.272)	1.374 (0.709)	0.883 (0.328)	2.675 (0.984)

Notes. The reported estimates are computed by least squares regressions of $\hat{a}_i = \bar{y}_{i\circ} - \bar{\mathbf{q}}'_{i\circ}\hat{\boldsymbol{\psi}}$, on an intercept and $\mathbf{z}_i = (z_{i1}, z_{i2}, z_{i3}, z_{i4})'$, for $i = 1, 2, ..., n = 100$, where $\bar{\mathbf{q}}_{i\circ}$is the time averages of lagged output and capital variables for country i, $\hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\psi}}_{TWFE}$ or $\hat{\boldsymbol{\psi}}_{DCCEP}$ are TWFE and DCCEP estimators obtained using ARDL(2,2) panel regressions in log output and log capital (either per capita or per employee). Data on $\mathbf{z}_i$ are from Kremer et al. (2022) and are all divided by 100 to scale up the estimates provided. Standard errors are heteroskedasticity robust. The maximum time period used for estimation of $\hat{\boldsymbol{\psi}}$ was 68 years. $T_{\min} = 48$ and $T_{ave} = 63$ years for regressions based per capita measures, and $T_{\min} = 38$ and $T_{ave} = 61$ years for regressions based on per employee measures.

10 Concluding remarks

The empirical results presented in the previous section are in line with the theory and show that large biases can result from wrongly assuming parallel trends and/or homogeneous dynamics. Barro or pooled regression estimates of the speed of convergence are biased towards zero unless all slope coefficients in the panel are homogeneous and the heterogeneity of the fixed effects can be modelled perfectly: an impossible undertaking. TWFE estimates allow for intercept heterogeneity, but continue to be substantially biased if its two main assumptions of parallel trends and dynamic homogeneity are invalidated. Most crucially, when the latent factor is trended (deterministic or stochastic) and its effects differ across countries, the TWFE estimate of the speed of convergence tends to zero, no matter what is its true value.

To deal with nonparallel trends and dynamic heterogeneity we consider DCCEP and DCCEMG estimators that are the dynamic versions of the CCE estimator originally proposed by Pesaran (2006). DCCEP relaxes the parallel trends assumption but maintains dynamic homogeneity, whilst DCCEMG allows for both. We show that DCCEP produces consistent estimates even if the latent factors driving the growth process are trended, with or without unit roots.

In the empirical applications and in line we our theoretical results, we find that the mean convergence time of output to its steady state trend falls from hundreds of years if one uses Barro estimates, to 27 years if one uses TWFE, and to 6 years if one uses DCCEMG estimates, which closely match the business cycle frequency in OECD countries. We have relatively quick convergence to a country specific steady state trend, but there is no evidence of global cross-country convergence.

When capital is added it has a sensible long run coefficient. Relaxing the parallel trends assumption reduces the endogeneity problem, since the introduction of a latent global factor with a country specific effect reduces the omitted variable problem arising from common factors driving both output and capital.

We also investigated the effects of time-invariant variables on average deviations of output from its steady state trend, and found statistically significant positive effects from higher levels of *protection from expropriation* and statistically significant negative effects from high levels of *ethnolinguistic fractionalization*.

There remain a number of areas for further research both in the econometric methods and in the determinants of the differences in output across countries. These include

- The high dimensional selection problem in panels for a more comprehensive empirical analysis of the determinants of output from the large number of fast and slow moving variables suggested in the literature.

- Inference on the role of time-invariant effects and how to separate their effects on the level of output (α_i) and the trend output growth (γ_i).

- How to distinguish between time-varying and time-invariant effects of climate. The time-varying effects have been extensively investigated in the literature. See, for example, Kahn, Mohaddes, Ng, Pesaran, Raissi, and Yang (2021), and references cited therein.

- Convergence between countries, perhaps within clubs, can be investigated by looking at pairwise convergence, as in Pesaran (2007).

Although this paper has focussed on economic growth similar issues come up in other research areas. For example, Westerlund, Karabiyik, Narayan and Narayan (2022) consider similar econometric issues in the context of applications to corporate finance.

References

Acemoglu, D., S. Johnson, J.A. Robinson (2001) The colonial origins of comparative development: an empirical investigation, *The American Economic Review*, 91: 1369-1401.

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

Aghion, P., P. Howitt (1992) A model of growth through creative destruction, *Econometrica* 60:323-351.

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106: 407-44.

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

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

Baumol, W.J. (1986) Productivity growth, convergence, and welfare: what the long-run data show, *American Economic Review*, 76: 1072-1085.

Binder, M., M.H. Pesaran (1999) Stochastic growth models and their econometric implications, *Journal of Economic Growth*, 4: 139-183.

Chauvet, M., C. Yu (2006) International business cycles: G7 and OECD countries, *Economic Review*, Federal Reserve Bank of Atlanta, First Quarter: 43-54.

Chudik, A., M.H. Pesaran (2015a) Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors, *Journal of Econometrics*, 188: 393-420.

Chudik, A., M.H. Pesaran (2015b) Large panel data models with cross-sectional dependence: A survey, Chapter 1, *The Oxford Handbook on Panel Data,* B. H. Baltagi (Editor), Oxford University Press, Oxford.

Chudik, A., M.H. Pesaran, R.P.Smith (2025) Analysis of multiple long run relations in panel data models. https://arxiv.org/abs/2506.02135v3

Chudik, A., M.H. Pesaran, J-C Yang (2018) Half-panel jackknife fixed effects estimation of linear panels with weakly exogenous regressors, *Journal of Applied Econometrics*, 33: 816-836.

De Vos, I., G. Everaert (2021) Bias-corrected common correlated effects pooled estimation in dynamic panels, *Journal of Business and Economic Statistics*, 39: 294-306.

Ditzen, J. (2021) Estimating long-run effects and the exponent of cross-sectional dependence: an update to xtdcce2, *The Stata Journal*, 21: 687-707.

Durlauf, S. (2009) The rise and fall of cross-country growth regressions, *History of Political Economy*, 41 (annual supplement): 315-333.

Everaert, G., T. De Groote (2016) Common correlated effects estimation of dynamic panels with cross-sectional dependence, *Econometric Reviews*, 35: 428-463.

Friedman M. (1992) Do old fallacies ever die? *Journal of Economic Literature*, 30: 2129-32.

Granger, C.W.J. (1980) Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics*, 14: 227–238.

Hausman, J.A., W.E. Taylor (1981) Panel data and unobservable individual effects, *Econometrica*, 49: 1377-1398

Hayakawa, K., M.H. Pesaran, L. V. Smith (2023) Short T dynamic panel data models with individual, time and interactive effects, *Journal of Applied Econometrics,* 38: 940-967.

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

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

Juodis, A,. H. Karabiyik, J. Westerlund (2021) On the robustness of the pooled CCE estimator, *Journal of Econometrics*, 220: 325-348.

Kahn, M.E., K. Mohaddes, R.N.C. Ng, M.H. Pesaran, M. Raissi, J-C. Yang (2021) Long-term macroeconomic effects of climate change: a cross-country analysis, *Energy Economics, 104: 105624.*

Kapetanios, G., M.H. Pesaran, T. Yamagata (2011) Panels with non-stationary multifactor error structures, *Journal of Econometrics*, 160: 326-384.

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

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

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

Mankiw, N.G., D. Romer, D.N. Weil (1992). A contribution to the empirics of economic growth, *Quarterly Journal of Economics*, 107: 407–37.

Maseland, R. (2021) Contingent Determinants, *Journal of Development Economics*, 151: 102654.

Moon, H.R., M. Weidner (2017) Dynamic linear panel regression models with interactive fixed effects, *Econometric Theory*, 33: 158-195.

Nerlove, M. (1998) Growth rate convergence, fact or artifact? An essay on the use and misuse of panel data dconometrics, WP 98-21, Department of Agricultural and Resource Economics, The University of Maryland.

Nerlove, M. (1999) Properties of alternative estimators of dynamic panel models: an empirical analysis of cross-country data for the study of economic growth, in Chapter 6 of *Analysis of Panels and Limited Dependent Variable Models: In honour of G.S. Maddala*, by C. Hsiao, K. Lahiri, L-F Lee, and M. H. Pesaran (Editors), Cambridge University Press, Cambridge.

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

Ong, K. (2024) Do countries converge to their steady states at different Rates? *Open Economies Review* 35: 723–749

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

Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* 74: 967-1012.

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

Pesaran, M.H., A. Chudik (2014) Aggregation in large dynamic panels, *Journal of Econometrics*, 178: 273-285.

Pesaran, M.H., R.P. Smith (1995) Estimating long-run relationships from dynamic heterogenous panels, *Journal of Econometrics, 68: 79-113.*

Pesaran, M.H., L.Yang (2024) Heterogeneous autoregressions in short T panel data models, *Journal of Applied Econometrics*, 39: 1359-1378.

Pesaran, M.H., Q. Zhou (2018) Estimation of time-invariant effects in static panel data models, *Econometrics Reviews*, 37: 1137-1171.

Pesaran M.H., Y. Xie (2026) How to detect network dependence in latent factor models? A bias-corrected CD test, *Econometric Theory*, published online.

Pesaran, M.H., L.Yang (2021) Estimation and inference in spatial models with dominant units, *Journal of Econometrics*, 221: 591-615.

Robinson, P.M. (1978), Statistical inference for a random coefficient autoregressive model, *Scandinavian Journal of Statistics*, 5: 163–168.

Sarafidis, V., T. Wansbeek (2021) Celebrating 40 years of panel data analysis: Past, present and future, *Journal of Econometrics*, 220: 215-226.

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

Temple, J. (1999) The new growth evidence, *Journal of Economic Literature*, 37: 112-156.

Westerlund, J. (2018) CCE in panels with general unknown factors, *The Econometrics Journal*, 21: 264-276.

Westerlund, J., H. Karabiyik, P.K. Narayan and S. Narayan (2022) Estimating the speed of adjustment of leverage in the presence of interactive effects, Journal of Financial Econometrics, 20: 942-960.