European Economic Review 147 (2022) 104173

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

# Democracy, growth, heterogeneity, and robustness ☆

Markus Eberhardt *

*School of Economics, University of Nottingham, UK*
*Centre for Economic Policy Research, UK*

ARTICLE INFO

*JEL classification:*
O10
P16

*Keywords:*
Democracy
Growth
Political development
Difference-in-difference estimator
Interactive fixed effects

ABSTRACT

I motivate and empirically investigate differential long-run growth effects of democratisation across countries. While the existing literature recognises the potential for such heterogeneity, empirical implementations to date unanimously assume a common democracy-growth nexus across countries. Adopting novel methods for causal inference in policy evaluation I relax the homogeneity assumption. My results confirm that in the long-run democracy has a positive and significant average effect on per capita income, albeit at 10% this is at best half the magnitude of recent estimates in the literature. Guided by existing theories, additional analysis probes the patterns of the heterogeneous 'democratic dividend' across countries. Adopting two rule-based robustness exercises I furthermore demonstrate that, in contrast to recent contributions to the literature, my approach yields empirical findings that are robust to substantial changes to the sample.

## 1. Introduction

Recent empirical work has suggested that in the long run the economic 'returns' to democratisation are statistically significant and large: using a continuous measure of democracy and data from 1820 to 2000 the 2SLS regressions in Madsen et al. (2015) suggest a doubling of income per capita for a one standard deviation improvement in democracy. Adopting a new binary measure for democracy during 1960 to 2010 the 2SLS regressions in Acemoglu et al. (2019, henceforth ANRR) suggest a 31% increase in income per capita for permanent regime change. These results are novel and important since previous contributions to the literature[1] frequently failed to establish a significant positive relationship (Helliwell, 1994; Barro, 1996; Baum and Lake, 2003 and Murtin and Wacziarg, 2014 for continuous democracy indices and Giavazzi and Tabellini, 2005; Rodrik and Wacziarg, 2005 for binary measures).[2]

☆ I am indebted to Matteo Cervellati, Marc Chan, Rodolphe Desbordes, Jonathan Dingel, Doug Gollin, Brian Knight, Simon Kwok, Gerard Padro i Miquel, Andrea Presbitero, Daniel Seidmann, Jon Temple, Daniel Treisman, Nicolas van de Sijpe, Dietz Vollrath, and Joakim Westerlund for comments and suggestions on earlier drafts which helped improve the paper significantly. Seminar/session participants at Lund, Nottingham, Middlesex, and the Annual International Conference of the Research Group on Development Economics in Berlin provided useful feedback. Changes in the light of detailed comments from two anonymous reviewers have significantly improved the paper. The usual disclaimers apply.

* Correspondence to: School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, UK.
*E-mail address:* markus.eberhardt@nottingham.ac.uk.

[1] Exceptions include the large positive effects in Papaioannou and Siourounis (2008) who like ANRR emphasise the importance of taking care in *defining* regime change events, but in comparison with the latter can less credibly claim their results constitute causal effects. Similarly, Knutsen (2013) uses a continuous Freedom House Index but employs the Arellano and Bond (1991) GMM estimator which many researchers regard with scepticism in the context of cross-country regressions.

[2] A separate strand of the literature studies democratic capital and finds significant effects in the long-run using two-way fixed effects models (e.g. Gerring et al., 2005; Persson and Tabellini, 2009) — for these and other contributions see the review in Appendix Table B-1. A broader set of papers is reviewed in Dodsworth and Ramshaw (2021).

https://doi.org/10.1016/j.euroecorev.2022.104173
Received 23 March 2021; Received in revised form 19 December 2021; Accepted 20 April 2022
Available online 14 June 2022

All of the above studies adopt empirical models which treat the democracy-growth relationship[3] as *common* across countries.[4] Econometrically, the distinction between a common estimate derived from a pooled model, say $\hat{\beta}$, and an average estimate from a heterogeneous parameter model, say $\hat{\beta}^{MG} = \sum_i \hat{\beta}_i$, may seem innocuous. The authors of the aforementioned studies would likely not disagree with the principle of a heterogeneous democracy effect, instead pointing to their own estimates as *some form of cross-country average*.[5] Yet, the distinction matters greatly for identification in the most trusted form of empirical implementation: instrumental variable regression — it is well-known (but under-appreciated) that misspecifying a heterogeneous relationship as common violates the basic assumptions of 2SLS estimators (Pesaran and Smith, 1995, see Section 3.1 for more details).

Leaving identification aside, the distinction between 'pooled' and 'heterogeneous' effects also matters for policy making: if the average effect hides substantial differences in the impact of democracy *across countries*, then reporting an average effect is arguably at best only useful to make a point ('democracy *on average* causes growth') and at worst seriously misleading for policy decisions in individual countries. Maybe not all countries experience large positive growth effects of democratic regime change and it is self-evidently important to isolate the determinants of this heterogeneity. A range of studies in economics and political science has developed arguments in favour of heterogeneous economic effects of democratisation, based on *inter alia* the nature of the democratic transition, power structures, or populism (e.g. Cervellati and Sunde, 2014, Albertus and Menaldo, 2018, Funke et al., 2020, Treisman, 2020, Boucekkine et al., 2021), while the literature on sufficient conditions for democratisation has studied income or human capital thresholds as conditions for democracy to be a lasting (economic) success (Przeworski et al., 2000; Madsen et al., 2015, ANRR) — I empirically investigate these theories in Section 3.4 below. Arguments from a broader range of literature, including on democracy and structural change (Acemoglu et al., 2015) and the link between democracy and innovation (Aghion et al., 2014) as well as heterogeneous knowledge diffusion (Comin and Hobijn, 2004; De Visscher et al., 2020), similarly provide motivation for the use of *country-specific* empirical models in cross-country analysis more generally.

Hence, there is a wealth of arguments why the economic implications of democracy may differ across countries, why it matters to see whether average *heterogeneous* effects are still positive, large and significant, and why we may want to identify the patterns of this heterogeneous relationship. The empirical practice, however, is to estimate pooled models ignoring all of these important questions.

This paper represents a first attempt at shedding light on the implications of a heterogeneous 'democracy-growth' nexus. In order to benchmark the findings against the most important contribution in this literature, the ANRR study, I adopt the dataset and with this the definition of democracy from these authors. I study the causal link between democracy and growth when the long-run equilibrium relationship can differ across countries and investigate what a range of existing theories can tell us about the potential determinants of such a 'heterogeneous democracy dividend'. I adopt a factor-augmented difference-in-difference implementation which can account for heterogeneous trends prior to regime change and the endogenous selection of countries into democracy (Chan and Kwok, 2022). The implementation is quite simple: each 'treated' country regression is augmented with common factor proxies estimated from the control group of countries which never transitioned into democracy — these factors make up the macroeconomic variables in all countries and in combination with country-specific parameters can provide a great deal of flexibility in modelling *unobserved* heterogeneity.[6] Like in a pooled regression with time-*invariant* country fixed effects, the time-*varying* 'interactive fixed effects' (Bai, 2009) of the common factor setup and their country-specific parameters can be correlated with the other regressors, most notably the democracy dummy: this allows me to accommodate major challenges to identification such as selection into democracy and non-parallel trends across countries.

I further emphasise the robustness of empirical estimates to changes in the sample. Cross-country growth empirics typically fails to question the sample makeup, and to the best of my knowledge there are no examples of existing studies subjecting their findings to rigorous sample reduction exercises: robustness checks may drop certain 'types' of countries (e.g. former Socialist economies), but these exercises are not informed by the 'quality' or 'quantity' of the data at hand — for instance, the number of country observations.[7] Similarly, whether a sample ends in 2020 or 2015 or 2010 is typically not questioned, provided the end date is reasonably recent. While this situation is difficult to improve upon (there will always be countries with limited data), my concerns over robustness speak to recent research by Broderick et al. (2020) and Young (2022) who highlight that regression results can be heavily influenced by a small share of observations: how do results hold up when a small numbers of observations are systematically dropped? I devise two rule-based sample reduction strategies, dropping countries by the length of their time series or end years of the sample, and apply them to my empirical findings and for comparison to those of ANRR's seminal study.

[3] I follow ANRR in using 'growth' as a shorthand for economic development (per capita GDP) in the long-run and estimate levels, not growth equations — see Eberhardt and Teal (2011) for a detailed discussion of the interpretation of 'cross-country *growth* regressions'.

[4] Some of the papers cited here have employed interaction terms to highlight the differential growth impact of this or other characteristic studied. None however allow for the democracy-growth relationship to be estimated entirely flexibly (within the confines of the parametric model) as is done in this paper.

[5] It is well-known that in a static model the fixed effects estimator for $x$ is a weighted average of underlying country-specific slope coefficients (random coefficients following Swamy, 1970), where the weights are related to the $i$-specific variation in $x$ — see Sul (2016). In contrast, when the model is dynamic a fixed effects estimator is inconsistent for the coefficients on the lagged dependent and $x$ variables respectively, due to the presence of serial correlation in the residuals which are thus correlated with the lagged dependent variable (Pesaran and Smith, 1995).

[6] Although there are many differences between these methods, the analogy to the popular synthetic control method, in particular in its generalised version (Xu, 2017), can help with the intuition of this approach: untreated countries provide the building blocks (the common factors), which can then be flexibly assembled in the treated sample (by including them as additional covariates) to account for the challenges to identification, namely differential trends prior to and selection into 'treatment'.

[7] While it is tacitly acknowledged that the countries *included in the regression sample* are typically not a representative sample of the population of countries in the world, these concerns are brushed aside when the sample contains a large number of countries or the majority of a group of countries (e.g. OECD countries for the analysis of advanced economies).

There are three main findings: first, while the estimated long-run impact of democracy on income per capita in my analysis is still positive and significant, it is at best *less than half* the magnitude of the effect reported in ANRR.[8] Simple quartile estimates can demonstrate that differences in the long-run democracy-growth effect can be fairly large across countries. Second, my analysis of the patterns of this cross-country heterogeneity finds no evidence for the relevance of a democratic legacy, that only rich countries can make democracy 'work' for growth, or that democratisations orchestrated by autocratic elites result in lower growth effects. There is some evidence that the relationship between human capital and the magnitude of the democratic dividend could be convex (bad news for aspiring democracies with intermediate levels of human capital stock), while there are also indications that the democracy effect is a one-off levels effect (in line with ANRR) rather than a permanent growth effect. Somewhat surprisingly, democratisation 'by mistake' (Treisman, 2020), where autocratic incumbents unintentionally pave the way to democracy, appears to be associated with *higher* long-run growth. Third, the robust and consistent results for the democracy-growth nexus across different estimators in ANRR was not apparent in the rule-based sample reduction exercises I conducted: the 2SLS estimates, as well as the supporting evidence using GMM estimators, are highly sensitive to the exclusion of comparatively few observations, with between 3 and 7% of the total rendering the IV estimates insignificant.[9] In contrast, my own heterogeneous Diff-in-Diff estimates turn insignificant when 40% and one third of observations are dropped by country minimal observation count and end year, respectively, highlighting that these methods produce much more stable and robust empirical findings.

The remainder of this paper is organised as follows: in the next Section I briefly discusses the data sources and descriptive patterns of democratisation. In Section 3 I introduce my empirical methodology and the main results as well as the potential explanations for patterns of heterogeneity across country. The robustness exercises in Section 4 focus on whether my findings hold when I reduce the cross-section or time-series dimension of the panel. A conclusion follows.

## 2. Data and descriptives

Given the prominence of their work as well as the robustness of their empirical results across different (pooled) implementations, I adopt the dataset and sample from ANRR. Most importantly, this includes a new democracy indicator which combines information from two separate sources and is further argued to do away with the ex-post selection problem inherent in earlier studies, where researcher re-coded single-year democracy episodes as autocratic. The main sources are the Freedom House Index and the polity2 variable from PolityIV (Marshall et al., 2017) — the consolidated measure for democracy is equal to 1 when the former indicates a country is 'partially free' or 'free' *and* the polity2 variable is positive. When these indices are not both available the authors employ additional standard sources including Boix et al. (2013, BMR) and Cheibub et al. (2010, CGV).[10] The measure is refined by adjusting it to match the timings of permanent democratisations coded by Papaioannou and Siourounis (2008, PS). In my analysis I further explore the results using the indicators by BMR, CGV and PS (the latter is limited to permanent democratisations). A schematic representation in Appendix Figure A-1 highlights the different coverage of political institutions inherent in these different indicators.

Income per capita data (in real year 2000 US$ values), the share of gross investment in GDP and trade openness (the sum of exports and imports divided by GDP) come from the World Bank World Development Indicators (WDI) database. All of the above variables (democracy dummies, income and controls) are compiled by ANRR and provided for download from Daron Acemoglu's personal website.

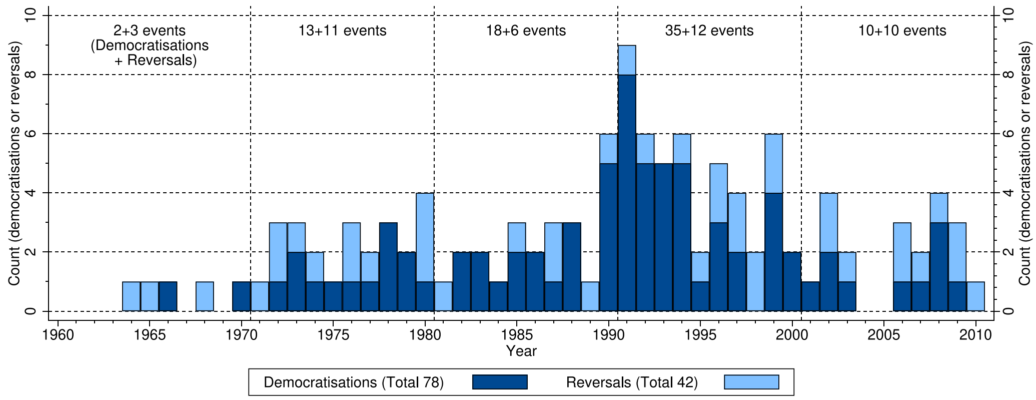
Fig. 1 Panel (a) plots the distribution of democratisation events as well as reversals to autocracy (following the ANRR definition), restricted to my dynamic regression sample of 61 countries (treatment sample). The average country had 1.3 democratisation events and 0.7 reversals; 16 countries experienced 2 or more democratisations (THA had 4), 8 countries had two or more reversals — in additional analysis below, Section 3.3, I will separately report the average long-run effect of democratisation for countries which experienced a single democratisation as well as that for other sub-groups. Panel (b) indicates the distribution of the number of years countries spent in democracy — this histogram is for democratic spells and distinguishes lasting democratisation from democratic episodes which were subsequently reversed. The latter dominate the left tail of the distribution for 10 or fewer years. The sample median (for lasting and overturned democratisations) is 12 years in democracy, while lasting democracies on their own have a median 18 years of 'treatment'.[11] Panel (c) provides a histogram for the total years spent in democracy per country, without any

[8] Like ANRR (see their Appendix A4) I find only relatively minor differences in this result when using alternative definitions of democracy (all binary) by Papaioannou and Siourounis (2008, PS), Cheibub et al. (2010, CGV), and Boix et al. (2013, BMR).

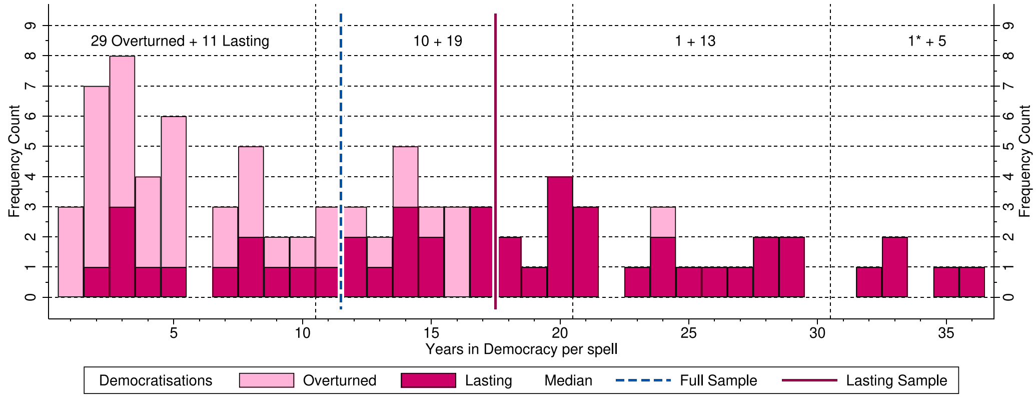
[9] The GMM and HHK estimates typically turn insignificant when 5% of observations are dropped. In selected specifications of Madsen et al. (2015) analysed using the same sample reduction exercises substantially more observations, between 11% and 34% of the total (the benchmark result in one exercise remains statistically significant throughout), need to be dropped in order for results to turn statistically insignificant.

[10] For more details see Appendix A1 of ANRR. Their democracy indicator is designed to "purge spurious changes" (50) in each of its constituent elements. Curiously, these 'spurious changes' in PS, CGV and BMR (ANRR and BMR's democracy dummies are in agreement in 93% of country observations where they are jointly defined, while for ANRR and CGV the figure is 92%), despite sample sizes which are 5%–9% smaller than ANRR's, still yield next to identical results for the preferred IV specification, with only the CGV implementation yielding a somewhat higher (40%) democratic dividend. Perhaps yet more curiously, the model dynamics for all four democracy indicators yield an *identical* persistence parameter of 0.964, with *identical* standard error of 0.005 (a $t$-statistic of 193)!

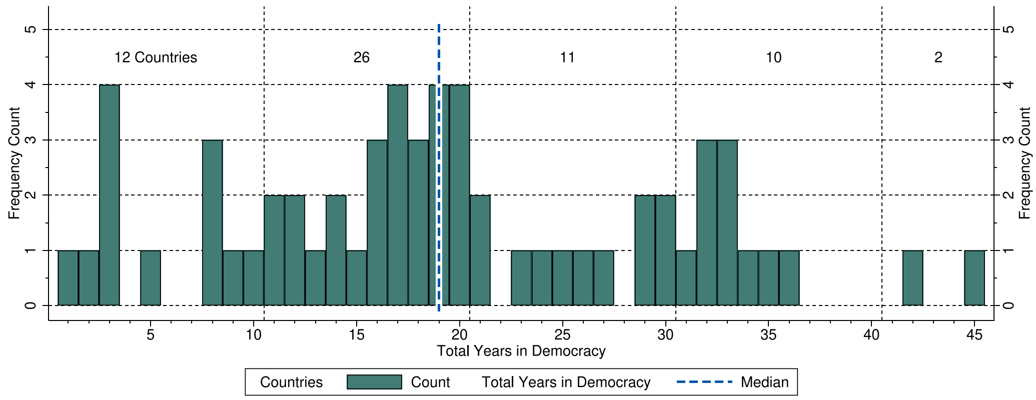
[11] For comparison, in the AB sample of ANRR with 6,161 observations these figures are 16.5 and 23. Note, however, that they include data from over 40 countries which were democratic *throughout the sample period*. When these are excluded (as is the case in my implementation) the median years in democracy of democratising countries are 11 and 17.5 for lasting democracies only — hence very similar to my above figures. Note that ANRR's preferred IV estimate rises from a long-run democracy effect of 30.5% to 42.3% (se 19.907) when the 'always-democracies' are excluded. A whopping democratic dividend given the median 11 years of treatment.




(a) Distribution of Events (Democratisation and Reversal)




(b) Distribution of Years Spent in Democracy by Spell




(c) Distribution of Years Spent in Democracy by Country

**Fig. 1.** Democratisation, Reversal, and Time in Democracy — 1960–2010. *Notes*: The histogram in panel (a) presents the distribution of democratisation events (dark blue) and reversals to autocracy (light blue) over the 1960 to 2010 period based on the ANRR definition of democracy — the sample here is that of the dynamic specification for which results are presented in Table 1, Panel (a). The histogram in panel (b) indicates the number of years spent in democracy for different episodes/spells of democracy, split into those which eventually reverted into autocracy ('overturned') and those which did not ('lasting'). ∗ For ease of illustration I omit one democratic episode of 45 years (which was overturned). The medians for the full sample (12 years) and the sample of lasting democratisations (18 years) is highlighted by the respective vertical lines. Panel (c) indicates the total number of years in democracy per country.

concern in how many spells; here the median is 19 years. With the exception of PS, who exclusively focus on *lasting* democracies (single spells), I am not aware of any existing research on the democracy-growth nexus which acknowledges the repeated 'back and forth' some countries experience with respect to democratisation and reversal.

**Table 1**
Main results—Dynamic specification (long-run estimates).

| | Plain vanilla | | With covariates | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Implementation | MG | C&K MG | MG | C&K MG |
| Parameters estimated | 5 × N | 14 × N | 13 × N | 22 × N |
| **(a) Democracy (ANRR)** | 16.624 | 7.692 | 7.712 | 10.074 |
| | (4.630)*** | (2.854)*** | (3.647)** | (3.651)*** |
| Observations | 2443 | 2443 | 2443 | 2443 |
| Countries (N) | 61 | 61 | 61 | 61 |
| Democratisations | 78 | 78 | 78 | 78 |
| Reversals | 42 | 42 | 42 | 42 |
| Avg Years in Dem | 19.6 | 19.6 | 19.6 | 19.6 |
| RMSE | 18.861 | 7.942 | 8.515 | 4.115 |
| **(b) Democracy (BMR)** | 15.130 | 7.322 | 9.983 | 11.118 |
| | (4.057)*** | (3.279)** | (2.843)*** | (3.612)*** |
| Observations | 2051 | 2051 | 2051 | 2051 |
| Countries | 55 | 55 | 55 | 55 |
| Democratisations | 66 | 66 | 66 | 66 |
| Reversals | 35 | 35 | 35 | 35 |
| Avg Years in Dem | 18.5 | 18.5 | 18.5 | 18.5 |
| RMSE | 17.976 | 6.466 | 8.200 | 3.726 |
| **(c) Democracy (CGV)** | 14.405 | 4.989 | 10.154 | 6.934 |
| | (4.572)*** | (4.059) | (3.190)*** | (4.179)* |
| Observations | 1922 | 1922 | 1922 | 1922 |
| Countries | 50 | 50 | 50 | 50 |
| Democratisations | 68 | 68 | 68 | 68 |
| Reversals | 34 | 34 | 34 | 34 |
| Avg Years in Dem | 20.0 | 20.0 | 20.0 | 20.0 |
| RMSE | 19.366 | 7.080 | 8.711 | 4.117 |
| **(d) Democracy (PS)** | 27.915 | 4.749 | 11.501 | 11.936 |
| | (4.813)*** | (4.594) | (4.445)*** | (5.072)** |
| Observations | 1670 | 1670 | 1670 | 1670 |
| Countries | 41 | 41 | 41 | 41 |
| Democratisations | 41 | 41 | 41 | 41 |
| Reversals | 0 | 0 | 0 | 0 |
| Avg Years in Dem | 21.6 | 21.6 | 21.6 | 21.6 |
| RMSE | 17.351 | 7.888 | 8.336 | 4.135 |

*Notes*: The table presents Robust mean estimates from heterogeneous panel estimators using different definitions of democracy: (1) and (3) simple Mean Group estimator, (2) and (4) Chan and Kwok (C&K) DID Mean Group estimator — all are estimated using least squares. We hold the sample fixed across the four specifications, but not when comparing different definitions of democracy. All estimates presented are long-run (ATET) estimates for the causal effect of democracy on income per capita (in percent), derived from a CS-DL model (Chudik et al., 2016). The models in (3) and (4) include gross investment ratio and trade/GDP as additional covariates. The four alternative democracy dummies are by Acemoglu et al. (2019) – ANRR, Boix et al. (2013) – BMR, Cheibub et al. (2010) – CGV, and Papaioannou and Siourounis (2008) – PS.

My empirical implementation discussed below discards countries which remained democracies throughout the 1960–2010 sample period, which amounts to 1,892 observations in 45 countries. The control sample is for all those countries which never transitioned into democracy, for which there are 1,194 observations in 38 economies — these will be used to construct the cross-section averages of income per capita, the gross investment share/GDP and trade openness (see next section). Prominent economies included in this control sample are China, Malaysia and Viet Nam. See Appendix Table A-1 for details on the control sample makeup.

The potential treatment sample has 3,017 observations in 84 countries. However, estimating dynamic heterogeneous panel models is demanding and hence I have to impose a minimum time-series observation count of 24 — given that the focus of the analysis is on the long-run, I do not feel that this introduces undue distortions.[12] My dynamic specifications are for a maximum of 61 countries with 2,433 observations and cover 78 democratisation events as well as 42 reversals — see Appendix Table A-2 for

[12] Compared with the ANRR dynamic IV regressions this omits 16 countries with 19 democratisation events and 6 reversals (given additional data availability for gross investment and trade): *ARM* (18 out of 20 available years in democracy for two democratisations), DJI (9/17), GEO (7/23), *GIN (1/24),* HRV (11/19), HTI (9/19 for two democratisations), KGZ (5/20 for two democratisations), LBN (6/21), LVA (18/20), *PRY (1/19),* RUS (11/21), SLB (7/11), SVK (18/23), *SVN (19/20),* UKR (17/21) — for the four countries in italics it is questionable whether a democracy effect could be identified given the small number of observations in the pre- or post-treatment regime. If I reduce the number of parameters to be estimated in each country by specifying a more parsimonious dynamic specification with single lags of the covariates and up to two lags of the cross-section averages, ten of these 16 remain in the sample. My robust long-run mean estimate for the C&K MG estimator in this larger sample is 7.868*** (se 2.378), around 2 percentage points lower than in my main sample.

details on countries in the 'treated' sample.[13] The samples for the alternative definitions of democracy are smaller and capture fewer events.

## 3. Heterogeneity

### 3.1. Methodology: Heterogeneous treatment effects

I this paper I study the impact of observable and unobservable heterogeneity on empirical estimates of the democracy-growth nexus. My model builds on the panel time series econometric literature which has emphasised heterogeneous parameters across panel members (Pesaran and Smith, 1995) and, more recently, the presence of strong cross-section dependence (e.g. Pesaran, 2006; Bai, 2009), a form of unobserved, time-varying heterogeneity.[14] This literature has taken to specifying a multi-factor error structure, $\lambda_i' \boldsymbol{F}_t$, where $\boldsymbol{F}$ is a set of common factors with associated heterogeneous factor loadings $\lambda$, to capture this dependence. The most recent contributions to this econometric literature have been able to build bridges to the work on policy evaluation using the synthetic control methodology (Xu, 2017) or the difference-in-difference specifications (Gobillon and Magnac, 2016; Chan and Kwok, 2022) most suited to the present empirical setup. What distinguishes these latest approaches from their canonical predecessors is the adoption of the multi-factor error structure to address three challenges to identification in these popular methods: (i) the presence of uncommon trends prior to the 'treatment' evaluated, (ii) endogenous selection into 'treatment', and (iii) the possibility that, following the 'treatment', treated and control samples are affected by common shocks, albeit with heterogeneous impact (e.g. the differential effect of the Global Financial Crisis across countries).

*Intuition.* In this paper I adopt the Chan and Kwok (2021) Principle Component Difference in Difference (PCDID) estimator,[15] although as I explain below, my implementation does not use principle component analysis but an augmentation following Pesaran (2006) instead. The intuition of this approach, which builds on the general motivation for the use of 'common factors' in cross-country growth regressions, is as follows: there are many variables and processes — some of them measurable, like investment, others less so, like culture or absorptive capacity — which can be argued to influence economic development, far too many to accommodate in a cross-country study with annual data.[16] Since we typically account for some proxy for investment and population the remainder of determinants is what growth economists refer to as total factor productivity (TFP). Since TFP and its influence is pervasive, we know that if we fail to capture this appropriately in our regression then the variable of interest, here the democracy dummy, is likely subject to significant bias ('transmission bias', Marschak and Andrews, 1944). In the early days of cross-country empirics, Robert Barro (1991) simply added as many control variables as possible to the estimation equation, but this 'kitchen sink' approach is commonly blamed for failing to produce robust results (Durlauf, 2020) and lately frowned upon (though still surprisingly common). Coming up with credible time-varying instrumental variables to avoid the transmission bias, such as ANRR's regional waves of democratisation, is hard, and the GMM estimators by Arellano and Bond (1991) and Blundell and Bond (1998) which adopt *internal* instruments have fallen out of favour, again for the large variability in the resulting empirical findings. This is where the common factor setup comes in: instead of trying to capture every single process and variable as in the 'Barro regressions', we can use a dimensionality-reducing device such as principle component analysis (PCA) to estimate a small number of 'common factors', which, in combination with country-specific 'factor loadings,' can capture the richness and diversity of country-specific TFP.[17]

The Chan and Kwok (2022) estimator uses the factor structure to allow for correlation between observable and unobservable determinants of income per capita on the one hand, and the democracy dummy on the other (selection effect), as well as to capture differential trends across countries, most importantly between treated and control countries. From country regressions in the *control group*, made up of all countries which never transitioned to democracy, they use the error terms to estimate common factors (via PCA), which are then added to the country regressions of the *treated group*.[18] One complication in the present context is that PCA in unbalanced panel data is cumbersome, since missing observations have to be estimated using an expectation maximisation methodology. The panel literature however provides an alternative avenue to capturing the common factors via PCA (à la Bai, 2009), namely by including cross-section averages of all variables (à la Pesaran, 2006) — see details below.

Finally, in this literature the distinction between short-run and long-run clearly matters. In the short-run, we may see that democratic regime change leads to upheaval, 'democratic overload' and a 'tumultuous youth' (Gerring et al., 2005; Papaioannou and Siourounis, 2008; Cervellati and Sunde, 2014; Acemoglu et al., 2019), whereas with longer time 'in regime' governments learn from experience, improve policy making, establish meritocratic bureaucracy, and become legitimised in the eyes of the populace (Gerring

[13] For the static specifications, result of which are relegated to an appendix, I can draw on 83 countries with 3,052 observations — see Appendix Table A-3.

[14] Strong cross-sectional correlation is distinct from weaker forms of dependence, e.g. spatial correlation, and can lead to serious bias in the estimated coefficients on observable variables (Phillips and Sul, 2003; Andrews, 2005).

[15] Previous work analysing the democracy-growth nexus using difference-in-difference specifications includes Giavazzi and Tabellini (2005), Papaioannou and Siourounis (2008) and Cervellati and Sunde (2014).

[16] Durlauf et al.'s (2005: Appendix B) survey lists around 150 separate determinants.

[17] The literature on macro forecasting was among the first to recognise that a few common factors could replace hundreds of macroeconomic variables, which made it feasible to include the information from the latter in a forecasting model in a parsimonious fashion (Stock and Watson, 2002).

[18] This PCA idea goes back to Bai (2009), but the somewhat surprising insight in the work of Chan and Kwok (2022) is that the estimated PCs are 'good enough' proxies in this setup, whereas in Bai's implementation they (and the parameters on the variable of interest) are biased and need to be estimated iteratively.

et al., 2005). Econometrically, since the results from a *static* estimator are not readily comparable with those from the *dynamic* specifications investigated in the existing literature cited above, I focus primarily on results for a dynamic 'CS-DL' version (cross-section-augmented distributed lag; Chudik et al., 2016) of the Chan and Kwok (2022) estimator, relegating the static results to Appendix C.1.

*An important aside on pooled 2SLS regressions.* It has been known for some time that 'heterogeneity misspecification', modelling a heterogeneous relationship with a pooled (homogeneous) model, violates the basic assumptions of 2SLS estimators (Pesaran and Smith, 1995): if the true coefficient on the variable of interest $x_{it}$ is $\beta_i$, yet the implementation imposes $\beta$, the error term $\varepsilon$ *by construction* contains $(\beta_i - \beta)x_{it}$. It is now easy to see that due to the presence of $(\beta_i - \beta)x_{it}$ in the error no potential instrument $z$ can both be relevant, $E[z_{it}x_{it}] \neq 0$, and valid, $E[z_{it}\varepsilon_{it}] = 0$. This econometric argument highlights the serious implications for any claims of 'causal inference' when heterogeneity is ignored.

*Formal model and implementation.* In the following I develop the formal econometric specification. In the potential outcomes framework, democratic regime change (the 'treatment') for country $i$ at time $T_{0i}$ can be written as

$$y_{it} = \text{Dem}_{it}\, y_{it}(0) + (1 - \text{Dem}_{it})\, y_{it}(1) = \Theta_{it}\mathbf{1}_{\{i\in E\}}\mathbf{1}_{\{t>T_{0i}\}} + y_{it}(0) \tag{1}$$

$$\text{where} \qquad y_{it}(0) = \varsigma_i + \boldsymbol{\beta}_i'\boldsymbol{X}_{it} + \lambda_i'\boldsymbol{F}_t + \widetilde{\epsilon}_{it}. \tag{2}$$

The two indicator variables $\mathbf{1}_{\{\cdot\}}$ refer to the country and the year treated, $\Theta_{it}$ is the time-varying heterogeneous treatment effect, $\boldsymbol{X}$ is a vector of observed covariates with associated country-specific parameters $\boldsymbol{\beta}_i$, $\lambda_i'\boldsymbol{F}_t$ represents a set of unobserved common factors $\boldsymbol{F}_t$ with country-specific factor loadings $\lambda_i$,[19] and $\widetilde{\epsilon}_{it}$ is the error term.

The treatment effect is country-specific and time-varying, but we assume it follows a decomposition $\Theta_{it} = \overline{\Theta}_i + \widetilde{\Theta}_{it}$, where $E(\widetilde{\Theta}_{it}|t > T_{0i}) = 0$ as $\widetilde{\Theta}_{it}$ is the cross-sectionally demeaned idiosyncratic component of $\Theta_{it}$; $\overline{\Theta}_i$ is the ITET, *the treatment effect of country $i$ averaged over the post-intervention period*. The reduced form model is then

$$y_{it} = \overline{\Theta}_i\mathbf{1}_{\{i\in E\}}\mathbf{1}_{\{t>T_{0i}\}} + \varsigma_i + \boldsymbol{\beta}_i'\boldsymbol{X}_{it} + \lambda_i'\boldsymbol{F}_t + \epsilon_{it}. \tag{3}$$

The composite error term, $\epsilon_{it} = \widetilde{\epsilon}_{it} + \widetilde{\Theta}_{it}\mathbf{1}_{\{i\in E\}}\mathbf{1}_{\{t>T_{0i}\}}$, has zero mean due to the decomposition assumption made above, but it can be heteroskedastic (perhaps due to spatial correlation) and/or serially correlated.

In the Chan and Kwok (2022) implementation common factor proxies estimated from the control group sample of 'never-democracies' are added as additional covariates to the country-specific equation for treated (democratising) countries, which is then simply estimated by least squares. The setup in Eq. (3) allows for non-parallel trends between treated countries and control sample. It can also accommodate various correlations between different elements of the equation, most notably between the treated units or timing of treatment and the factor loadings or the observed covariates $\boldsymbol{X}$. This implies that *democratisation can be endogenous to observed variables* (gross investment/GDP and trade openness) *and unobserved common factors*: time-varying latent driving forces of economic development such as culture or absorptive capacity can be correlated with democratisation.

The most important assumptions underpinning this approach are (i) that the unobservables can be captured by the common factor structure, as is laid out in the panel time series literature I cite above and Athey et al. (2021), among others; and (ii) that the composite error term $\epsilon$ is orthogonal to $\boldsymbol{X}$, $\boldsymbol{F}$, parameters and factor loadings, as well as the treatment dummies: any selection into democracy is fully captured by the other elements of the model (most notably, the common factors).

For all countries which experienced regime change (from democracy to autocracy or vice versa)[20] I specify the following static regression model

$$y_{it} = \alpha_i + \theta_i\,\text{Dem}_{it} + \boldsymbol{\beta}_i'\boldsymbol{X}_{it} + \delta_i^y\overline{y}_t + \boldsymbol{\delta}_i^{X'}\overline{\boldsymbol{X}}_t + \varepsilon_{it}, \tag{4}$$

where $y$ is per capita GDP (in logs and multiplied by 100), Dem is the democracy dummy, and $\boldsymbol{X}$ is a set of additional controls (gross investment share of GDP and trade openness). $\overline{y}$ and $\overline{\boldsymbol{X}}$ are the cross-section averages of the observed variables *but for those countries which never experienced democracy during the sample period* (the control group).[21] As was shown by Pesaran (2006) and Westerlund and Urbain (2015), the use of cross-section averages is very simple yet powerful in capturing a common factor structure.

The dynamic variant of Eq. (4) is:

$$\begin{aligned} y_{it} \;=\; & \alpha_i + \theta_i^*\,\text{Dem}_{it} + \boldsymbol{\beta}_i^{*'}\boldsymbol{X}_{it} + \sum_{\ell=0}^{p-1}\omega_{i\ell}^{D}\Delta\text{Dem}_{i,t-\ell} + \sum_{\ell=0}^{p-1}\boldsymbol{\omega}_{i\ell}^{X'}\boldsymbol{\Delta X}_{i,t-\ell} \\ & + \sum_{\ell=0}^{p_{\bar{y}}}\delta_{i\ell}^{*y}\overline{y}_{t-\ell} + \sum_{\ell=0}^{p_{\overline{X}}}\boldsymbol{\delta}_{i\ell}^{*X'}\overline{\boldsymbol{X}}_{t-\ell} + \varepsilon_{it}, \end{aligned} \tag{5}$$

where the two terms involving sums in the first line capture the short-run effects, while $\theta_i^*$ and $\boldsymbol{\beta}_i^*$ represent the long-run coefficients for the effects of democracy and additional controls on income per capita, respectively — I use stars to indicate that the interpretation

[19] One feature of this empirical approach is that it allows for nonstationary common factors $F$.

[20] With the exception of PS it is common in this literature to lump together single and multiple regime switchers, including countries which only reversed to democracy during the sample period — in Section 3.3 I will have a closer look at the implications of this convention for my results.

[21] Country and time fixed effects represent a special case of the interactive effects $\lambda_i'F_t$ captured by these cross-section averages. Note that by construction there is no cross-section average for the democracy variable, since this is always zero in the control group from which these are computed.

of the ITET and the covariate coefficients is different from that in Eq. (4): here, these are *long-run* estimates derived from a dynamic specification. The sums in the second line capture the multifactor error structure using cross-section averages, which like in the static model are constructed from those countries which never experienced democracy during the sample period. The use of this 'CS-DL' version of the Chan and Kwok (2022) approach is convenient since the long-run democracy coefficient, $\theta_i^*$, can be estimated in a single step rather than two as in an error-correction specification or the ANRR ARDL implementations.[22] Following suggestions in Chudik et al. (2016) I adopt $p_{\bar{y}} = 0$ and $p = p_{\overline{X}} = int(T^{1/3}) = 3$, where $T$ is the time dimension of the panel. My presentation below will focus on average estimates of $\hat{\theta}^*$ in the dynamic case (which can be interpreted as ATET estimates); in line with the literature I adopt robust regression (Hamilton, 1992) to compute outlier-robust means. In the sources of heterogeneity analysis in Section 3.4 I employ the country-specific ITET estimates $\hat{\theta}_i^*$. Inference for all 'Mean Group' estimates is based on standard errors computed non-parametrically, following Pesaran and Smith (1995). Observed covariates $\boldsymbol{X}$ are not included in what I refer to as the 'plain vanilla' Chan and Kwok (2022) implementation — the covariate cross-section averages from the control sample, $\bar{y}$ and $\overline{X}$, are however always included.[23] For comparison, I also estimate simple Mean Group models (Pesaran and Smith, 1995) which exclude the cross-section averages in Eq. (5).

### 3.2. Main empirical results

In Table 1 I provide the robust mean estimates (ATET) for two alternative specifications of two heterogeneous estimators: in the first two columns the 'plain vanilla' empirical models do not include the observed values for gross investment share of GDP and trade openness as regressors, in the final two columns they do; MG is a simple 'mean group' estimator of a model which excludes the cross-section averages in the second line of Eq. (5), whereas C&K MG is the Chan and Kwok (2022) estimator — the latter is the preferred implementation. All results presented are long-run estimates derived from the dynamic specification. The different panels present results using alternative definitions of democracy (in all cases dummy variables), with the conceptually preferred ANRR definition at the top, followed by BMR, CGB, and PS.[24]

My results for the ANRR definition cover 61 countries, which experienced 78 democratisation events in 2,433 country-year observations (48% of which are 'in democracy'). When additional covariates are excluded, the MG estimate for the long-run effect of democracy on growth is 16.6%. Accounting for pre-treatment non-parallel trends and selection into democratic regime change pushes this estimate down to around 7.7%. The models with investment share and trade openness as additional covariates find the reverse pattern, with the simple MG long-run estimate at 7.7% and the C&K MG estimate at 10.1%.[25]

In panel (b) I adopt the BMR democracy indicator, which despite conceptual differences (see Figure A-1) and a different sample makeup yields remarkably similar long-run estimates in all four models. The preferred Chan and Kwok (2022) Mean Group estimate in column (4) at 11% is only marginally higher than when adopting the ANRR definition of democracy. Results in panel (c) for the CGV democracy indicator deviate somewhat, in that the two Chan and Kwok (2022) MG estimates in (2) and (4) indicate weaker democracy effects as well as much less precise estimates, while the two heterogeneous models which ignore selection and non-parallel trends in (1) and (3) are very similar to previous results (and highly statistically significant) — this sample ends in 2008 rather than 2010 (ANRR), but since the BMR sample ends in 2007 and the average years in democracy are actually *higher* in CGV than ANRR or BMR, this is unlikely to account for this deviation.[26] Panel (d) adopts the PS definition which is limited to 41 permanent democratisations and excludes any autocratic reversals, resulting in two more years spent in democracy for the average country and a long-run democracy effect of 12% in the preferred C&K MG model. It is notable that the standard MG model without adjustment for selection into democracy in column (1) arrives at an average democracy effect of 28%.

Hence, the average long-run effect of democratic transition in the preferred implementation ranges from 7 to 12% across these four different specifications. This translates into only one-half to one-third of the long-run effects found in ANRR, depending on their implementation: accounting for parameter heterogeneity still yields a positive average democracy-growth effect, but of considerably smaller magnitude. Since especially the PS definition of democracy varies substantially from the others, which allow multiple back and forth of democratisation and reversal (as well as cases of 'pure' reversal), I now shift my attention to the implications of such a 'mixed' treatment sample.

---

[22] In the ECM specification we obtain an estimate $\hat{\beta}_i$ for democracy and $\hat{\rho}_i$ for the lagged dependent variable (or $\sum_{\ell=1}^{p} \hat{\rho}_{i\ell}$ for $p$ lags), from which the long-run coefficient $\hat{\theta}_i = \hat{\beta}_i / -\hat{\rho}_i$ has to be computed. It is apparent from this that any finite sample bias in $\hat{\rho}_i$ will carry over to $\hat{\theta}_i$ (Chudik and Pesaran, 2015). The CS-DL obtains these estimates in a single step by adopting an alternative specification and avoids potential bias from dynamic misspecification.

[23] Merely adding $\bar{y}$ allows for a single unobserved common factor $f$, whereas inclusion of $\overline{X}$ allows for multiple common factors.

[24] The dynamic C&K MG implementation requires estimation of 21 parameters plus an intercept, however due to missing observations the minimum requirement for the ANRR definition of democracy is 24 observations, although the first country in the sample had a minimum of 26. In order to make the estimates across different implementations (i.e. across columns) directly comparable I fix the sample at the C&K MG minimum for $T_i$.

[25] In this and all the following cases, the latter implementation results in the lowest root mean squared error (RMSE), hence indicating that this cross-section averaged-augmented difference-in-difference estimator with additional controls provides the best fit for the data.

[26] If, in line with my analysis in Panel (b) of Table 2 I exclude countries which only reverse to democracy during the sample period (here: UGA), the robust mean estimate using the BMR definition is close to 8% with a $t$-statistic of 1.88.

**Table 2**
ANRR definition—Dynamic specifications.

| | Plain vanilla | | With covariates | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Implementation | MG | C&K MG | MG | C&K MG |
| Parameters estimated | 5 × N | 14 × N | 13 × N | 22 × N |
| **(a) Full sample (ANRR definition)** | 16.624 (4.630)*** | 7.692 (2.854)*** | 7.712 (3.647)** | 10.074 (3.651)*** |
| Observations | 2443 | 2443 | 2443 | 2443 |
| Countries (N) | 61 | 61 | 61 | 61 |
| Democratisations | 78 | 78 | 78 | 78 |
| Reversals | 42 | 42 | 42 | 42 |
| Avg Years in Dem | 19.6 | 19.6 | 19.6 | 19.6 |
| RMSE | 18.861 | 7.942 | 8.515 | 4.115 |
| **(b) Full sample excluding four Reversal-only Countries** | 18.646 (4.837)*** | 8.231 (2.895)*** | 8.746 (4.048)** | 11.970 (3.798)*** |
| Observations | 2294 | 2294 | 2294 | 2294 |
| Countries | 57 | 57 | 57 | 57 |
| Democratisations | 78 | 78 | 78 | 78 |
| Reversals | 38 | 38 | 38 | 38 |
| Avg Years in Dem | 19.5 | 19.5 | 19.5 | 19.5 |
| RMSE | 19.098 | 8.052 | 8.662 | 4.176 |
| **(c) Single democratisation without reversal** | 27.393 (6.456)*** | 4.655 (4.544) | 9.702 (5.972) | 13.636 (5.688)** |
| Observations | 1115 | 1115 | 1115 | 1115 |
| Countries | 28 | 28 | 28 | 28 |
| Democratisations | 28 | 28 | 28 | 28 |
| Reversals | 0 | 0 | 0 | 0 |
| Avg Years in Dem | 20.4 | 20.4 | 20.4 | 20.4 |
| RMSE | 18.557 | 7.874 | 8.735 | 4.269 |
| **(d) Single democratisation** | 19.487 (6.009)*** | 5.673 (3.538) | 8.462 (5.339) | 12.270 (5.097)** |
| Observations | 1675 | 1675 | 1675 | 1675 |
| Countries | 41 | 41 | 41 | 41 |
| Democratisations | 41 | 41 | 41 | 41 |
| Reversals | 14 | 14 | 14 | 14 |
| Avg Years in Dem | 18.8 | 18.8 | 18.8 | 18.8 |
| RMSE | 18.582 | 7.576 | 8.973 | 4.161 |
| **(e) Two or more democratisations** | 16.557 (8.079)** | 13.654 (4.511)*** | 9.088 (5.851) | 12.965 (4.690)*** |
| Observations | 619 | 619 | 619 | 619 |
| Countries | 16 | 16 | 16 | 16 |
| Democratisations | 37 | 37 | 37 | 37 |
| Reversals | 24 | 24 | 24 | 24 |
| Avg Years in Dem | 21.6 | 21.6 | 21.6 | 21.6 |
| RMSE | 20.430 | 9.218 | 7.758 | 4.217 |

*Notes*: The table presents Robust mean estimates from heterogeneous panel estimators using for the ANRR definition of democracy (see Table 1 for further details). The different result panels refer to different samples of 'treated' countries: all countries; excluding four 'reveral-only' countries; countries which experienced a Single democratisation event and no reversal; countries which experienced a Single democratisation event (but allowing for reversals); countries with two or more democratisation events (and reversals).

### *3.3. Multiple democratisations and reversals*

In Table 2 I focus on the ANRR definition of democracy and provide robust mean long-run estimates for several subsamples of countries. As before, I focus mainly on the Chan and Kwok (2022) Mean Group estimate in column (4) as my preferred implementation. Panel (a) provides the benchmark full sample result. In panel (b) I exclude the four countries which only reverted from democracy to autocracy but did not experience a democratisation event during the sample period (GMB, UGA, VEN, ZWE): the average long-run growth effect of democracy increases by almost two percentage points to 12%. These four countries are also excluded in all further models presented in this table. Panel (c) follows the spirit of PS and focuses on a subsample of 28 countries which experienced exactly one democratisation during the sample period (and no reversals to autocracy). The robust mean effect for this group of countries is now 13.6%. In Panel (d) I still prescribe a single democratisation but also allow for reversal, with the result that the magnitude of the average democracy coefficient for the 41 countries in this sample is only marginally lower than that in the previous panel. Finally, in panel (e) I only include those 16 countries which experienced two or more democratisations. Perhaps somewhat surprisingly, this yields the largest long-run democracy effect of 13%, despite an average of 1.5 reversals per country. It should be noted, though, that the average number of years spent in democracy at 21.6 years is highest in this subsample, higher even than in the single-democratisation sample in panel (c).

Overall, this analysis would seem to suggest that if we exclude 'pure' reversal cases — the four countries dropped in panel (b) — then the magnitude of the democracy-growth effect in the long-run is fairly stable, regardless of whether countries experienced a single or multiple democratisations, provided they still manage to spend substantial time 'in treatment'.

### 3.4. Sources of heterogeneity

The empirical exercises presented in the previous subsections suggest that allowing for parameter heterogeneity as well as dynamics and selection into democracy arrives at robust results for an average long-run 'democratic dividend' of around 10%–12%, depending on whether we include or exclude the 'pure reversal' cases. But are there any further insights beyond these *average* effects of democracy across (possibly) heterogeneous countries and time?

Before I investigate existing theories I illustrate the heterogeneity of 'treatment effects' by presenting the different democracy effects across the distribution: Table 3 indicates that lower quartile, median and upper quartile treatment effect estimates differ substantially, regardless of which definition of democracy or empirical implementation was adopted. For instance, using the preferred Chan and Kwok (2022) specification with covariates, the ANRR definition of democracy has a negative albeit insignificant lower quartile average, a median estimate of 11%, and an upper quartile estimate in excess of 30% higher income per capita.

Although there a significant differences in assumptions and implementation with the Chan and Kwok (2022) estimator, I also adopt the (Xu, 2017) generalised synthetic control approach for illustrative purposes: Appendix Figure F-1 charts the country-specific treatment effect evolution from a model with investment and openness, assuming four common factors. Even focusing the attention on the series in dark pink (for countries with a single democratic regime change), these results represent a kaleidoscope of differential economic experiences following regime change and a strong motivation for a heterogeneous democracy-growth nexus.

Turning to existing theories, I first study competing explanations in political science whereby democratisations can be distinguished as 'elite-biased' or 'popular' (Albertus and Menaldo, 2018),[27] or to have occurred 'by mistake' rather than intention (Treisman, 2020). Panels (a) and (b) of Fig. 2 study these two explanations, presenting simple mean estimates (by least squares, median and robust regression) of the heterogeneous treatment effects for the two groups, respectively. The estimated long-run coefficients[28] do not provide statistically significant evidence for lower long-run growth in 'elite-biased' democracies,[29] although the mean estimates do point in that direction. The analysis of 'democracy by mistake' confirms that countries in which democracy came about against the intentions but by the actions of the incumbents have higher democracy coefficients: one- and two-sided tests for equality of means reject this null at the 5% and 10% level, respectively. Perhaps the answer for this significant difference can be provided by a broader interpretation of the insignificant result in panel (a): a democracy where regime change came about by mistake is more likely to be somewhat removed from the power structures of the autocratic regime, and hence provides less scope for elite-capture and more incentives for the average citizen to try to 'make it' in the new era.

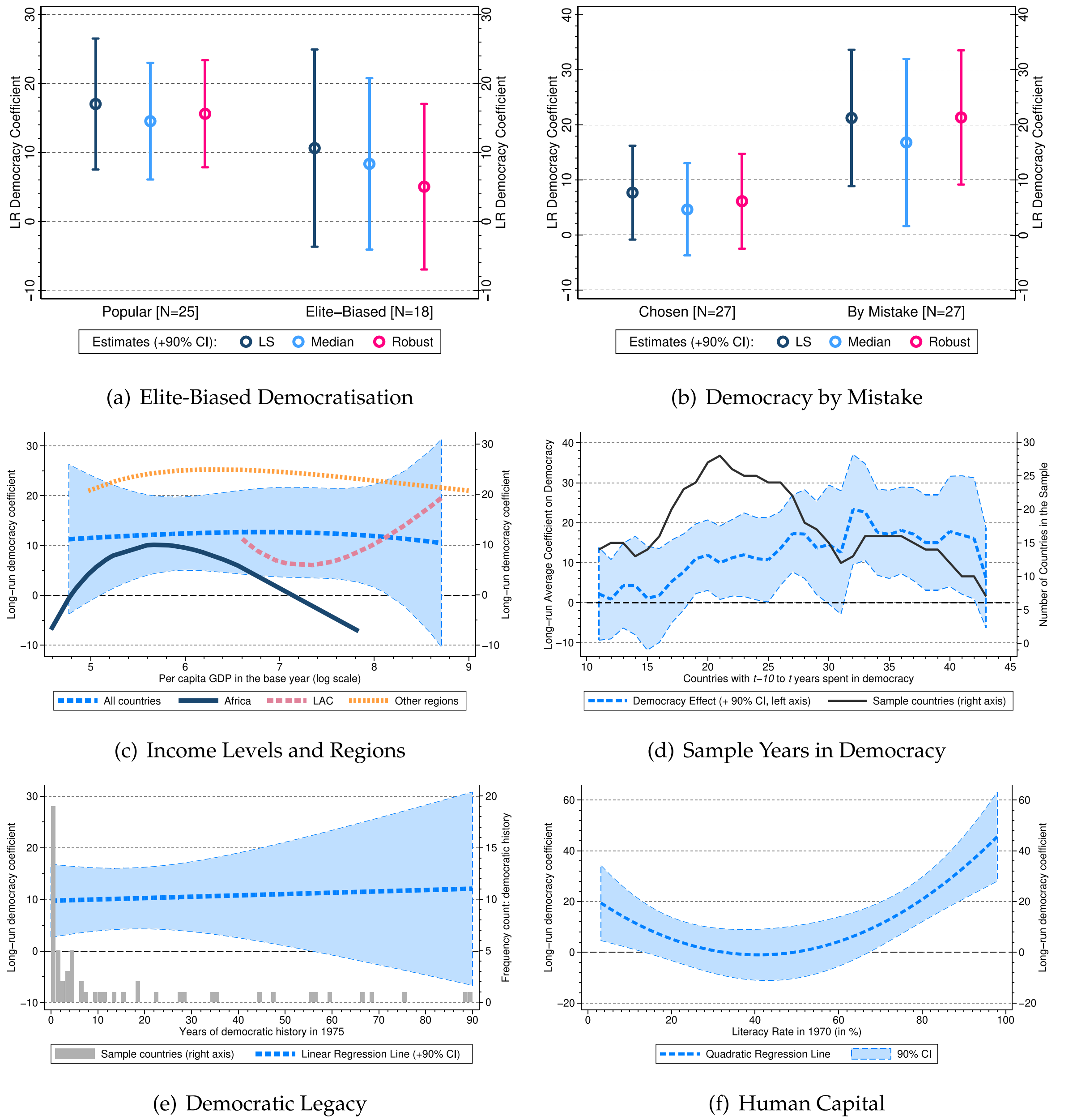
Second, I attempt to provide some insights into the cross-country heterogeneity of the long-run democracy estimates as well as initial conditions. I use fractional polynomial regressions of the country-specific long-run coefficient on base year per capita GDP (in logs), where the base year is the first sample year for each country; I provide this for the full sample as well as three geographical sub-samples (Sub-Saharan Africa, Latin America and the Caribbean, and Other regions) — for the full-country plot I omit the outlier countries with the largest and the smallest base-year GDP, respectively. Panel (c) of Fig. 2 shows the resulting plot for all countries in this sample using a dashed blue line and a shaded blue 90% confidence interval: although the regression line has a minimal hump, the wide confidence interval suggest that no matter whether countries were initially rich or poor, on average the long-run democracy coefficient is around 10%. But looking at three distinct geographical regions (Africa in blue, Latin America and the Caribbean (LAC) in red, and 'Other regions' in orange) yields very different patterns for countries with similar base year GDP (log values 6.6 to 8): low and declining for Africa, high and U-shaped for LAC, and high and stable/rising somewhat for the 'Other' regions. Note that the median share of years in democracy for these regions are 33% for Africa and almost exactly twice that for both LAC and 'others', so that at least for the latter two my simple analysis is not distorted by democratic experience. This analysis would not seem to support the notion that a democracy-growth nexus hinges on a certain minimum-level of income (Przeworski et al., 2000). The regional analysis points to substantial heterogeneity but should not be (mis)read as advocating geographic determinism.

Third, I assume that different length of 'treatment' (years spent in democracy) results in heterogeneous long-run estimates across countries, speaking to the 'experience' argument of Gerring et al. (2005). In panel (d) of Fig. 2 I estimate the robust mean democracy coefficient for different country groups, where group membership is defined by the number of years a country has spent in democracy. The band for each country group is arbitrarily set to eleven years, i.e. the first estimate is for all those countries which have spent between one and eleven years in democracy, the second for those with between two and twelve years, etc. — a strategy

[27] These authors suggest that up to two-thirds of new democracies in the 20th century were 'captured' by the pre-transition autocratic elite, building on constitutions designed by outgoing autocrats, and hence not only were "not *for* the people... [but] also not *of* or *by* the people" (Albertus and Menaldo, 2018, 7, emphases in original).

[28] In this and all of the below analysis I adopt the country-specific long-run democracy estimates from the above preferred regression model in Table 1, Panel (a), Column (4).

[29] One-sided or two-sided $t$-tests cannot reject the null of no difference in means.




(a) Elite-Biased Democratisation

(b) Democracy by Mistake

(c) Income Levels and Regions

(d) Sample Years in Democracy

(e) Democratic Legacy

(f) Human Capital

**Fig. 2.** Analysing Heterogeneity (Long-run Coefficients). *Notes*: Unless indicated the long-run democracy estimates employed are those from Table 1, top panel for ANRR, column (4). Panel (a) compares the mean democracy coefficient for elite-biased and 'popular' democratisation (Albertus and Menaldo, 2018) using least squares, median and robust regression. Panel (b) does the same for Treisman's (2020) 'democracy by mistake'. Panel (c) highlights the cross-country heterogeneity of the democracy-growth nexus by correlating the estimated coefficients with base year per capita GDP for all countries (blue dashed line and 90% CI) and for three geographic regions. Africa has the lowest experience of democracy (median 33% of observations), but LAC and the 'Other' category are very similar in this regard at around 67%. The figure in Panel (d) presents the robust mean estimates for the long-run effect of democracy on income per capita for different samples, based on the number of years a country has spent in democracy. The left-most estimate with value 11 on the $x$-axis is for countries which have spent between 1 and 11 years in democracy, the value 12 is for 2 to 12 years, etc. The shaded area is a 90% confidence interval around a mean estimate (CI based on nonparametric standard errors following Pesaran and Smith, 1995). Since the distribution of years in democracy is not uniform the black line indicates how many countries are contained in each specific regression. Panel (e) takes Polity IV data to plot the long-run democracy coefficient against the years spent in democracy *prior to 1975* – 19 countries enter with no history of democracy. For ease of illustration I exclude one country with a democratic legacy of 121 years. Panel (f) adopts the 1970 literacy rates from Madsen et al. (2015) and fits a quadratic regression line (with 90% CI) to the democracy coefficient-literacy data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**
Main results—Dynamic specifications—Quartile TE.

| | Plain vanilla | | With covariates | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Implementation | MG | C&K MG | MG | C&K MG |
| **(a) Democracy (ANRR)** | | | | |
| Robust mean | 16.624 | 7.692 | 7.712 | 10.074 |
| | (4.630)*** | (2.854)*** | (3.647)** | (3.651)*** |
| Lower quartile | −1.713 | −4.156 | −6.006 | −2.591 |
| | (7.708) | (3.257) | (6.725) | (4.963) |
| Median | 15.299** | 7.067* | 8.209 | 10.872* |
| | (5.430) | (3.445) | (4.219) | (4.381) |
| Upper quartile | 40.153*** | 22.403*** | 26.525*** | 31.099*** |
| | (7.980) | (5.427) | (7.031) | (5.551) |
| Countries (N) | 61 | 61 | 61 | 60 |
| **(b) Democracy (BMR)** | | | | |
| Robust mean | 15.130 | 7.322 | 9.983 | 11.118 |
| | (4.057)*** | (3.279)** | (2.843)*** | (3.612)*** |
| Lower quartile | −1.679 | −5.540 | −2.758 | −7.200 |
| | (6.679) | (5.052) | (3.752) | (6.339) |
| Median | 16.858*** | 6.727 | 8.879* | 9.092 |
| | (4.787) | (3.609) | (3.630) | (4.954) |
| Upper quartile | 34.598*** | 21.814*** | 24.753*** | 30.344** |
| | (7.241) | (5.898) | (4.163) | (9.074) |
| Countries (N) | 55 | 55 | 55 | 55 |
| **(c) Democracy (CGV)** | | | | |
| Robust mean | 14.405 | 4.989 | 10.154 | 6.934 |
| | (4.572)*** | (4.059) | (3.190)*** | (4.179)* |
| Lower quartile | −2.684 | −12.480 | −4.832 | −9.935 |
| | (7.293) | (6.651) | (7.351) | (7.200) |
| Median | 15.505** | 5.354 | 8.457 | 5.499 |
| | (5.161) | (4.289) | (4.363) | (4.908) |
| Upper quartile | 33.493** | 19.371* | 27.378*** | 22.390* |
| | (9.585) | (8.206) | (5.227) | (8.811) |
| Countries (N) | 49 | 49 | 49 | 49 |
| **(d) Democracy (PS)** | | | | |
| Robust mean | 27.915 | 4.749 | 11.501 | 11.936 |
| | (4.813)*** | (4.594) | (4.445)*** | (5.072)** |
| Lower quartile | 12.520 | −9.936 | −0.712 | −6.512 |
| | (10.037) | (7.126) | (7.592) | (9.023) |
| Median | 28.962*** | 7.235 | 10.278 | 10.715 |
| | (5.987) | (6.239) | (5.452) | (5.946) |
| Upper quartile | 46.503*** | 22.478** | 27.539** | 33.168*** |
| | (9.618) | (7.726) | (8.243) | (7.388) |
| Countries (N) | 41 | 41 | 41 | 41 |

*Notes*: The table presents Robust mean estimates from heterogeneous panel estimators using different definitions of democracy: (1) and (3) simple Mean Group estimator, (2) and (4) Chan and Kwok (C&K) DID Mean Group estimator — all are estimated using least squares. We hold the sample fixed across the four specifications, but not when comparing different definitions of democracy. All estimates presented are long-run (ATET) estimates for the causal effect of democracy on income per capita (in percent), derived from a CS-DL model (Chudik et al., 2016). The models in (3) and (4) include gross investment ratio and trade/GDP as additional covariates. The four alternative democracy dummies are by Acemoglu et al. (2019) – ANRR, Boix et al. (2013) – BMR, Cheibub et al. (2010) – CGV, and Papaioannou and Siourounis (2008) – PS.

which artificially increases the number of observations (long-run democracy estimates) in each of the constituent regressions. The maximum year $t$ of each band is printed along the $x$-axis of the plot, the implied minimum number of years is simply $t-10$. The dashed blue line represents the robust mean estimate for the effect of democracy on growth across bands (left scale), the blue shaded area the 90% confidence interval.[30] Using a band of eleven years leads to different sample sizes as the in-sample democracy

[30] This is constructed from the robust mean estimates in each band following Pesaran and Smith (1995) as in the main results in Section 3.2 above.

experience increases, and I therefore indicate the sample size with the black solid line (right scale). This analysis reveals the slow emergence of a democratic dividend, with a positive significant effect taking around 20 years (after transition) to manifest itself with statistical significance; the effect plateaus sometime after 30 years, but the sample size changes too much in the final years to make a convincing claim about stability or decline. Nevertheless, the profile *appears* closer to a concave than a linear relationship, which implies that democracy has a one-off levels effect (in line with the assumptions in ANRR) and not a perpetual growth effect.

Fourth, instead of studying the in-sample experience of democracy, I gauge the significance of a democratic legacy since 1800, proxied by the number of years spent in democracy by 1975 — this cut-off maximises the data availability in the Polity IV dataset.[31] Gerring et al. (2005), among others, argue that political regimes are historical legacies, with cumulative effects of institutions (only) coming to bear over long time horizons. Panel (e) of Fig. 2 shows a fitted linear regression line[32] for the relationship between the long-run democracy coefficient and democratic legacy in years (blue line, shaded 90% CI), together with a histogram for the latter variable. There is no clear advantage or disadvantage of democratic legacy for the 'treated' countries which transitioned into or out of democracy during the sample period. Having said that, as the histogram indicates there is a mere sprinkling of countries with legacies in excess of twenty years by 1975. A second, perhaps more meaningful, conclusion from this exercise is that the 19 countries with no democratic legacy have a statistically significant long-run effect of democracy around 10% — hence no different from the average sample effect.[33]

In panel (f) I study the relationship between human capital endowment and the long-run democracy coefficient (e.g. Glaeser et al., 2004): to maximise coverage I adopt the literacy data for 1970 from Madsen et al. (2015), which still represents a reasonably early 'base-year' observation. I fit a quadratic regression line along with its 90% confidence interval to reveal some evidence for a convex relationship. However, while positive significant growth effects of democracy at low levels of literacy have overlapping confidence intervals with intermediate levels indicating zero growth effects, the graph suggests that countries with initially high rates of literacy were able to extract on average a higher democratic dividend.

## 4. Robustness

In this Section I investigate how robust my main findings are to changes in the sample makeup and compare the 'performance' to that of the seminal ANRR study. I motivate two exercises: one focused on the number of time series observations available in each country, and a second on the time period covered by the sample.

I first drop countries by their observation count, $T_i$: having few(er) observations potentially over-emphasises individual shocks to the economy and arguably makes it harder to empirically capture the *long-run equilibrium relationship*. Since fewer observations on average also means fewer observations *in democracy*,[34] the long-run estimate is in effect an *extrapolated* effect of democracy, given that the median country in my treated sample merely experienced a spell of twelve years in democracy.[35] Although I adopt a dynamic model to capture the dip in economic performance observed before and in the immediate aftermath of democratisation (see ANRR and PS), the dynamics may be misspecified and hence the long-run effect is potentially underestimated (if the dynamics are not captured sufficiently) or overestimated (if more elaborate dynamics translate into less precise estimates given the limited time series data). Studying the evolution of estimates as the sample is restricted to countries with larger and larger minimum observation counts should go some way to address these concerns.

I then shift my attention to restricting the sample by moving the end year of analysis: my data, taken from ANRR, covers 1960 to 2010, and hence includes the Global Financial Crisis (GFC) — the most significant global macroeconomic shock since the 1930s — as part of the final sample years. Although the impact of the GFC was substantial, it was by no means uniform across countries. The same could be said for the post-crisis recovery. Recent work by Young (2022) and Broderick et al. (2020) has highlighted the fragility of statistical inference in many applications which often rests on a mere handful of observations.[36] It is straightforward to develop an argument whereby some autocracies like China or Viet Nam (included in the control sample) for reasons other than political regime were substantially less affected by the crisis than economies included in our treated sample of democratisers, such as South Korea, Chile, or, most notably, Greece. If the economic shock is substantial then my long-run estimates may not adequately capture the equilibrium relationship and hence under-estimate the democratic dividend. Conversely, the effect may be overestimated if the bounce-back from the crisis was systematically swifter and/or more substantial in democracies than autocracies and this 'spike' at the end of the sample may have tilted the fitted regression line upwards. In order to guard against either possibility, I systematically restrict the sample by shifting the end year forward one year at a time.

---

31 Around 20% of the observations used to derive the long-run democracy effects in this exercise are from 1961–75.

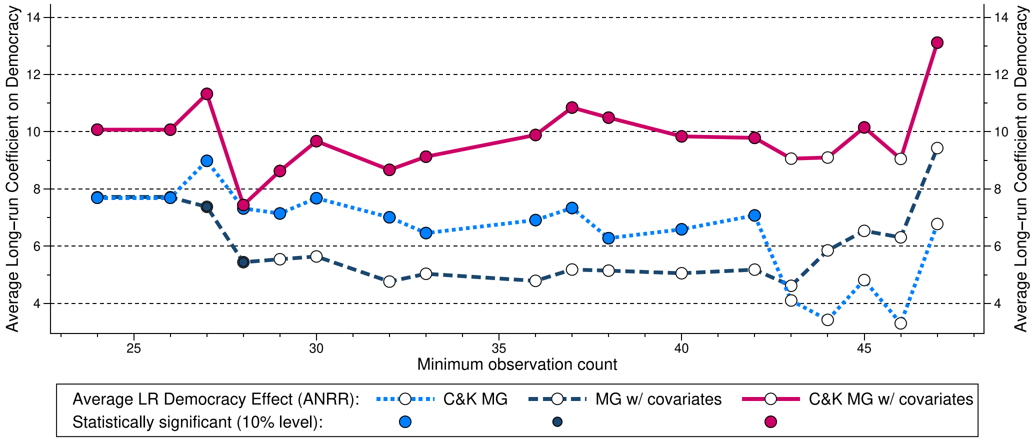
32 A fractional polynomial points to this linear relationship but has a wider confidence interval in the right (but not the left) tail. I adopt the linear fit for ease of presentation.

33 These countries are predominantly Sub-Saharan African ex-colonies. Further note that for the 27 out of 38 countries in the 'control group' (never democratic during the sample period) for which these 1975 data are available in PolityIV, 20, equivalent to three-quarters, have no democratic legacy. In my exercise for economies transitioning in or out of democracy the 'no legacy' countries amount to one-third of the sample.

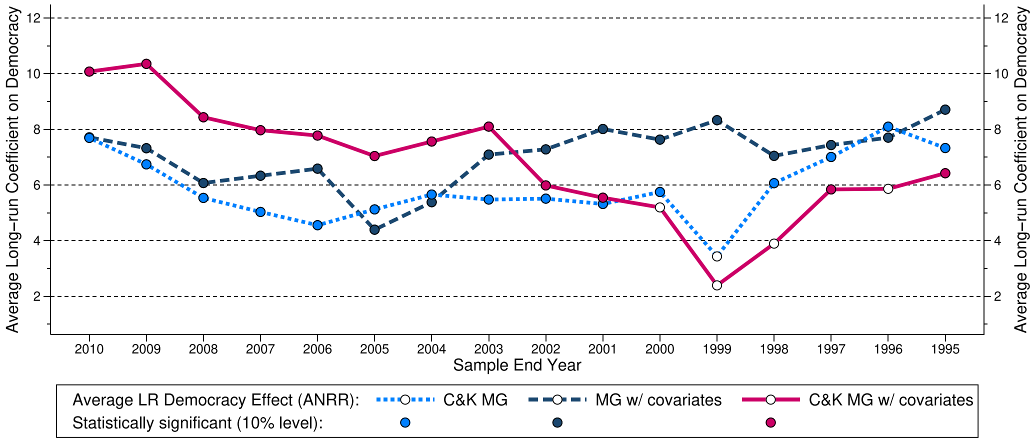
34 The ten countries with $T_i$ of 26 or 27 have a median of 17 years in democracy, for the nine countries with $T_i$ between 40 and 44 and the 24 countries with $T_i$ of 47 the medians are 19 and 24, respectively.

35 This analysis is based on the spell data presented in Fig. 1 Panel (b).

36 In Appendix Table E-3 I employ a naive and *ad hoc* procedure to find that ANRR's AB, HHK and IV results turn insignificant when the observations from three or four (predominantly central Asian) countries are omitted from the sample, constituting between 0.78% and 0.99% of the full sample. Of course I cannot rule out that dropping further countries using a similar *ad hoc* approach will not restore statistical significance.




(a) Sample reduction by $T_i$ count




(b) Sample reduction by end year

**Fig. 3.** Sample Reduction Exercises — Heterogeneous Diff-in-Diff Models. *Notes*: The figure presents the robust mean estimates for a variety of heterogeneous Difference-in-Difference estimators as the regression sample is constrained, using the minimum count of country observations as the selection mechanism in panel (a) and the sample end year in panel (b). The unconstrained sample is made up of a maximum of 61 economies which transitioned into democracy at least once during the sample period. The estimates for the Chan and Kwok (CK) approaches further build on the information contained in a sample of 42 countries which *never* experienced democracy during the sample period. A filled (white) marker indicates that the coefficient on democracy is statistically (in)significant at the 10% level. 'MG' presents results for models which ignore (strong) cross-section correlation and/or uncommon pre-democratisation trends; 'C&K MG w/ covariates' presents results for a model including country observations for gross investment and trade as covariates to the ANRR democracy dummy and the various cross-section averages detailed in the text; 'C&K MG' only includes democracy as observed regressand alongside cross-section averages as detailed in the text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### *4.1. Sample reduction by minimum observation count*

Panel (a) of Fig. 3 presents the results from dynamic specifications of three heterogeneous parameter models for the first sample reduction exercise by country observation count, adopting the ANRR definition of democracy — the plots for the BMR, CGV and PS alternatives are provided in Appendix D. In this and the plot in panel (b) a filled (hollow) circle indicates statistically (in-)significant difference from zero at the 10% level, the *x*-axis reports the minimum observation count $T_i$ for inclusion in the sample and the *y*-axis the average long-run democracy coefficient (in percent). Table 4 reports the estimates and sample characteristics for the full sample, the sample when the long-run democracy coefficient turns insignificant, and the balanced panel sample in all of the four definitions of democracy adopted in this study — an equivalent table for the ANRR parametric estimators is relegated to Appendix E to safe space.

The estimates from the empirical model ignoring any potential factor structure and thus selection, uncommon trends and/or common shocks with heterogeneous impact (in short teal-coloured dashed) demonstrate comparatively little robustness, given that the democracy effect turns insignificant at $T_i = 29$ when around 13% of observations (for 12 of 61 countries) are dropped. The

**Table 4**
Sample reduction exercises.

| Definition | Sample reduction by $T_i$ count | | | | Sample reduction by end year | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) ANRR | (2) BMR | (3) CGV | (4) PS | (5) ANRR | (6) BMR | (7) CGV | (8) PS |
| **Panel A: Full sample estimate** | | | | | | | | |
| Long-run effect | 10.074 | 11.118 | 6.934 | 11.936 | 10.074 | 11.118 | 6.934 | 11.936 |
| | (3.651)*** | (3.612)*** | (4.179)* | (5.072)** | (3.651)*** | (3.612)*** | (4.179)* | (5.072)** |
| min $T_i$/End year | 24 | 23 | 24 | 24 | 2010 | 2007 | 2008 | 2010 |
| Countries | 61 | 55 | 50 | 41 | 61 | 55 | 50 | 41 |
| Observations | 2,443 | 2,051 | 1,922 | 1,670 | 2,443 | 2,051 | 1,922 | 1,670 |
| Share of full sample | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **Panel B: Estimate insignificant (10% significance level)** | | | | | | | | |
| Long-run effect | 9.060 | 8.213 | 6.084 | 9.222 | 5.194 | 4.715 | 5.540 | 5.760 |
| | (5.551) | (5.602) | (4.014) | (6.147) | (3.220) | (3.135) | (3.520) | (5.116) |
| min $T_i$/End year | 43 | 40 | 25 | 40 | 1999 | 1999 | 2007 | 2002 |
| Countries | 32 | 29 | 49 | 27 | 46 | 42 | 50 | 33 |
| Observations | 1,488 | 1,262 | 1,898 | 1,239 | 1,546 | 1,402 | 1,875 | 1,205 |
| Share of full sample | 0.61 | 0.62 | 0.99 | 0.74 | 0.63 | 0.68 | 0.98 | 0.72 |
| **Panel C: Balanced panel estimate** | | | | | | | | |
| Long-run effect | 13.113 | 10.036 | 10.420 | 9.649 | n/a | n/a | n/a | n/a |
| | (6.088)** | (6.591) | (5.795)* | (7.719) | | | | |
| min $T_i$/End year | 47 | 44 | 45 | 47 | | | | |
| Countries | 24 | 23 | 22 | 19 | | | | |
| Observations | 1,128 | 1,012 | 990 | 893 | | | | |
| Share of full sample | 0.46 | 0.49 | 0.52 | 0.53 | | | | |

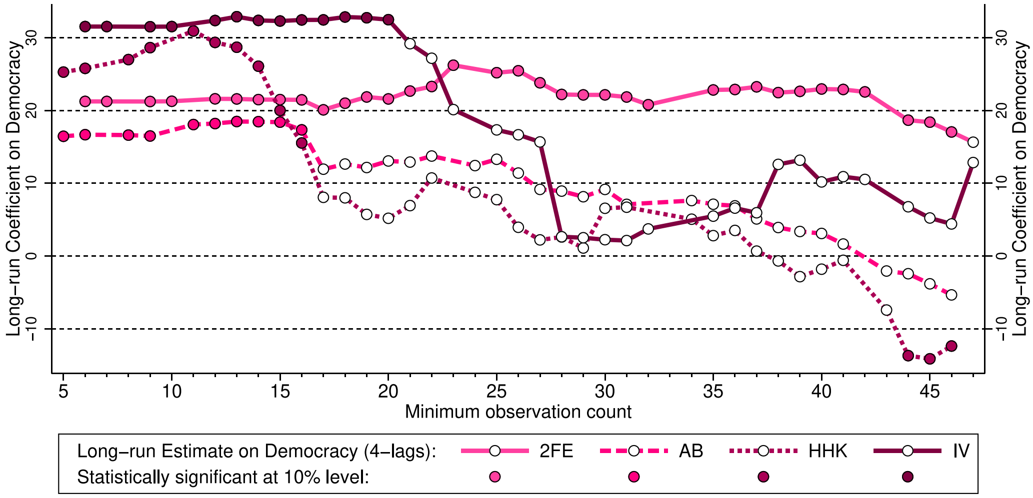
*Notes*: The table presents estimates for the two sample reduction exercises in columns (1)-(4) and (5)-(8), respectively (definition of democracy as indicated). All estimates the robust long-run coefficient for democracy from the Chan and Kwok (2022) model with additional covariates (standard errors are computed via the Delta method). Results in Panel A are identical to those in column (4) of Table 1 above. Full estimates for ANRR are presented in Fig. 3 and for all other definitions of democracy in Appendix Figs. D-1 and D-2. The sample end year reduction strategy in columns (5)–(8) does not lead to a balanced panel like the sample reduction by minimum observation count in columns (1)–(4). Statistical significance at the 10%, 5% and 1% level are indicated as *, **, and ***, respectively.

plain vanilla Chan and Kwok (2022) model accounting for these distortions (in short blue dashes) yields more stable long-run estimates at around 6%–8% until the coefficient turns insignificant when $T_i = 43$. The model including additional covariates (gross investment ratio and trade openness) lifts the estimate somewhat to around 10% throughout the sample reduction exercise. Statistical insignificance first occurs again when $T_i = 43$, where almost 40% of sample observations have been dropped, though the coefficient magnitude is still stable.

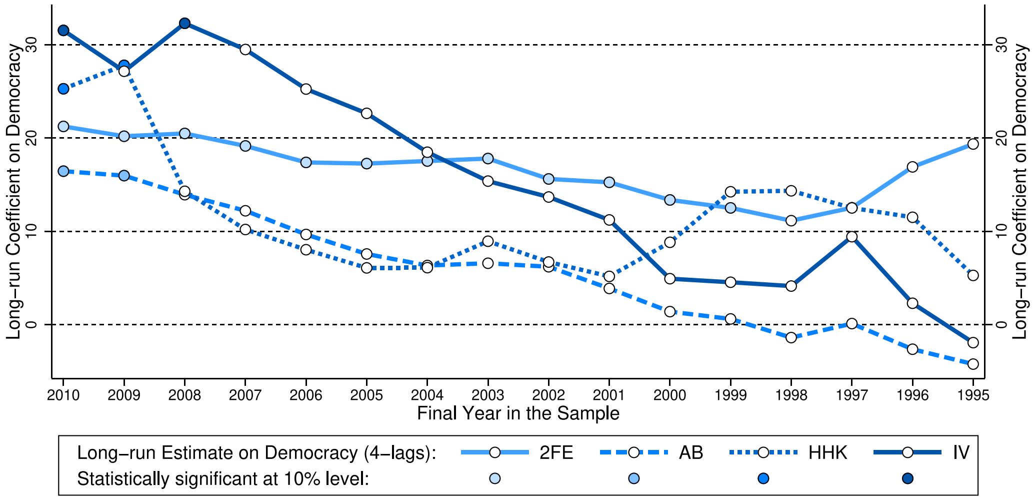
Columns (2)-(4) in Panel (B) of Table 4 report a similar level of robustness for the results when adopting the alternative definitions of BMR and PS for democracy: for BMR the democracy coefficient turns insignificant when 38% of observations are dropped, for PS the figure is 26%. The CGV results turn insignificant when fewer than 1% of observations are omitted — however, the graph in Panel (b) of Appendix Figure D-1 charting the step-by-step results for this definition indicates that this is an anomaly, given that further sample restriction yields stable and statistically significant long-run democracy estimates until one-third of observations are omitted. Balanced panel results for all these alternative definitions, reported in Panel (C) of the same table, yield long-run estimates of around 10% (statistically significant for ANRR and CGV definitions): dropping 50% of the sample still yields qualitatively similar results for my empirical implementation.

Panel (a) of Fig. 4 reports the results for the same sample reduction exercise using the four principle parametric estimators in ANRR.[37] Focusing on the 2SLS results (solid dark pink line) long-run democracy estimate turns statistically insignificant when fewer than 7% of observations are omitted (the coefficient merely drops from 31.5% to 29%) — a histogram provided in Appendix E indicates that this does not affect the number of democratisation events included in the sample in any substantial way. Estimates for further sample restrictions swiftly drop to long-run estimate between 5 and 15% (all insignificant). The balanced panel IV estimate (for a 42% sample reduction) has a democracy estimate of 13% with a standard error almost twice that size. The other implementations, with the exception of the two-way fixed effects estimator, demonstrate very similar patterns, although their coefficient magnitudes collapse and turn negative in the balanced panel results.

[37] In Appendix E I report the findings when I carry out the same sample reduction exercises for a number of implementations in Madsen et al. (2015), including the preferred IV specifications. The Madsen et al. (2015) results, which are derived from decadal data for 1820–2000, are much more robust to this form of sample reduction: the benchmark estimate of 96% higher income per capita for a standard deviation increase in the continuous democracy measure drops to an insignificant 60% when over one-third of observations are omitted, alternative IV specifications turn insignificant when 13% and 26% of observations are dropped, respectively.




(a) Sample reduction by $T_i$ count




(b) Sample reduction by end year

**Fig. 4.** Sample Reductions — ANRR (Parametric) Models. *Notes*: The figure presents the long-run estimates for democracy from varying empirical samples for the 2FE, AB, HHK and 2SLS estimators, computed as $\hat{\beta}^{LR} = \hat{\beta}/(1 - \sum_{\ell=1}^{4} \hat{\rho}_{i,t-\ell})$, where $\hat{\beta}$ is the estimate on the democracy dummy and the $\hat{\rho}$ are estimates for the lags of per capita GDP (standard errors are constructed via the Delta method). A filled (white) circle marker indicates that the long-run coefficient is statistically (in)significant at the 10% level. All estimates are for the specification with four lags of GDP (and four lags of the instrument for 2SLS) preferred by ANRR. Alternative specifications yield qualitatively identical results (available on request). The 'left-most' estimates replicate the results in ANRR's Table 2, column (3) for 2FE, (7) for AB, and (11) for HHK, and Table 6, column (2) Panel A for 2SLS. In Panel (a) the $x$-axis indicates the minimum number of observations required to be included in the sample, in Panel (b) it indicates the end year included in the sample. In panel (a) the 2FE, AB, HHK and IV estimates turn statistically insignificant when 41%, 5%, 5% and 7% of country-observations are excluded. In panel (b) the equivalent figures are 30%, 25%, 5% and 3%.

### *4.2. Sample reduction by sample end year*

Panel (b) of Fig. 3 presents the mean heterogeneous Diff-in-Diff results when observations are omitted by sample end year. In this graph for the ANRR democracy dummy, and in the equivalent graphs in Appendix D for BMR, CGV and PS definitions, the $x$-axis is in *reverse* chronological order. Here, the standard MG estimate of the long-run democracy effect (in dark blue dashes) is remarkably stable and remains statistically significant throughout the years displayed. With one exception in 1999 the same is true for the 'plain vanilla' Chan and Kwok (2022) implementation (in light blue dashes). This is somewhat surprising, given that the average number of years spent in democracy substantially *declines* as the sample end year is moved further and further back in time. The preferred Chan and Kwok (2022) estimator with additional covariates (in dark pink) follows a more logical pattern of a declining magnitude for the democracy effect as the sample is curtailed. It turns statistically insignificant in the sample restricted to 1960–1999, when one-third of the full sample observations are discarded. For the BMR and PS definitions of democracy, the same happens when 32% and 28% of observations are omitted, with end years 1999 and 2002, respectively. For the CGV definition the omission of the final year for which these data are available, 2008, turns the marginally significant 7% long-run effect into an insignificant 5.6%.

Comparing these patterns to the findings for the same exercise in the ANRR models presented in Panel (b) of Fig. 4 highlights the comparative robustness of my estimates. Omitting a single year, 2010, from the analysis in ANRR, equivalent to fewer than 3% of observations, turns the preferred IV estimate statistically insignificant. Although this estimate returns to statistical significance when 2009 is also excluded, any further restrictions show a declining coefficient which is always insignificant. The two-way fixed effects results excepted, the alternative implementations of ANRR again do not improve much on this finding: the HHK estimate is insignificant with fewer than 5% of observations omitted, though for the AB estimate this figure is a more substantial 25%.[38]

Like in the first sample reduction exercise, my results are remarkably stable, with the exception of the models adopting the CGV definition of democracy. This robustness is particularly marked compared with the collapse of significance in the ANRR implementations.

## 5. Concluding remarks

In this paper I motivated the idea that the 'democratic dividend', the long-run growth effect from democratisation, is likely to differ across countries, and provided empirical analysis of the 'heterogeneous democracy-growth nexus' using novel difference-in-difference estimators. My findings do not challenge the recently-emerging consensus that 'democracy causes growth', but it qualifies the magnitude of the average income effect of democratic regime change to be much more modest than the 20%–30% found in a recent paper by ANRR. My empirical results are robust to substantial changes to the sample, whereas these authors' estimates turn insignificant when a mere 3%–7% of observations are omitted. Further analysis cautions that, as so often, realities on the ground for individual countries may look less straightforward: it would seem that not all countries benefit equally from a 'democratic dividend', and as is previously acknowledged in the literature, the *mode* of democratic regime change as well as prior human capital endowment may play important roles in determining the magnitude of this dividend.

In addition to further exploring these specific factors, future work building on the finding of a heterogeneous democracy-growth nexus should pay closer attention to the relationship with 'length of treatment', as well as the definition and underlying 'building blocks' of the concept of democracy adopted in the empirical analysis: alongside the idea of electoral democracy captured in all measures of democracy considered here, some definitions, in particular the Freedom House Index, ignore executive constraints but include notions of civil liberties and equality before the law ('rule of law'); others, like the PolityIV polity2 index ignore the rule of law but incorporate executive constraints (see Appendix Figure A-1). Electoral democracy on its own, as captured in the Boix et al. (2013) and Cheibub et al. (2010) indicators, appears to yield similar economic effects of regime change (at least in the former) to the ANRR definition, which combines polity2 and the Freedom House Index. In the past, the economics literature has put significant emphasis on 'institutions' in the determinants of long-run growth (e.g. Acemoglu et al. 2001, Glaeser et al. 2004, and in particular Rodrik et al.'s 2004 'institutions rule'), adopting measures of rule of law or executive constraints as empirical proxies. This begs the question why definitions of democracy accommodating such concepts of institutional change do not appear to yield significantly higher economic benefits. ANRR talk of their 'meta' indicator as primarily capturing qualities of electoral democracy, but this clearly ignores the rule of law element in the Freedom House Index meshed into their indicator. ANRR (and previously Papaioannou and Siourounis, 2008) combine indices from different sources, based on different data collection and aggregation practices. A more meaningful strategy to investigate the important institutional building blocks of the positive democracy-growth nexus could adopt the Varieties of Democracy (V-Dem) 'liberal democracy' index (Coppedge et al., 2021) and its hierarchical components to determine which underlying institutions matters (most) for economic prosperity — I aim to address this question in ongoing research (Boese and Eberhardt, 2022).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.euroecorev.2022.104173.

[38] Madsen et al.'s (2015) benchmark specification is again remarkably stable and does not turn statistically insignificant in the periods considered in Appendix Figure E-3, Panel (b). The two alternative implementations however yield substantially reduced coefficients which are insignificant when the decadal observations for 2000 are removed (11% and 17% of all observations, respectively).

## References


Acemoglu, D., Johnson, S., Robinson, J.A., 2001. The colonial origins of comparative development: An empirical investigation. Amer. Econ. Rev. 91 (5), 1369–1401.

Acemoglu, D., Naidu, S., Restrepo, P., Robinson, J.A., 2015. Democracy, redistribution, and inequality. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, Vol. 2. Elsevier, pp. 1885–1966.

Acemoglu, D., Naidu, S., Restrepo, P., Robinson, J.A., 2019. Democracy does cause growth. J. Polit. Econ. 127 (1), 47–100.

Aghion, P., Akcigit, U., Howitt, P., 2014. What do we learn from Schumpeterian growth theory? In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth, Vol. 2. Elsevier, pp. 515–563.

Albertus, M., Menaldo, V., 2018. Authoritarianism and the Elite Origins of Democracy. Cambridge University Press.

Andrews, D.W., 2005. Cross-section regression with common shocks. Econometrica 73 (5), 1551–1585.

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

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., Khosravi, K., 2021. Matrix completion methods for causal panel data models. J. Amer. Statist. Assoc. 116 (536), 1716–1730.

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

Barro, R.J., 1991. Economic growth in a cross section of countries. Q. J. Econ. 106 (2), 407–443.

Barro, R.J., 1996. Democracy and growth. J. Econ. Growth 1 (1), 1–27.

Baum, M.A., Lake, D.A., 2003. The political economy of growth: Democracy and human capital. Am. J. Political Sci. 47 (2), 333–347.

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

Boese, V.A., Eberhardt, M., 2022. Which Institutions Rule? Unbundling the Democracy-Growth Nexus. V-Dem Working Paper 131, Varieties of Democracy Institute.

Boix, C., Miller, M., Rosato, S., 2013. A complete data set of political regimes, 1800–2007. Comp. Political Stud. 46 (12), 1523–1554.

Boucekkine, R., Desbordes, R., Melindi-Ghidi, P., 2021. A theory of elite-biased democracies. Math. Social Sci. 112, 159–166.

Broderick, T., Giordano, R., Meager, R., 2020. An automatic finite-sample robustness metric: Can dropping a little data change conclusions? arXiv:2011.14999.

Cervellati, M., Sunde, U., 2014. Civil conflict, democratization, and growth: Violent democratization as critical juncture. Scand. J. Econ. 116 (2), 482–505.

Chan, M.K., Kwok, S.S., 2022. The PCDID approach: Difference-in-differences when trends are potentially unparallel and stochastic. J. Bus. Econom. Statist. 40 (3), 1216–1233.

Cheibub, J.A., Gandhi, J., Vreeland, J.R., 2010. Democracy and dictatorship revisited. Public Choice 143 (1–2), 67–101.

Chudik, A., Mohaddes, K., Pesaran, M.H., Raissi, M., 2016. Long-run effects in large heterogeneous panel data models with cross-sectionally correlated errors. In: Gonzalez-Rivera, G., Hill, R.C., Lee, T.-H. (Eds.), Advances in Econometrics: Essays in Honour of Aman Ullah. Emerald Group Publishing Limited, pp. 85–135.

Chudik, A., Pesaran, M.H., 2015. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. J. Econometrics 188 (2), 393–420.

Comin, D., Hobijn, B., 2004. Cross-country technology adoption: Making the theories face the facts. J. Monetary Econ. 51 (1), 39–83.

Coppedge, M., Gerring, J., Knutsen, C.H., Lindberg, S.I., Teorell, J., Altman, D., Bernhard, M., Cornell, A., Fish, M.S., Gastaldi, L., et al., 2021. V-Dem Dataset v11. Technical report, http://dx.doi.org/10.23696/vdemds21.

De Visscher, S., Eberhardt, M., Everaert, G., 2020. Estimating and testing the multicountry endogenous growth model. J. Int. Econ. 125, 103325.

Dodsworth, S., Ramshaw, G., 2021. Democracy's development dividend. J. Democracy 32 (1), 126–138.

Durlauf, S.N., 2020. Institutions, development, and growth: Where does evidence stand? In: Baland, J.-M., Bourguignon, F., Platteau, J.-P., Verdier, T. (Eds.), The Handbook of Economic Development and Institutions. Princeton University Press, pp. 189–217.

Durlauf, S.N., Johnson, P.A., Temple, J.R., 2005. Growth econometrics. In: Handbook of Economic Growth, Vol. 1. Elsevier, pp. 555–677.

Eberhardt, M., Teal, F., 2011. Econometrics for grumblers: A new look at the literature on cross-country growth empirics. J. Econ. Surv. 25 (1), 109–155.

Funke, M., Schularick, M., Trebesch, C., 2020. Populist Leaders and the Economy. CEPR Discussion Paper No. 15405.

Gerring, J., Bond, P., Barndt, W.T., Moreno, C., 2005. Democracy and economic growth: A historical perspective. World Politics 57 (3), 323–364.

Giavazzi, F., Tabellini, G., 2005. Economic and political liberalizations. J. Monetary Econ. 52 (7), 1297–1330.

Glaeser, E.L., La Porta, R., Lopez-de Silanes, F., Shleifer, A., 2004. Do institutions cause growth? J. Econ. Growth 9 (3), 271–303.

Gobillon, L., Magnac, T., 2016. Regional policy evaluation: Interactive fixed effects and synthetic controls. Rev. Econ. Stat. 98 (3), 535–551.

Hamilton, L.C., 1992. How robust is robust regression? Stata Techn. Bull. 1 (2).

Helliwell, J.F., 1994. Empirical linkages between democracy and economic growth. Br. J. Political Sci. 225–248.

Knutsen, C.H., 2013. Democracy, state capacity, and economic growth. World Dev. 43, 1–18.

Madsen, J.B., Raschky, P.A., Skali, A., 2015. Does democracy drive income in the world, 1500–2000? Eur. Econ. Rev. 78, 175–195.

Marschak, J., Andrews, W.H., 1944. Random simultaneous equations and the theory of production. Econometrica 12 (3/4), 143–205.

Marshall, M.G., Gurr, T.R., Jaggers, K., 2017. Polity IV project: Political regime characteristics and transitions, 1800-2016. Available at http://www.systemicpeace.org/inscrdata.html.

Murtin, F., Wacziarg, R., 2014. The democratic transition. J. Econ. Growth 19 (2), 141–181.

Papaioannou, E., Siourounis, G., 2008. Democratisation and growth. Econom. J. 118 (532), 1520–1551.

Persson, T., Tabellini, G., 2009. Democratic capital: The nexus of political and economic change. Am. Econ. J.: Macroecon. 1 (2), 88–126.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74 (4), 967–1012.

Pesaran, M.H., Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. J. Econometrics 68 (1), 79–113.

Phillips, P.C., Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. Econom. J. 6 (1), 217–259.

Przeworski, A., Alvarez, M., Cheibub, J.A., Limongi, F., 2000. Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990, Vol. 3. Cambridge University Press, Cambridge, UK.

Rodrik, D., Subramanian, A., Trebbi, F., 2004. Institutions rule: The primacy of institutions over geography and integration in economic development. J. Econ. Growth 9 (2), 131–165.

Rodrik, D., Wacziarg, R., 2005. Do democratic transitions produce bad economic outcomes? Am. Econ. Rev., Pap. Proc. 95 (2), 50–55.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97 (460), 1167–1179.

Sul, D., 2016. Pooling is harmful sometimes. Unpublished Mimeo, University of Texas At Dallas.

Swamy, P.A., 1970. Efficient inference in a random coefficient regression model. Econometrica 311–323.

Treisman, D., 2020. Democracy by mistake: How the errors of autocrats trigger transitions to freer government. Am. Political Sci. Rev. 114 (3), 792–810.

Westerlund, J., Urbain, J.-P., 2015. Cross-sectional averages versus principal components. J. Econometrics 185 (2), 372–377.

Xu, Y., 2017. Generalized synthetic control method: Causal inference with interactive fixed effects models. Political Anal. 25 (1), 57–76.

Young, A., 2022. Consistency without inference: Instrumental variables in practical application. Eur. Econ. Rev. 104112, forthcoming.