

Democracy, Growth, Heterogeneity, and Robustness*

Markus Eberhardt[†]

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

Final Version: April 21, 2022

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.

JEL Classification: O10, P16

Keywords: Democracy, Growth, Political Development, Difference-in-Difference Estimator, Interactive Fixed Effects

*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: Markus Eberhardt, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, UK; Tel: +44 115 8468416, Email: markus.eberhardt@nottingham.ac.uk

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 frequently¹ failed to establish a significant positive relationship ([Helliwell 1994](#), [Barro 1996](#), [Baum & Lake 2003](#) and [Murtin & Wacziarg 2014](#) for continuous democracy indices and [Giavazzi & Tabellini 2005](#), [Rodrik & Wacziarg 2005](#) for binary measures).²

All of the above studies adopt empirical models which treat the democracy-growth relationship³ 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 & 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

¹Exceptions include the large positive effects in [Papaioannou & 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 & Bond \(1991\)](#) GMM estimator which many researchers regard with scepticism in the context of cross-country regressions.

²A 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 & Tabellini 2009](#)) — for these and other contributions see the review in Appendix Table B-1. A broader set of papers is reviewed in [Dodsworth & Ramshaw \(2021\)](#).

³I 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 & Teal \(2011\)](#) for a detailed discussion of the interpretation of ‘cross-country growth regressions’.

⁴Some 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.

⁵It 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 & Smith 1995](#)).

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 & Sunde 2014](#), [Albertus & 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 & 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 & Kwok 2021](#)). 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

⁶Although there are many differences between these methods, the analogy to the popular synthetic control method, in particular in its generalized 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'.

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 (2020) 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.

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 unintention-

⁷While 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).

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

ally 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 discuss 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 & Siourounis (2008, PS). In my analysis I further explore the results using the indicators by BMR, CGV and PS (the latter is

⁹The 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.

¹⁰For 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)!

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.

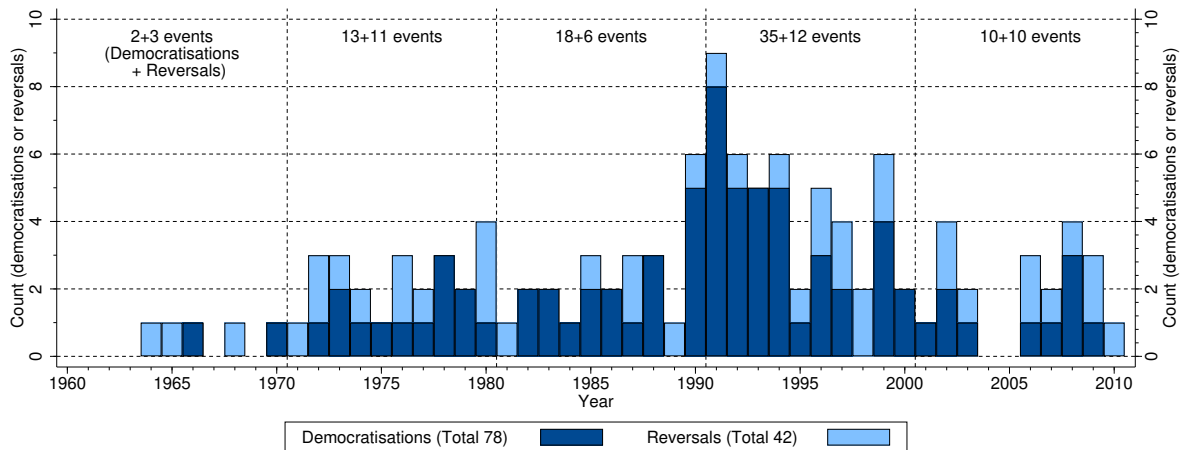
Figure 1 Panel (a) plots the distribution of democratisation events as well as reversals to autocracy (following the ANRR definition), restricted to my dynamic regression sample of 61 countries (treatment sample). The average country had 1.3 democratisation events and 0.7 reversals; 16 countries experienced 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 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.

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.

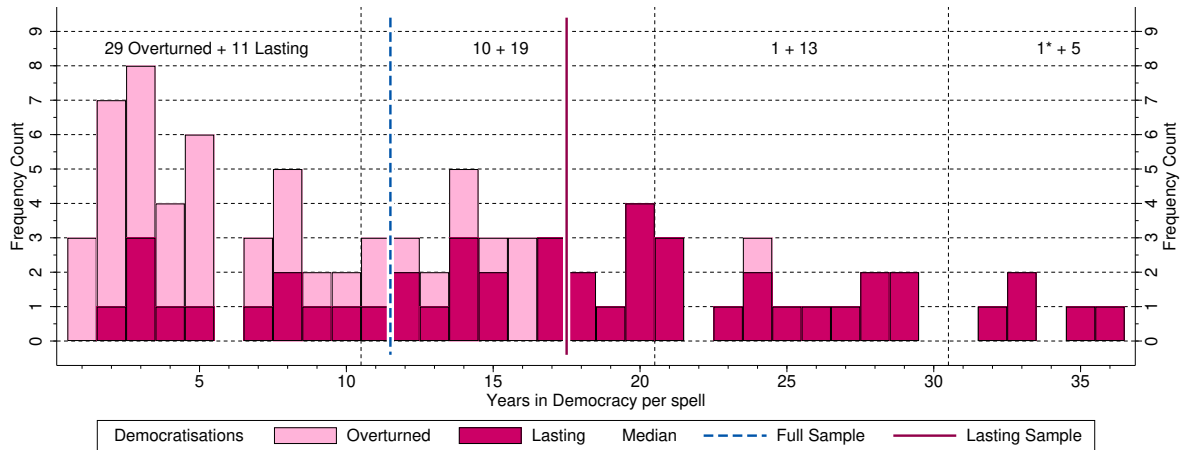
The potential treatment sample has 3,017 observations in 84 countries. However, es-

¹¹For 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.

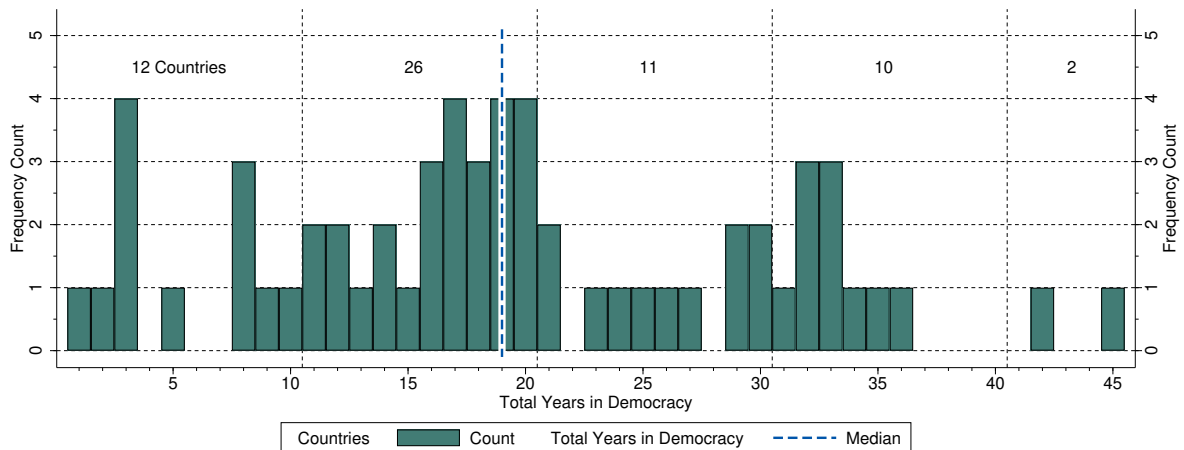
Figure 1: Democratisation, Reversal, and Time in Democracy — 1960-2010



(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

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.

timating 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 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

In 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 & 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, $\lambda'_i F_t$, where F is a set of common factors with associated heterogeneous factor loadings λ , 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 & Magnac 2016, Chan & Kwok 2021) 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).

¹²Compared 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.

¹³For 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.

¹⁴Strong 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 & Sul 2003, Andrews 2005).

Intuition In this paper I adopt the [Chan & 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 & 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 & Bond \(1991\)](#) and [Blundell & 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 & Kwok \(2021\)](#) 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

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

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

¹⁷The 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 & Watson 2002](#)).

¹⁸This PCA idea goes back to [Bai \(2009\)](#), but the somewhat surprising insight in the work of [Chan & Kwok \(2021\)](#) is that the estimated PCs are ‘good enough’ proxies in this setup, whereas in Bai’s implementation they (and the

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 & Siourounis 2008](#), [Cervellati & 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 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 & Kwok \(2021\)](#) 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 & Smith 1995](#)): if the true coefficient on the variable of interest x_{it} is β_i , yet the implementation imposes β , the error term ε 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_0 can be written as

$$y_{it} = \text{Dem}_{it} y_{it}(0) + (1 - \text{Dem}_{it}) y_{it}(1) = \Theta_{it} \mathbf{1}_{\{i \in E\}} \mathbf{1}_{\{t > T_0\}} + y_{it}(0) \quad (1)$$

$$\text{where} \quad y_{it}(0) = \varsigma_i + \beta'_i \mathbf{X}_{it} + \lambda'_i \mathbf{F}_t + \tilde{\varepsilon}_{it}. \quad (2)$$

The two indicator variables $\mathbf{1}_{\{\cdot\}}$ refer to the country and the year treated, Θ_{it} is the time-varying heterogeneous treatment effect, \mathbf{X} is a vector of observed covariates with associated country-specific parameters β_i , $\lambda'_i \mathbf{F}_t$ represents a set of unobserved common factors \mathbf{F}_t with country-specific factor loadings λ_i ,¹⁹ and $\tilde{\varepsilon}_{it}$ is the error term.

parameters on the variable of interest) are biased and need to be estimated iteratively.

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

The treatment effect is country-specific and time-varying, but we assume it follows a decomposition $\Theta_{it} = \bar{\Theta}_i + \tilde{\Theta}_{it}$, where $E(\tilde{\Theta}_{it}|t > T_0) = 0$ as $\tilde{\Theta}_{it}$ is the cross-sectionally demeaned idiosyncratic component of Θ_{it} ; $\bar{\Theta}_i$ is the ITET, *the treatment effect of country i averaged over the post-intervention period*. The reduced form model is then

$$y_{it} = \bar{\Theta}_i \mathbf{1}_{\{i \in E\}} \mathbf{1}_{\{t > T_0\}} + \varsigma_i + \beta_i' \mathbf{X}_{it} + \lambda_i' \mathbf{F}_t + \epsilon_{it}. \quad (3)$$

The composite error term, $\epsilon_{it} = \tilde{\epsilon}_{it} + \tilde{\Theta}_{it} \mathbf{1}_{\{i \in E\}} \mathbf{1}_{\{t > T_0\}}$, 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 & Kwok \(2021\)](#) 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 equation (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 \mathbf{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 ϵ is orthogonal to \mathbf{X} , \mathbf{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

$$y_{it} = \alpha_i + \theta_i \text{Dem}_{it} + \beta_i' \mathbf{X}_{it} + \delta_i^y \bar{y}_t + \delta_i^{X'} \bar{\mathbf{X}}_t + \epsilon_{it}, \quad (4)$$

where y is per capita GDP (in logs and multiplied by 100), Dem is the democracy dummy, and \mathbf{X} is a set of additional controls (gross investment share of GDP and trade openness). \bar{y} and $\bar{\mathbf{X}}$ are the cross-section averages of the observed variables *but for those countries which never*

²⁰With 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.

experienced democracy during the sample period (the control group).²¹ As was shown by Pesaran (2006) and Westerlund & Urbain (2015), the use of cross-section averages is very simple yet powerful in capturing a common factor structure.

The dynamic variant of equation (4) is:

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

where the two terms involving sums in the first line capture the short-run effects, while θ_i^* and β_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 of the ITET and the covariate coefficients is different from that in equation (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 & Kwok (2021) approach is convenient since the long-run democracy coefficient, θ_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 $p = p_{\bar{X}} = \text{int}(T^{1/3}) = 3$, 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 & Smith (1995). Observed covariates \mathbf{X} are not included in what I refer to as the ‘plain vanilla’ Chan & Kwok (2021) implementation — the covariate cross-section averages from the control sample, \bar{y} and $\bar{\mathbf{X}}$, are however always included.²³ For comparison, I also estimate simple Mean Group models (Pesaran & Smith 1995) which exclude the cross-section averages in equation (5).

²¹Country 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.

²²In 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 & Pesaran 2015). The CS-DL obtains these estimates in a single step by adopting an alternative specification and avoids potential bias from dynamic misspecification.

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

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 (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) Democracy (BMR)	15.130 (4.057) ^{***}	7.322 (3.279) ^{**}	9.983 (2.843) ^{***}	11.118 (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 (4.572) ^{***}	4.989 (4.059)	10.154 (3.190) ^{***}	6.934 (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 (4.813) ^{***}	4.749 (4.594)	11.501 (4.445) ^{***}	11.936 (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 [Papaiouannou & Siourounis \(2008\)](#) – PS.

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 equation (5), whereas C&K MG is the [Chan & Kwok \(2021\)](#) 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 & Kwok \(2021\)](#) 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 & Kwok \(2021\)](#) 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

²⁴The 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 .

²⁵In 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.

²⁶If, 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.

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.

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 & Kwok \(2021\)](#) 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 12.7%. 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’.

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).

3.4 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 & Kwok \(2021\)](#) 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 are significant differences in assumptions and implementation with the [Chan & Kwok \(2021\)](#) estimator, I also adopt the [\(Xu 2017\)](#) generalized 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.

Table 3: Main Results – Dynamic Specifications – Quartile TE

Implementation	Plain Vanilla		With Covariates	
	(1) MG	(2) C&K MG	(3) MG	(4) C&K MG
(a) Democracy (ANRR)				
Robust Mean	16.624 (4.630)***	7.692 (2.854)***	7.712 (3.647)**	10.074 (3.651)***
Lower Quartile	-1.713 (7.708)	-4.156 (3.257)	-6.006 (6.725)	-2.591 (4.963)
Median	15.299** (5.430)	7.067* (3.445)	8.209 (4.219)	10.872* (4.381)
Upper Quartile	40.153*** (7.980)	22.403*** (5.427)	26.525*** (7.031)	31.099*** (5.551)
Countries (N)	61	61	61	60

(Continued overleaf)

Table 3: Main Results – Dynamic Specifications – Quartile TE (Continued)

Implementation	Plain Vanilla		With Covariates	
	(1)	(2)	(3)	(4)
	MG	C&K MG	MG	C&K MG
(b) Democracy (BMR)				
Robust Mean	15.130 (4.057)***	7.322 (3.279)**	9.983 (2.843)***	11.118 (3.612)***
Lower Quartile	-1.679 (6.679)	-5.540 (5.052)	-2.758 (3.752)	-7.200 (6.339)
Median	16.858*** (4.787)	6.727 (3.609)	8.879* (3.630)	9.092 (4.954)
Upper Quartile	34.598*** (7.241)	21.814*** (5.898)	24.753*** (4.163)	30.344** (9.074)
Countries (N)	55	55	55	55
(c) Democracy (CGV)				
Robust Mean	14.405 (4.572)***	4.989 (4.059)	10.154 (3.190)***	6.934 (4.179)*
Lower Quartile	-2.684 (7.293)	-12.480 (6.651)	-4.832 (7.351)	-9.935 (7.200)
Median	15.505** (5.161)	5.354 (4.289)	8.457 (4.363)	5.499 (4.908)
Upper Quartile	33.493** (9.585)	19.371* (8.206)	27.378*** (5.227)	22.390* (8.811)
Countries (N)	49	49	49	49
(d) Democracy (PS)				
Robust Mean	27.915 (4.813)***	4.749 (4.594)	11.501 (4.445)***	11.936 (5.072)**
Lower Quartile	12.520 (10.037)	-9.936 (7.126)	-0.712 (7.592)	-6.512 (9.023)
Median	28.962*** (5.987)	7.235 (6.239)	10.278 (5.452)	10.715 (5.946)
Upper Quartile	46.503*** (9.618)	22.478** (7.726)	27.539** (8.243)	33.168*** (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.

Turning to existing theories, I first study competing explanations in political science whereby democratisations can be distinguished as ‘elite-biased’ or ‘popular’ (Albertus & Menaldo 2018),²⁷ or to have occurred ‘by mistake’ rather than intention (Treisman 2020). Panels (a) and (b) of Figure 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 coefficients²⁸ 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 Figure 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 sup-

²⁷These 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 & Menaldo 2018, 7, emphases in original).

²⁸In 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).

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

port 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 Figure 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 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 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.

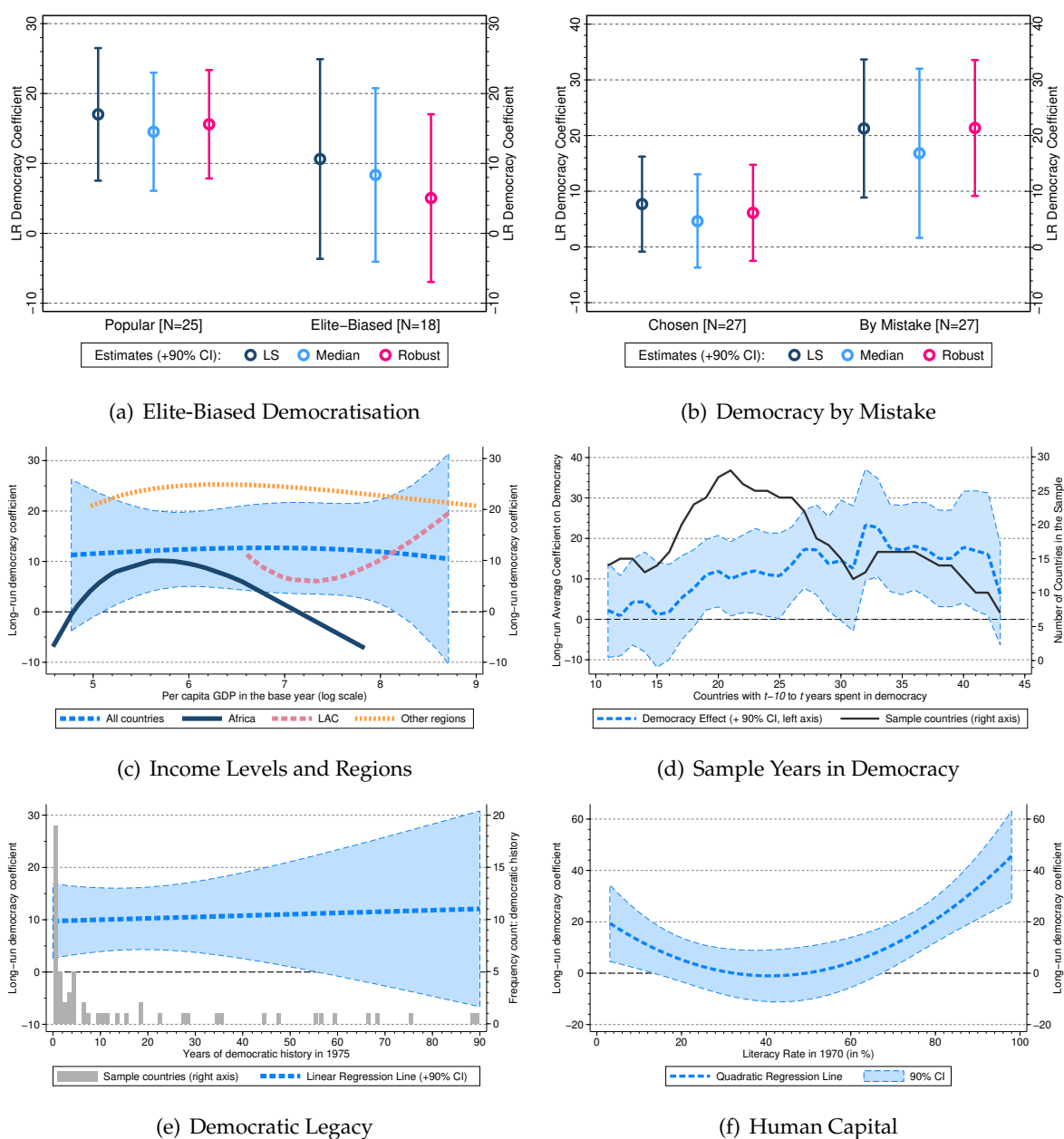
Fourth, instead of studying the in-sample experience of democracy, I gauge the significance of a democratic legacy since 1800, proxied by the number of years spent in democracy by 1975 — this cut-off maximises the data availability in the Polity IV dataset.³¹ Gerring et al. (2005), among others, argue that political regimes are historical legacies, with cumulative effects of institutions (only) coming to bear over long time horizons. Panel (e) of Figure 2 shows a fitted linear regression line³² 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

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

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

³²A 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.

Figure 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 & Menaldo 2018) using least squared, median and robust regression. Panel (b) does the same for Treisman’s (2020) ‘democracy by mistake’. The figure in Panel (c) 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 & 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 (d) 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 (e) 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%. 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.

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

³³These 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.

³⁴The 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.

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

(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 \(2020\)](#) 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.

4.1 Sample reduction by minimum observation count

Panel (a) of Figure 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 save 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 plain vanilla [Chan & Kwok \(2021\)](#) model accounting for these distortions (in short blue dashes) yields more stable long-run estimates at around 6-8% until the coefficient

³⁶In 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.

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.

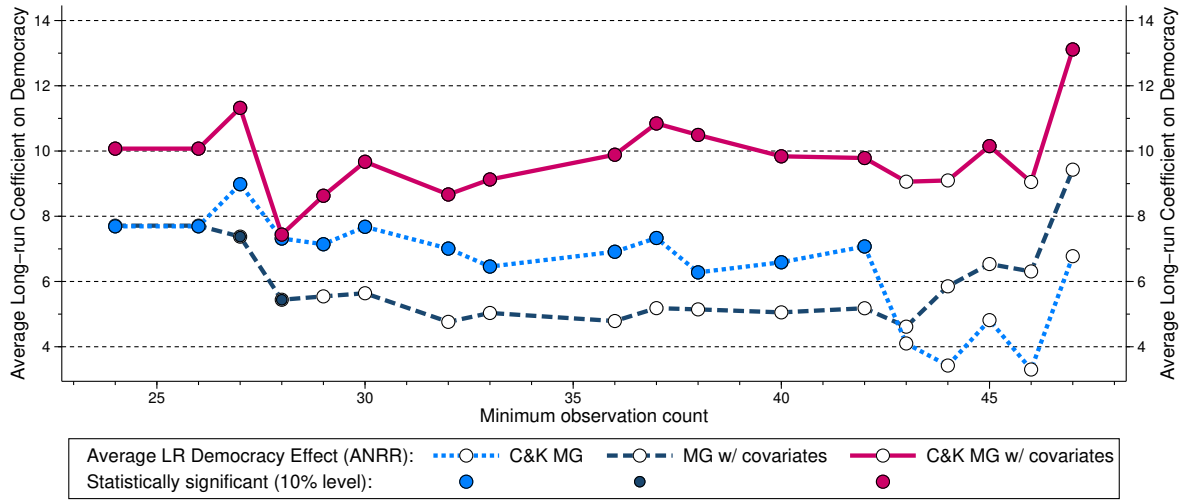
Panel (a) of Figure 4 reports the results for the same sample reduction exercise using the four principle parametric estimators in ANRR.³⁷ Focusing on the 2SLS results (solid dark pink line) long-run democracy estimate turns statistically insignificant when fewer than 7% of observations are omitted (the 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.

4.2 Sample reduction by sample end year

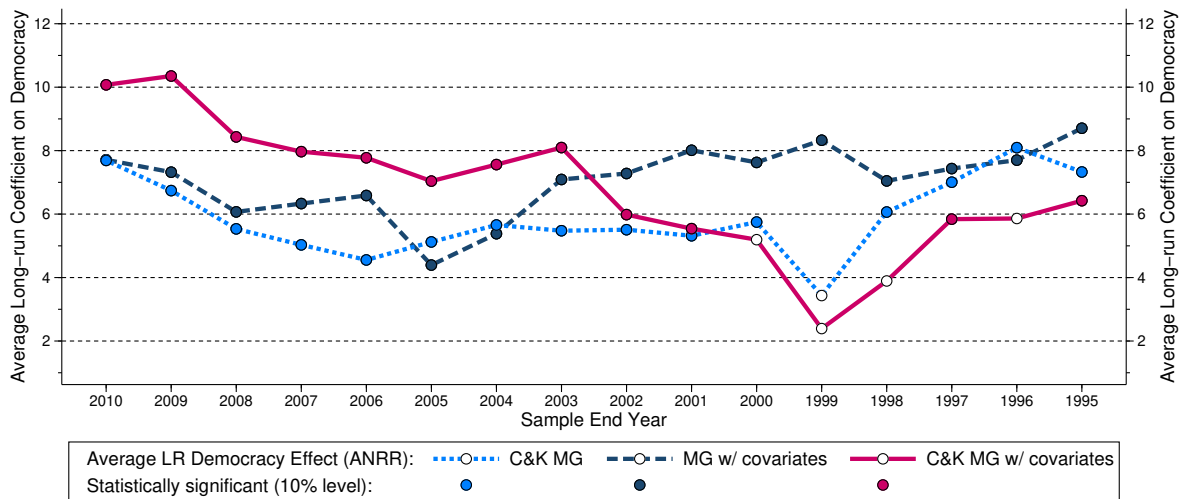
Panel (b) of Figure 3 presents the mean heterogeneous Diff-in-Diff results when observations are omitted by sample end year. In this graph for the ANRR democracy dummy, and in the equivalent graphs in Appendix D for BMR, CGV and PS definitions, the x -axis is in *reverse* chronological order. Here, the standard MG estimate of the long-run democracy effect (in dark

³⁷In 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.

Figure 3: Sample Reduction Exercises — Heterogeneous Diff-in-Diff Models



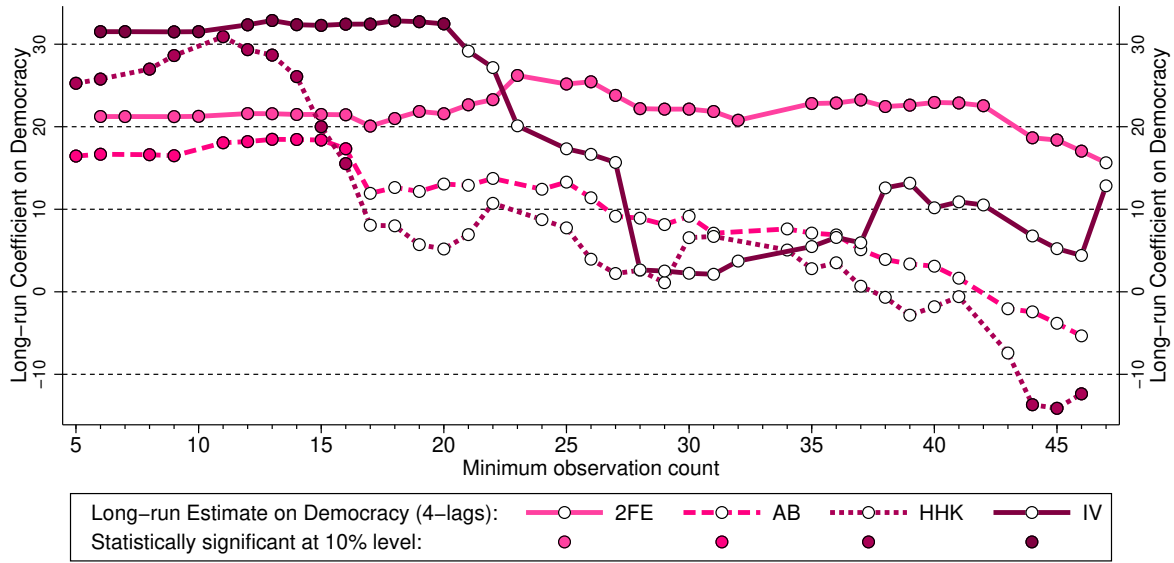
(a) Sample reduction by T_i count



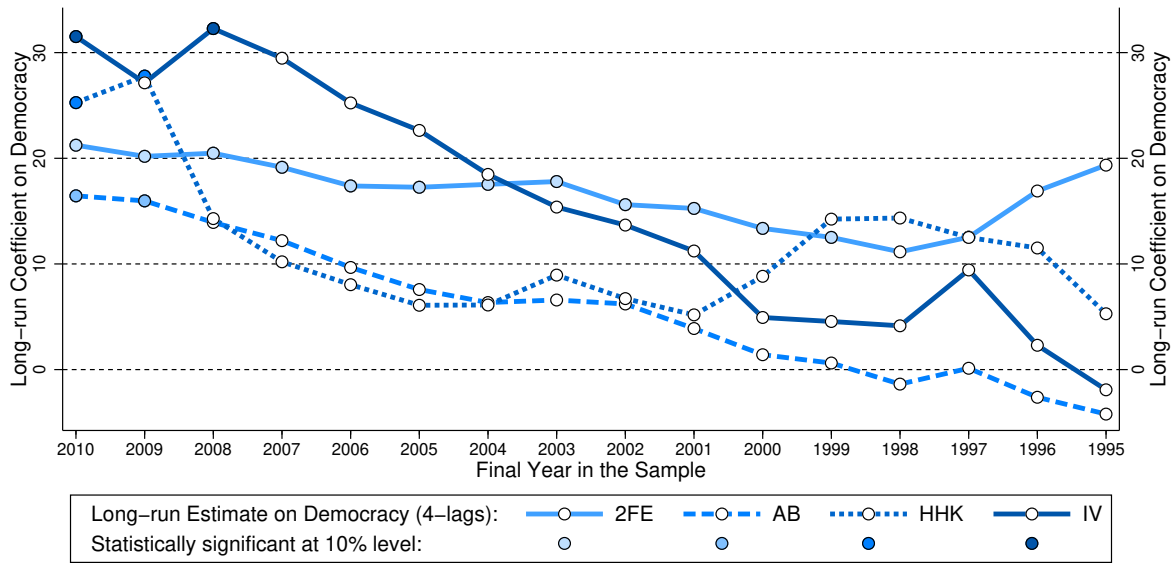
(b) Sample reduction by end year

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.

Figure 4: Sample Reductions — ANRR (Parametric) Models



(a) Sample reduction by T_i count



(b) Sample reduction by end year

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%.

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 (3.651)***	11.118 (3.612)***	6.934 (4.179)*	11.936 (5.072)**	10.074 (3.651)***	11.118 (3.612)***	6.934 (4.179)*	11.936 (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 (5.551)	8.213 (5.602)	6.084 (4.014)	9.222 (6.147)	5.194 (3.220)	4.715 (3.135)	5.540 (3.520)	5.760 (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 (6.088)**	10.036 (6.591)	10.420 (5.795)*	9.649 (7.719)	n/a	n/a	n/a	n/a
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 & Kwok \(2021\)](#) 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 Figure 3 and for all other definitions of democracy in Appendix Figures 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.

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 & Kwok \(2021\)](#) 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 & Kwok \(2021\)](#) 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 Figure 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, while the analysis of decadal data over a much longer time horizon in [Madsen et al. \(2015\)](#) holds up quite well.

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

³⁸[Madsen et al.’s \(2015\)](#) benchmark specification is again remarkably stable and does not turn statistically insignificant in the periods considered in 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).

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 & 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 & Eberhardt 2022).

References

Acemoglu, D., Johnson, S. & Robinson, J. A. (2001), 'The colonial origins of comparative development: An empirical investigation', *American Economic Review* 91(5), 1369–1401.

- Acemoglu, D., Naidu, S., Restrepo, P. & Robinson, J. A. (2015), Democracy, redistribution, and inequality, in F. B. Anthony B. Atkinson, ed., 'Handbook of Income Distribution', Vol. 2, Elsevier, pp. 1885–1966.
- Acemoglu, D., Naidu, S., Restrepo, P. & Robinson, J. A. (2019), 'Democracy does cause growth', *Journal of Political Economy* **127**(1), 47–100.
- Aghion, P., Akcigit, U. & Howitt, P. (2014), What do we learn from Schumpeterian growth theory?, in P. Aghion & S. N. Durlauf, 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', *Review of Economic Studies* **58**(2), 277–297.
- Athey, S., Bayati, M., Doudchenko, N., Imbens, G. & Khosravi, K. (2021), 'Matrix completion methods for causal panel data models', *Journal of the American Statistical Association* .
- 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', *Quarterly Journal of Economics* **106**(2), 407–443.
- Barro, R. J. (1996), 'Democracy and growth', *Journal of Economic Growth* **1**(1), 1–27.
- Baum, M. A. & Lake, D. A. (2003), 'The political economy of growth: democracy and human capital', *American Journal of Political Science* **47**(2), 333–347.
- Blundell, R. & Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of 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', *Comparative Political Studies* **46**(12), 1523–1554.
- Boucekkine, R., Desbordes, R. & Melindi-Ghidi, P. (2021), 'A theory of elite-biased democracies', *Mathematical Social Sciences* **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', *Scandinavian Journal of Economics* **116**(2), 482–505.
- Chan, M. K. & Kwok, S. S. (2021), 'The PCDID approach: difference-in-differences when trends are potentially unparallel and stochastic', *Journal of Business & Economic Statistics* (forthcoming).
- 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 T.-H. L. Gloria Gonzalez-Rivera, R. Carter Hill, ed., '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', *Journal of Econometrics* **188**(2), 393–420.
- Comin, D. & Hobijn, B. (2004), 'Cross-country technology adoption: making the theories face the facts', *Journal of Monetary Economics* **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, Available at: <https://doi.org/10.23696/vdemds21>.
- De Visscher, S., Eberhardt, M. & Everaert, G. (2020), 'Estimating and testing the multicountry endogenous growth model', *Journal of International Economics* **125**, 103325.
- Dodsworth, S. & Ramshaw, G. (2021), 'Democracy's development dividend', *Journal of Democracy* **32**(1), 126–138.
- Durlauf, S. N. (2020), Institutions, development, and growth: Where does evidence stand?, in J.-P. J.-M Baland, F Bourguignon & T. Verdier, eds, 'The Handbook of Economic Development and Institutions', Princeton University Press, p. 189–217.
- Durlauf, S. N., Johnson, P. A. & Temple, J. R. (2005), 'Growth econometrics', *Handbook of Economic Growth* **1**, 555–677.

- Eberhardt, M. & Teal, F. (2011), 'Econometrics for grumblers: a new look at the literature on cross-country growth empirics', *Journal of Economic Surveys* **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', *Journal of Monetary Economics* **52**(7), 1297–1330.
- Glaeser, E. L., La Porta, R., Lopez-de Silanes, F. & Shleifer, A. (2004), 'Do institutions cause growth?', *Journal of economic Growth* **9**(3), 271–303.
- Gobillon, L. & Magnac, T. (2016), 'Regional policy evaluation: Interactive fixed effects and synthetic controls', *Review of Economics and Statistics* **98**(3), 535–551.
- Hamilton, L. C. (1992), 'How robust is robust regression?', *Stata Technical Bulletin* **1**(2).
- Helliwell, J. F. (1994), 'Empirical linkages between democracy and economic growth', *British Journal of Political Science* pp. 225–248.
- Knutsen, C. H. (2013), 'Democracy, state capacity, and economic growth', *World Development* **43**, 1–18.
- Madsen, J. B., Raschky, P. A. & Skali, A. (2015), 'Does democracy drive income in the world, 1500–2000?', *European Economic Review* **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', *Journal of Economic Growth* **19**(2), 141–181.
- Papaioannou, E. & Siourounis, G. (2008), 'Democratisation and growth', *Economic Journal* **118**(532), 1520–1551.
- Persson, T. & Tabellini, G. (2009), 'Democratic capital: The nexus of political and economic change', *American Economic Journal: Macroeconomics* **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', *Journal of Econometrics* **68**(1), 79–113.
- Phillips, P. C. & Sul, D. (2003), 'Dynamic panel estimation and homogeneity testing under cross section dependence', *The Econometrics Journal* **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', *Journal of Economic Growth* **9**(2), 131–165.
- Rodrik, D. & Wacziarg, R. (2005), 'Do democratic transitions produce bad economic outcomes?', *American Economic Review, Papers & Proceedings* **95**(2), 50–55.
- Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **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* pp. 311–323.
- Treisman, D. (2020), 'Democracy by mistake: How the errors of autocrats trigger transitions to freer government', *American Political Science Review* **114**(3), 792–810.
- Westerlund, J. & Urbain, J.-P. (2015), 'Cross-sectional averages versus principal components', *Journal of Econometrics* **185**(2), 372–377.
- Xu, Y. (2017), 'Generalized synthetic control method: Causal inference with interactive fixed effects models', *Political Analysis* **25**(1), 57–76.
- Young, A. (2020), 'Consistency without inference: Instrumental variables in practical application', *unpublished mimeo* .

Appendix – Not intended for publication

A Data Source and Sample Makeup

Income (ANRR) Data are taken from the World Bank World Development Indicators (WDI) database for real GDP per capita in year 2000 US\$. The GDPpc variable is transformed into logarithms and multiplied by 100, which eases the interpretation of the coefficients on the democracy dummy.

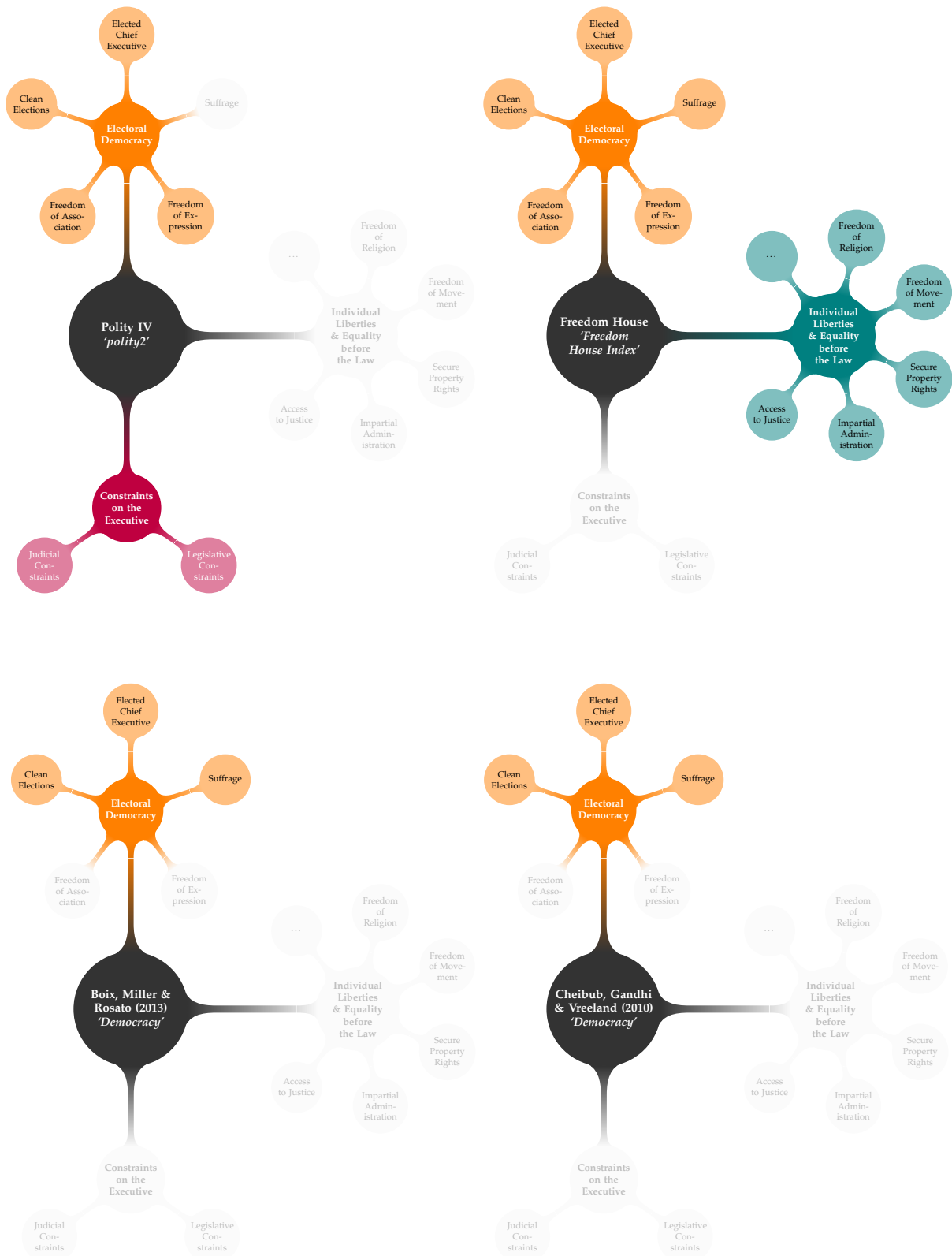
Democracy (ANRR) Data are combined from the Polity IV project, Freedom House and a number of alternative sources. The primary strategy for construction of the democracy dummy prescribes a positive Polity score in addition to a ‘free’ or ‘partially free’ label in Freedom House. Further strategies are described in detail in an appendix to ANRR. In Figure [A-1](#) below I highlight the ‘coverage’ of BMR, CGV as well as the polity2 (PolityIV) and FHI measures in terms of different aspects of political institutions: by combining the latter two measures ANRR come closer to the V-Dem definition of ‘liberal democracy’ which captures electoral democracy, individual liberties and constraints to the executive.

Investment (ANRR) Data are taken from the World Bank World Development Indicators (WDI) database for the share of gross investment in GDP.

Trade Openness (ANRR) Data are taken from the World Bank World Development Indicators (WDI) database for the sum of imports and exports expressed as a share of GDP.

All of the above variables are compiled by ANRR and provided for download (along with the Stata do-files used in the analysis) from Daron Acemoglu’s personal website. Table [A-2](#) indicates the sample makeup for the analysis of each of the four definitions of democracy (treatment sample), focusing on the dynamic specifications presented in the main section of the paper. Table [A-1](#) presents the 38 countries which make up the control group (countries which never transitioned into democracy.) For reference, I also provide the sample makeup (treated sample) for static specifications in Table [A-3](#).

Figure A-1: Alternative Empirical Measures of Democracy



Notes: The figure compares four popular measures for democracy with the V-Dem conceptual framework for 'liberal democracy', where faint gray aspects are not covered by the democracy measure in question. Note that for the Freedom House FHI the index *does* include aspects of executive constraints but that these are given much less significance than in the Polity IV or V-Dem data. This visualisation merely covers the elements covered by each measure for democracy, not the substantial variation in the aggregation procedure. The ANRR measure is a combination of the PolityIV and FHI, checked against the PS measure for permanent democratisation.

Table A-1: Control Sample — Dynamic Specifications

	wbcode	obs	start	end	M		wbcode	obs	start	end	M
Angola	AGO	23	1986	2010	2	Oman	OMN	39	1968	2008	2
Bahrain	BHR	28	1981	2008		Qatar	QAT	10	2000	2009	
Bosnia & Herzegovina	BIH	16	1995	2010		Rwanda	RWA	49	1962	2010	
Brunei Darussalam	BRN	20	1990	2009		Saudi Arabia	SAU	42	1969	2010	
PR China	CHN	40	1971	2010		Singapore	SIN	45	1966	2010	
Cameroon	CMR	45	1966	2010		Swaziland	SWZ	35	1976	2010	
Cuba	CUB	40	1971	2010		Syria	SYR	50	1961	2010	
Algeria	DZA	49	1962	2010		Chad	TCD	47	1961	2010	3
Egypt	EGY	50	1961	2010		Togo	TGO	50	1961	2010	
Eritrea	ERI	15	1993	2007		Tajikistan	TJK	22	1989	2010	
Gabon	GAB	38	1970	2007		Turkmenistan	TKM	16	1993	2010	2
Equatorial Guinea	GNQ	19	1990	2010	2	Tonga	TON	30	1981	2010	
Iran	IRN	42	1966	2007		Tunisia	TUN	49	1962	2010	
Jordan	JOR	34	1977	2010		Tanzania	TZA	20	1991	2010	
Kazakhstan	KAZ	18	1993	2010		Uzbekistan	UZB	20	1991	2010	
Kuwait	KWT	13	1995	2007		Vietnam	VNM	24	1987	2010	
Lao PDR	LAO	15	1985	2010	11	Yemen	YEM	20	1991	2010	
Libya	LBY	10	1999	2008							
Morocco	MAR	50	1961	2010		Totals	38	1,194			
Maldives	MDV	11	1995	2005							
Malaysia	MYS	50	1961	2010							

Notes: This table provides sample details for the set of control countries from which the common factor proxies are constructed (cross-section averages for per capita GDP, gross investment rate, trade openness). M indicates the number of missing observations in the time series.

Table A-2: Regression Sample — Dynamic Specifications

	ANRR (1961-2010)					BMR (1961-2007)					CGV (1961-2008)					PS (1961-2010)									
	In Democracy		Events			In Democracy		Events			In Democracy		Events			In Democracy		Events							
	0	1	Obs	D	A	0	1	Obs	D	A	0	1	Obs	D	A	0	1	Obs	D	A					
1	Albania	ALB	9	18	27	2	1	ALB	9	15	24	2	1	ALB	7	18	25	1	0	ALB	8	19	27	1	0
2	Argentina	ARG	16	31	47	2	1	ARG	14	30	44	2	2	ARG	14	31	45	2	2	ARG	19	28	47	1	0
3	Antigua & Barbuda	ATG						ATG	19	4	23	1	0	ATG						ATG					
4	Burundi	BDI	38	8	46	1	0	BDI	40	3	43	1	0	BDI	37	7	44	2	1	BDI					
5	Benin	BEN	27	20	47	1	0	BEN	27	17	44	1	0	BEN	27	18	45	1	0	BEN	27	20	47	1	0
6	Burkina Faso	BFA	44	3	47	1	1	BFA						BFA						BFA					
7	Bangladesh	BGD	18	18	36	2	1	BGD	13	21	34	1	1	BGD	14	21	35	1	1	BGD	16	20	36	1	0
8	Bulgaria	BGR	7	20	27	1	0	BGR	6	18	24	1	0	BGR	6	19	25	1	0	BGR	7	20	27	1	0
9	Bolivia	BOL	8	29	37	1	0	BOL	7	27	34	2	1	BOL	7	28	35	2	1	BOL	8	29	37	1	0
10	Brazil	BRA	21	26	47	1	1	BRA	15	29	44	1	1	BRA	21	24	45	1	1	BRA	21	26	47	1	0
11	Bhutan	BTN	24	3	27	1	0	BTN						BTN						BTN					
12	Central African Rep	CAF	37	10	47	1	1	CAF	34	10	44	1	1	CAF	35	10	45	1	1	CAF					
13	Chile	CHL	17	30	47	1	1	CHL	17	27	44	1	1	CHL	17	28	45	1	1	CHL	26	21	47	1	0
14	Cote d'Ivoire	CIV	43	2	45	1	1	CIV						CIV						CIV					
15	Rep of Congo	COG	42	5	47	1	1	COM	22	2	24	1	0	COM	40	5	45	1	1	COM					
16	Comores	COM	10	16	26	3	2	COM						COM	15	10	25	2	1	COM					
17	Cape Verde	CPV	7	20	27	1	0	CPV	7	17	24	1	0	CPV	6	19	25	1	0	CPV	7	20	27	1	0
18	Cyprus	CYP						CYP						CYP	4	26	30	1	0	CYP					
19	Dominican Rep	DOM	14	33	47	1	0	DOM	2	42	44	1	0	DOM	2	43	45	1	0	DOM	14	33	47	1	0
20	Ecuador	ECU	15	32	47	1	0	ECU	18	26	44	2	1	ECU	17	28	45	2	1	ECU	15	32	47	1	0
21	Spain	ESP	5	33	38	1	0	ESP	4	31	35	1	0	ESP	4	32	36	1	0	ESP	5	33	38	1	0
22	Ethiopia	ETH	11	15	26	1	1	ETH						ETH						ETH	10	16	26	1	0
23	Fiji	FJI	6	30	36	1	2	FJI	21	14	35	0	1	FJI	28	8	36	1	1	FJI					
24	Ghana	GHA	28	19	47	3	2	GHA	29	15	44	3	2	GHA	24	21	45	3	2	GHA	32	15	47	1	0
25	Gambia	GMB	17	24	41	0	1	GMB	16	22	38	1	1	GMB						GMB					
26	Guinea-Bissau	GNB	18	8	26	2	1	GNB	22	4	26	1	1	GNB	23	3	26	1	0	GNB					
27	Greece	GRC	7	36	43	1	0	GRC	6	34	40	1	0	GRC	6	35	41	1	0	GRC	7	36	43	1	0
28	Granada	GRD	3	27	30	1	0	GRD	3	24	27	1	0	GRD	3	25	28	1	0	GRD	3	27	30	1	0
29	Guatemala	GTM	14	33	47	2	1	GTM	6	38	44	2	1	GTM	6	39	45	2	1	GTM	32	15	47	1	0
30	Guyana	GUY	23	14	37	1	0	GUY	23	14	37	1	0	GUY						GUY	23	14	37	1	0
31	Honduras	HND	18	29	47	1	0	HND	17	27	44	2	1	HND	17	28	45	2	1	HND	18	29	47	1	0
32	Hungary	HUN	25	21	46	1	0	HUN	25	18	43	1	0	HUN	25	19	44	1	0	HUN	25	21	46	1	0
33	Indonesia	IDN	35	12	47	1	0	IDN	35	9	44	1	0	IDN	35	10	45	1	0	IDN	35	12	47	1	0
34	Kenya	KEN	36	9	45	1	0	KEN	36	6	42	1	0	KEN	32	11	43	1	0	KEN					
35	South Korea	KOR	24	23	47	1	0	KOR	24	20	44	1	0	KOR	24	21	45	1	0	KOR	24	23	47	1	0

Table A-3: Regression Sample — Static Specifications

	ANRR (1961-2010)			BMR (1961-2007)			CGV (1961-2008)			PS (1961-2010)		
	In Democracy	Events		In Democracy	Events		In Democracy	Events		In Democracy	Events	
	0	1	Obs	0	1	Obs	0	1	Obs	0	1	Obs
1 Albania	ALB	12	18	ALB	12	15	ALB	10	18	ALB	11	12
2 Argentina	ARG	19	31	ARG	15	32	ARG	15	33	ARG	22	21
3 Armenia	ARM	2	18	ARM	ARM		ARM	ARM		ARM	7	6
4 Antigua & Barbuda	ATG	ATG	19	ATG	22	4	ATG	ATG		ATG	ATG	13
5 Azerbaijan	AZE	AZE	19	AZE	AZE		AZE	AZE		AZE	AZE	1
6 Burundi	BDI	41	8	BDI	43	3	BDI	40	7	BDI	40	7
7 Benin	BEN	30	20	BEN	30	17	BEN	30	18	BEN	30	13
8 Burkina Faso	BFA	47	3	BFA	BFA		BFA	BFA		BFA	BFA	43
9 Bangladesh	BGD	19	20	BGD	16	21	BGD	17	21	BGD	19	13
10 Bulgaria	BGR	10	20	BGR	9	18	BGR	9	19	BGR	10	13
11 Belarus	BLR	16	4	BLR	14	3	BLR	BLR		BLR	BLR	32
12 Bolivia	BOL	11	29	BOL	10	27	BOL	10	28	BOL	11	22
13 Brazil	BRA	21	29	BRA	15	32	BRA	21	27	BRA	24	19
14 Bhutan	BTN	27	3	BTN	BTN		BTN	26	2	BTN	24	19
15 Central African Rep	CAF	40	10	CAF	37	10	CAF	38	10	CAF	10	13
16 Chile	CHL	17	33	CHL	17	30	CHL	17	31	CHL	29	14
17 Cote d'Ivoire	CIV	46	2	CIV	CIV		CIV	CIV		CIV	29	14
18 Rep of Congo	COG	43	7	COG†	45	2	COG	41	7	COG	11	22
19 Comores	COM	13	16	COM	25	2	COM	18	10	COM	24	19
20 Cape Verde	CPV	10	20	CPV	10	17	CPV	9	19	CPV	10	13
21 Cyprus	CYP	CYP	2	CYP	1	31	CYP	7	26	CYP	10	13
22 Czech Rep	CZE	2	18	CZE	CZE		CZE	2	16	CZE	10	13
23 Djibouti	DJI	8	9	DJI	DJI		DJI	DJI		DJI	8	5
24 Dominican Rep	DOM	17	33	DOM	5	42	DOM	5	43	DOM	17	26
25 Ecuador	ECU	18	32	ECU	19	28	ECU	18	30	ECU	18	25
26 Spain	ESP	8	33	ESP	7	31	ESP	7	32	ESP	8	26
27 Ethiopia	ETH	14	15	ETH	ETH		ETH	ETH		ETH	13	9
28 Fiji	FJI	6	33	FJI†	21	17	FJI	31	8	FJI	8	5
29 Georgia	GEO	7	16	GEO	16	4	GEO	16	5	GEO	4	9
30 Ghana	GHA	31	19	GHA	32	15	GHA	27	21	GHA	35	8

Table A-3: Regression Sample — Static Specifications (continued)

	ANRR (1961-2010)					BMR (1961-2007)					CGV (1961-2008)					PS (1961-2010)									
	In Democracy		Events		A	In Democracy		Events		A	In Democracy		Events		A	In Democracy		Events		A					
	0	1	Obs	D		0	1	Obs	D		0	1	Obs	D		0	1	Obs	D		0	1	Obs	D	A
31	Guinea	GIN	23	1	24	1	0	GIN	19	22	41	1	1	GIN	7	41	48	2	2	GTM	35	8	43	1	0
32	Gambia	GMB†	17	27	44	0	1	GMB	25	4	29	1	1	GMB	26	3	29	1	0	GUY	26	12	38	1	0
33	Guinea-Bissau	GNB	21	8	29	2	1	GNB	7	36	43	1	1	GNB	7	37	44	1	1	HND	21	22	43	1	0
34	Greece	GRC	8	38	46	1	1	GRC	5	25	30	1	1	GRC	5	26	31	1	1	HRV	8	4	12	1	0
35	Granada	GRD	5	28	33	1	1	GRD	7	40	47	2	2	GRD	7	41	48	2	2	HTI	35	8	43	1	0
36	Guatemala	GTM	17	33	50	2	1	GTM	26	14	40	1	0	GTM	26	14	40	1	0	HND	21	22	43	1	0
37	Guyana	GUY	26	14	40	1	0	GUY	8	8	16	1	0	GUY	18	30	48	2	2	KHM†	16	1	17	0	1
38	Honduras	HND	21	29	50	1	0	HND	25	18	43	1	0	HND	25	19	44	1	0	KOR	27	21	48	1	1
39	Croatia	HRV	8	11	19	1	0	HRV	38	9	47	1	0	HRV	38	10	48	1	0	LBN	15	6	21	1	0
40	Haiti	HTI	9	10	19	2	2	HTI	39	6	45	1	0	HTI	35	11	46	1	0	LBR	4	7	11	1	0
41	Hungary	HUN	25	21	46	1	0	HUN	6	2	8	1	0	HUN	6	2	8	1	0	LKA	14	29	43	1	1
42	Indonesi	IDN	38	12	50	1	0	IDN	14	29	43	1	1	IDN	12	32	44	1	1	LSO	28	17	45	2	1
43	Kenya	KEN	39	9	48	1	0	KEN	36	6	42	1	0	KEN	36	6	42	1	0	LTU	2	18	20	1	0
44	Kyrgyzstan	KGZ	15	5	20	2	1	KGZ	32	15	47	1	0	KGZ	32	16	48	1	0	LVA	2	18	20	1	0
45	Cambodia	KHM†	16	1	17	0	1	KHM	39	8	47	1	0	KHM	39	9	48	1	0	MDA	3	17	20	1	0
46	South Korea	KOR	27	23	50	1	1	KOR	32	15	47	1	0	KOR	32	16	48	1	0	MDG	33	16	49	1	1
47	Lebanon	LBN	15	6	21	1	0	LBN	39	8	47	1	0	LBN	39	8	47	1	0	MEX	36	14	50	1	0
48	Liberia	LBR	4	7	11	1	0	LBR	27	20	47	1	1	LBR	27	21	48	1	1	MKD	36	14	50	1	1
49	Sri Lanka	LKA	14	29	43	1	1	LKA	24	16	40	1	0	LKA	24	17	41	1	0	MLI	24	19	43	1	0
50	Lesotho	LSO	28	17	45	2	1	LSO	8	18	26	1	0	LSO	8	19	27	1	0	MNG	11	18	29	1	0
51	Lithuania	LTU	2	18	20	1	0	LTU	17	10	27	1	1	LTU	17	10	27	1	1	MOZ	13	17	30	1	0
52	Latvia	LVA	2	18	20	1	0	LVA	1	16	17	1	0	LVA	1	16	17	1	0	MRT	49	1	50	1	1
53	Moldava	MDA	3	17	20	1	0	MDA	2	15	17	1	0	MDA	2	15	17	1	0						
54	Madagascar	MDG	33	16	49	1	1	MDG	32	15	47	1	0	MDG	32	16	48	1	0						
55	Mexico	MEX	36	14	50	1	0	MEX	39	8	47	1	0	MEX	39	9	48	1	0						
56	Macedonia	MKD						MKD						MKD											
57	Mali	MLI	24	19	43	1	0	MLI	24	16	40	1	0	MLI	24	17	41	1	0						
58	Mongolia	MNG	11	18	29	1	0	MNG	8	18	26	1	0	MNG	8	19	27	1	0						
59	Mozambique	MOZ	13	17	30	1	0	MOZ	17	10	27	1	1	MOZ	17	10	27	1	1						
60	Mauritania	MRT	49	1	50	1	1	MRT	47	1	48	1	1	MRT	47	1	48	1	1						

Table A-3: Regression Sample — Static Specifications (continued)

	ANRR (1961-2010)										BMR (1961-2007)										CGV (1961-2008)										PS (1961-2010)									
	In Democracy					Events					In Democracy					Events					In Democracy					Events					In Democracy					Events				
	0	1	Obs	D	A	0	1	Obs	D	A	0	1	Obs	D	A	0	1	Obs	D	A	0	1	Obs	D	A	0	1	Obs	D	A										
61	Malawi	MWI	30	16	46	1	0	0	0	30	13	43	1	0	MWI	30	14	44	1	0	MWI	30	14	44	1	0	MWI	30	9	39	1	0								
62	Niger	NER	33	12	45	2	1	0	0	35	10	45	2	1	NER	36	9	45	2	1	NER	36	9	45	2	1	NER	36	14	43	1	0								
63	Nicaragua	NIC	29	21	50	1	0	0	0	23	24	47	1	0	NIC	23	25	48	1	0	NIC	23	25	48	1	0	NIC	29	14	43	1	0								
64	Nepal	NPL	29	16	45	2	1	0	0	31	11	42	1	1	NPL	30	13	43	2	1	NPL	30	13	43	2	1	NPL	30	14	43	1	0								
65	Pakistan	PAK	24	19	43	3	2	0	0	24	16	40	2	2	PAK	24	17	41	3	2	PAK	24	17	41	3	2	PAK	24	17	41	3	2								
66	Panama	PAN	13	17	30	1	0	0	0	10	17	27	1	0	PAN	8	20	28	1	0	PAN	8	20	28	1	0	PAN	13	10	23	1	0								
67	Peru	PER	14	36	50	3	3	0	0	24	23	47	3	3	PER	24	24	48	3	3	PER	24	24	48	3	3	PER	19	24	43	1	0								
68	Philippines	PHL	22	28	50	1	1	0	0	21	26	47	1	1	PHL	21	27	48	1	1	PHL	21	27	48	1	1	PHL	26	17	43	1	0								
69	Portugal	PRT	6	35	41	1	0	0	0	6	32	38	1	0	PRT	6	33	39	1	0	PRT	6	33	39	1	0	PRT	6	28	34	1	0								
70	Paraguay	PRY	1	18	19	1	0	0	0	11	5	16	1	0	PRY	11	5	16	1	0	PRY	11	5	16	1	0	PRY	1	11	12	1	0								
71	Russia	RUS	10	11	21	1	1	0	0	11	7	18	1	1	RUS	11	7	18	1	1	RUS	11	7	18	1	1	RUS	2	11	13	1	0								
72	Sudan	SDN	36	3	39	1	1	0	0	33	3	36	1	1	SDN	34	3	37	1	1	SDN	34	3	37	1	1	SDN	19	24	43	1	0								
73	Senegal	SEN	39	11	50	1	0	0	0	39	8	47	1	0	SEN	39	9	48	1	0	SEN	39	9	48	1	0	SEN	39	4	43	1	0								
74	Solomon Isl.	SLB	4	7	11	1	1	0	0	6	5	11	1	1	SLB	6	5	11	1	1	SLB	6	5	11	1	1	SLB	6	28	34	1	0								
75	Sierra Leone	SLE	20	11	31	2	1	0	0	22	6	28	1	0	SLE	17	12	29	2	1	SLE	17	12	29	2	1	SLE	6	28	34	1	0								
76	El Salvador	SLV													SLV						SLV						SLV	7	10	17	1	0								
77	Suriname	SUR	9	22	31	2	2	0	0	9	22	31	2	2	SUR	9	22	31	2	2	SUR	9	22	31	2	2	SUR	16	13	29	1	0								
78	Slovakia	SVK	5	18	23	1	0	0	0	5	15	20	1	0	SVK	5	16	21	1	0	SVK	5	16	21	1	0	SVK	5	12	13	1	0								
79	Slovenia	SVN	1	19	20	1	0	0	0	1	17	18	1	0	SVN	1	18	19	1	0	SVN	1	18	19	1	0	SVN	1	12	13	1	0								
80	Thailand	THA	18	32	50	4	3	0	0	24	23	47	3	3	THA	20	28	48	4	3	THA	20	28	48	4	3	THA	31	12	43	1	0								
81	Turkey	TUR	5	45	50	3	2	0	0	3	44	47	2	1	TUR	3	45	48	2	1	TUR	3	45	48	2	1	TUR	22	21	43	1	0								
82	Tanzania	TZA													TZA						TZA						TZA	4	9	13	1	0								
83	Uganda	UGA†	26	3	29	0	1	0	0	23	3	26	0	1	UGA†	24	3	27	0	1	UGA	24	3	27	0	1	UGA	3	10	13	1	0								
84	Ukraine	UKR	4	17	21	1	0	0	0	1	17	18	1	0	UKR	1	18	19	1	0	UKR	1	18	19	1	0	UKR	3	10	13	1	0								
85	Uruguay	URY	13	35	48	1	1	0	0	12	35	47	1	1	URY	12	36	48	1	1	URY	12	36	48	1	1	URY	24	19	43	1	0								
86	Venezuela	VEN†	2	48	50	0	1	0	0	3	44	47	0	1	VEN†						VEN						VEN													
87	South Africa	ZAF	33	17	50	1	0	0	0	33	14	47	1	0	ZAF						ZAF						ZAF	33	10	43	1	0								
88	Zambia	ZMB	27	20	47	1	0	0	0						ZMB						ZMB						ZMB	27	13	40	1	0								
89	Zimbabwe	ZWE	26	9	35	1	1	0	0						ZWE						ZWE						ZWE													
Totals	83	1,568	1,484	3,052	105	58	68	68	68	1,283	1,190	2,473	81	47	58	1,150	1,104	2,254	79	44	54	955	735	1,690	54	0														

Notes: This table presents the sample make-up of the static regression models for the four alternative definitions of democracy (ANRR, BMR, CGV, PS). ‘In Democracy’ reports the number of observations in democracy (1) and autocracy (0) per country as well as the total observation count (obs). ‘Events’ refer to democratisations (D) and reversals to autocracy (A). For each of the four definitions a bold country isocode indicates that the country is included in the treatment sample. A number of countries only have reversals to democracy but no democratisation events — these are highlighted using †. Note that the PS sample is made up of countries which ‘permanently’ transitioned to democracy only.

B Schematic Review of the Literature

In Table B-1 I provide a schematic review of the empirical literature on democracy and growth.³⁹ This body of work can be categorised using two criteria: first, by the nature of the democracy proxy adopted, either in form of a continuous variable, or in form of a dichotomous variable; and second, by the identification strategy. Both of these criteria seem to follow a certain chronology, so this will be the main structural feature of this brief review.

Work published in the 1990s always adopts continuous variables for democracy (Bollen Index, Freedom House, early Polity data), in combination with either simple IV strategies arising from the panel structure (lagged variables as instruments) or even plain least squares. These studies show a wide range of results, typically pointing to a non-linear (concave) relationship between democracy and growth or no relationship at all. Papers published in the early 2000s adopt more refined democracy indicators or experiment with democracy stock variables, at times concluding a positive democratic dividend (Baum and Lake, 2003; Gerring et al, 2005); however, when implementation was more plausibly able to identify a *causal* relationship, such as in the work by Giavazzi and Tabellini (2005), the results become very fragile or disappear.⁴⁰ The latter authors were also among the first to adopt a dummy variable for democratisation, which became the standard in the economics literature thereafter (e.g. Rodrik and Wacziarg, 2005; Persson and Tabellini, 2006).⁴¹ The first paper to make the dummy variable approach ‘work’ was the study by Papaioannou and Siourounis (2008), who found strongly positive growth effects for democratisation — since many sample characteristics are not dissimilar to those in the Giavazzi and Tabellini (2005) paper, who had failed to find robust positive effects, this seemed to highlight the importance of careful construction of democracy dummies, comparing indices across a number of data sources. The same is still true for the most recent democracy-dummy paper by Acemoglu et al (2019) — their paper furthermore adopts a number of empirical strategies which in their sum total are argued to address the problems inherent in cross-country analysis (endogeneity, dynamics, linearity assumptions).

The more recent contributions adopting continuous democracy indicators tended to adopt the Arellano and Bond (1991, AB) or Blundell and Bond (1998, BB) estimators to argue for causal identification: the positive result of Knutsen (2013) in a small post-WWII sample of 44 countries using AB were undermined by the results for 69 countries in Murin and Wacziarg (2014) adopting BB. The latest contribution to this strand of the literature by Madsen et al (2015) adopts IV estimation (linguistic distance-weighted foreign democracy) to yield robustly positive and large effects for democratic change in historical and post-WWII samples.

Hence both strands adopting dichotomous and continuous measures for democratic change in the most recent iterations have yielded positive, large, and statistically significant causal effects.

³⁹Many of these studies, in particular the early work, carried out analysis of the growth-democracy as well as the democracy-growth relationship. More generally, while I do not present all results from all papers I believe the selection below is representative of the respective study. This is a snapshot of the main contributions in political science and economics; a broader literature and surveys are discussed in Dodsworth and Ramshaw (2021).

⁴⁰In terms of implementation the study by Tavares and Wacziarg (2001) is distinct from all others discussed, and while this does not diminish their contribution, it makes it difficult to compare with the other papers reviewed.

⁴¹The exception here is Persson and Tabellini (2009) who construct ‘democratic capital’ stock.

Table B-1: Literature on Democracy and Growth

Reference	Method	Democracy	Dep. variable	Specification	Sample	Results	Details
Helliwell (1994)	2SLS (lagged levels)	Continuous , Bollen index	Δ GDPpc 1960-85	GDP pc (log), investment, schooling (all in 1960, restrictions imposed following MIRW values)	N=n=90, 1960-85 (time-averaged or base year values)	– (insign.)	Table 3[2]
Barro (1996)	2SLS (lagged levels)	Continuous , Bollen and Gastil (Freedom House) indices	Δ GDPpc in non-overlapping 5-year periods	Elaborate controls, lagged levels as instruments	N=89, 1960-90	– (insign.)	Table 1[2]
	2SLS (lagged levels)	Continuous , Bollen and Gastil (Freedom House) indices, level and squared terms	dto.	dto.	N=89, 1960-90	concave (5% level)	Table 1[4]
Leblang (1997)	OLS w/ period FE	Continuous , institutionalised democracy from Polity II, lagged	Decadal average Δ GDPpc	GDP pc (log), primary and secondary school attainment (all in decade start year)	n=232, 1960-89	+ (5% level)	Table 2[2]
Minier (1998)	2SLS (lagged levels)	Continuous , Gastil (Freedom House) index, level and squared terms; dummies for positive and negative changes in democracy	Δ GDPpc in non-overlapping 5-year periods	GDP pc (log), schooling attainment, (all lagged by 5 years)	n=485, 1960-89	concave (5% level); <i>insig. +ve changes,</i> sig. –ve changes	Table 2[1]
Tavares & Wacziarg (2001)	3SLS	Continuous , Bollen and Freedom House indices	Annual Δ GDPpc	HC, inequality, instability, distortions,...	N=n=65, 1970-89	– (1% level)	Table 3[4]

Table continued overleaf

Table B-1: Literature on Democracy and Growth (continued)

Reference	Method	Democracy	Dep. variable	Specification	Sample	Results	Details
Baum & Lake (2003)	OLS w/ country FE	Continuous , Polity 98 index	Annual Δ GDPpc	GDP pc (log), life expectancy, investment, labour force, HC (all lagged), various lags of Δ GDPpc	N=128, n=548, 1967-97	+ (<i>insign.</i>)	Table 1[2]
Gerring et al (2005)	OLS w/ country FE	Continuous , Democracy stock (1900-2000) based on continuous polity2 (-10, +10)	Annual Δ GDPpc	Lagged GDP pc (log)	N=180, n=6,264, 1950-2000	+ (1% level)	Table 2[1]
	dto.	Continuous , Democracy stock (1900-2000) based on dummy (= 1 if polity2 > 4)	Annual Δ GDPpc	Lagged GDP pc (log)	N=180, n=6,264, 1950-2000	+ (1% level)	Table 2[6]
Giavazzi & Tabellini (2005)	Diff-in-Diff (OLS w/ year FE)	Dummy for polity2 > 0, all democratisation	Annual Δ GDPpc	Dummy for socialist regimes (interacted with democratisation), continent dummies	N=138, n=4,388, 1960-2000	+ (10% level)	Table 1[7]
	Diff-in-Diff (OLS w/ year FE)	Dummy for polity2 > 0, permanent democratisation	Annual Δ GDPpc	Dummy for socialist regimes (interacted with democratisation), continent dummies	N=138, n=4,387, 1960-2000	+ (<i>insign.</i>)	Table 1[8]
Rodrik & Wacziarg (2005)	OLS w/ country FE	Dummy : New Democracy, Established Democracy, etc (Polity IV-based)	Annual Δ GDPpc	Dummies for different regimes (new, established)	N=154, n=5,649, 1950-2000	+ (5% level) SR effect for democratisation in the past 5 yrs	Table 1[3]

Table continued overleaf

Table B-1: Literature on Democracy and Growth (continued)

Reference	Method	Democracy	Dep. variable	Specification	Sample	Results	Details
Person & Tabellini (2006)	OLS w/ period FE	Dummy for polity2 > 0	Annual Δ GDPpc	Continent dummies, legal origin, lagged GDP pc (log)	N=138, n=4,338, 1960-2000	+ (5% level)	Table 1[1]
	dto.	dto.	Annual Δ GDPpc	Continent dummies, legal origin, lagged GDP pc (log)	N=148, n=8,135, 1850-2000	+ (10% level)	Table 3[3]
Person & Tabellini (2009)	2FE	Continuous, domestic and foreign democratic stock (PIM, 1800-2000) based on dummy (= 1 if polity2 > 0)	Annual Δ GDPpc	lagged GDP pc (log)	n=8,379, 1820-2000	+ (1% level) domestic; <i>in-sign. foreign</i>	Table 5[1]
	Papaioannou & Siourounis (2008)	Dummy building on FHI and polity2	Annual Δ GDPpc	none	N=166, n=5,410, 1960-2005	+ (1% level)	Table 2[4]
Knutsen (2013)	Diff-in-Diff	dto.	Annual Δ GDPpc	lagged GDP pc (log) and lagged growth rate, investment	N=166, n=5,410, 1960-2005	+ (1% level)	Table 3[1]
	OLS w/ period FE	Continuous, Freedom House Index*	Annual Δ GDPpc	lagged GDP pc, population, regime duration (all in log)	N=44, n=1,289, 1972-2004	+ (1% level)*	Table 2[2]
	GMM AB	dto.	Annual Δ GDPpc	lagged GDP pc, population, regime duration (all in log)	N=44, n=1,234, 1972-2004	+ (1% level)*	Table 2[2]

Table continued overleaf

Table B-1: Literature on Democracy and Growth (continued)

Reference	Method	Democracy	Dep. variable	Specification	Sample	Results	Details
Murtin & Wacziarg (2014)	2FE	Continuous, re-scaled lagged polity2	Decadal GDPpc (log)	GDP pc (log), lagged by a decade	N=69, n=567, 1870-2000	+ (<i>insign.</i>)	Table 11[1]
	2FE	dto.	dto.	dto.	N=69, n=308, 1960-2000	- (<i>insign.</i>)	Table 11[10]
	GMM BB	dto.	dto.	dto.	N=69, n=489, 1870-2000	+ (<i>insign.</i>)	Table 11[3]
GMM BB	dto.	dto.	dto.	dto.	N=68, n=275, 1960-2000	+ (<i>insign.</i>)	Table 11[12]
Madsen et al. (2015)	2SLS-2FE	Continuous, re-scaled polity2	Decadal average GDPpc (log)	lagged GDP pc (log); IV linguistic-distance weighted democracy	N=141, n=1,143, 1820-2000	+ (5% level); 1sd → +96%	Table 4[1]
	2SLS-2FE	dto.	Decadal average GDPpc (log)	dto.	N=141, n=595, 1950-2000	+ (5% level)	Table 4[3]
Acemoglu et al (2019)	2FE	Dummy for polity2 > 0 plus other conditions	Annual GDPpc (log)	4 lags of GDP pc (log)	N=175, n=6,790, 1960-2010	+ (1% level); 21.2% LR effect	Table 2[3]
	GMM AB	dto.	dto.	dto.	N=175, n=6,161, 1960-2010	+ (5% level); 16.5% LR effect	Table 2[7]
	2SLS	dto.	dto.	dto., IV regional waves of democratisation	N=174, n=6,309, 1960-2010	+ (10% level); 31.5% LR effect	Table 6[2], Panel A
Non-para	dto.	dto.	dto.	4 lags of GDP pc (log)	1960-2010	+ (1% level); 23.7% (20-24 yrs)	Table 5 [6], Panel C

Notes: The table presents a subset of empirical results from the literature on democracy and growth. It is important to emphasise that I do not report all relevant results from each paper, but selectively picked the most general and most representative ones in each case. I highlight the distinction between a continuous and dichotomous proxy for democracy in bold. Regarding empirical results in the final column of the table, insignificant estimates are in italics, statistically significant estimates in bold. N refers to the number of countries, n the number of observations (if either or both are missing then it was not clearly reported in the study with reference to the specific result I present here), the time period of the sample is also indicated. * The FHI has a reversed scale compared with polity2, but here I adjust the 'democracy effect' (result) to be in line with logic of other indicators (higher value = more democracy). The final columns reports the Table and column for the result in the respective paper.

C Additional Results

C.1 Main Results — Static Specification

Table C-1: Main Results – Static Specifications

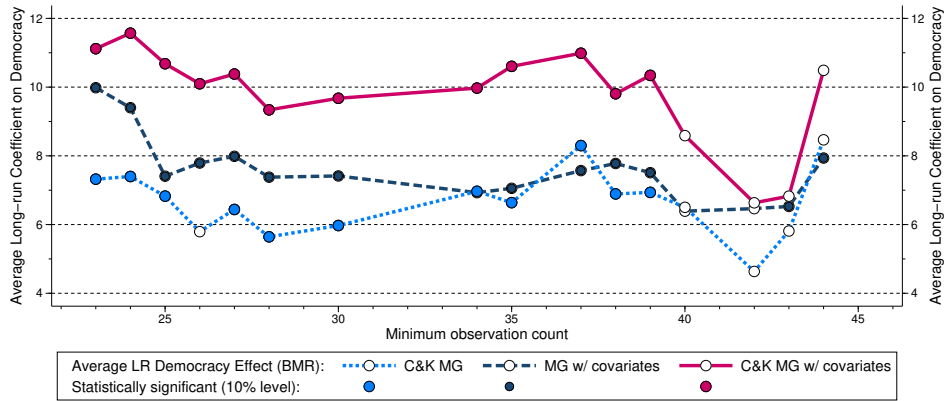
	Plain Vanilla		With Covariates	
	(1)	(2)	(3)	(4)
Implementation	MG	C&K MG	MG	C&K MG
Parameters estimated ‡	$2 \times N$	$5 \times N$	$4 \times N$	$7 \times N$
(a) Democracy (ANRR)	10.249 (3.277) ^{***}	4.402 (2.088) ^{**}	3.846 (2.598)	4.016 (1.983) ^{**}
Observations	3052	3052	3052	3052
Countries (N)	83	83	83	83
Democratisations	105	105	105	105
Reversals	58	58	58	58
Avg Years in Dem	17.9	17.9	17.9	17.9
RMSE	21.860	11.492	13.606	8.791
(b) Democracy (BMR)	10.629 (3.394) ^{***}	4.168 (2.242) [*]	5.505 (2.847) [*]	4.260 (2.166) [*]
Observations	2473	2473	2473	2473
Countries (N)	68	68	68	68
Democratisations	81	81	81	81
Reversals	47	47	47	47
Avg Years in Dem	18.5	18.5	18.5	18.5
RMSE	21.512	10.542	12.808	7.872
(c) Democracy (CGV)	12.849 (3.739) ^{***}	2.862 (2.730)	6.853 (2.837) ^{**}	4.991 (2.383) ^{**}
Observations	2254	2254	2254	2254
Countries (N)	58	58	58	58
Democratisations	79	79	79	79
Reversals	44	44	44	44
Avg Years in Dem	19.0	19.0	19.0	19.0
RMSE	22.725	10.550	13.361	8.236
(d) Democracy (PS)	21.990 (4.636) ^{***}	4.669 (3.146)	11.296 (3.538) ^{***}	4.874 (2.912)
Observations	2057	2057	2057	2057
Countries (N)	54	54	54	54
Democratisations	54	54	54	54
Reversals	0	0	0	0
Avg Years in Dem	20.3	20.3	20.3	20.3
RMSE	19.920	11.214	13.143	8.892

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.

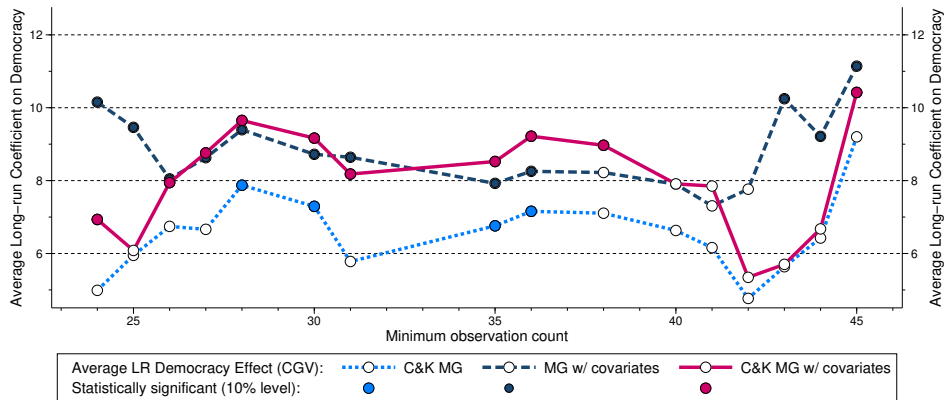
D Sample Reduction Exercises – More Results

D.1 Alternative Definitions of Democracy

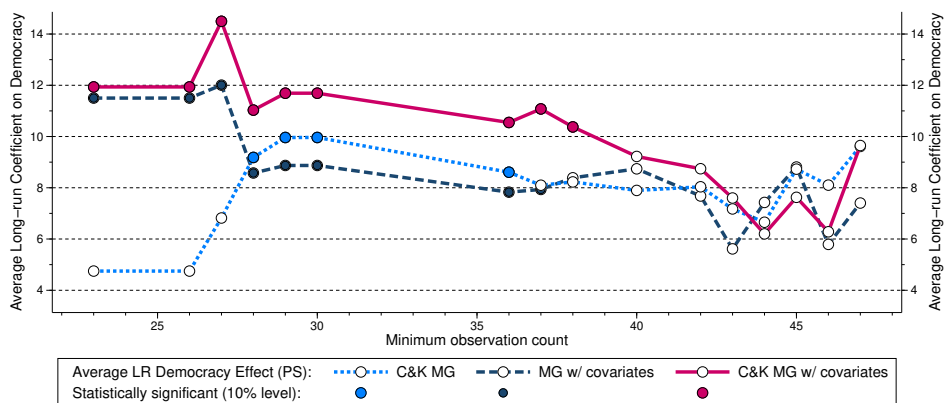
Figure D-1: Sample Reductions — minimum T_i



(a) Boix et al (2013)



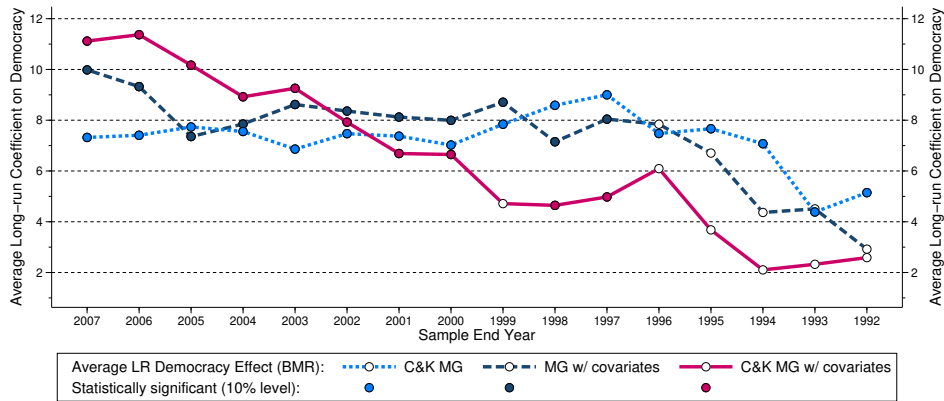
(b) Cheibub et al (2010)



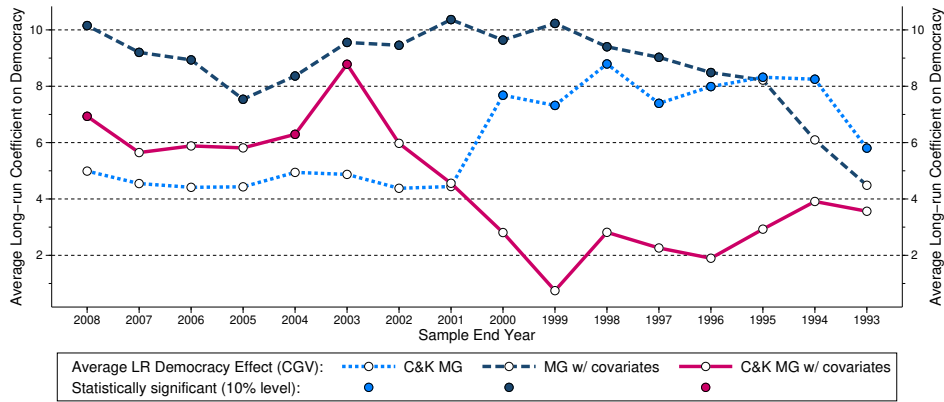
(c) Papaioannou and Siourounis (2008)

Notes: This figure provides sample reduction results for the static and dynamic Diff-in-Diff estimators using the alternative definition for democracy as indicated. This figure needs to be contrasted with Panel (a) of Figure 3 for a comparison with the results for the ANRR definition of democracy.

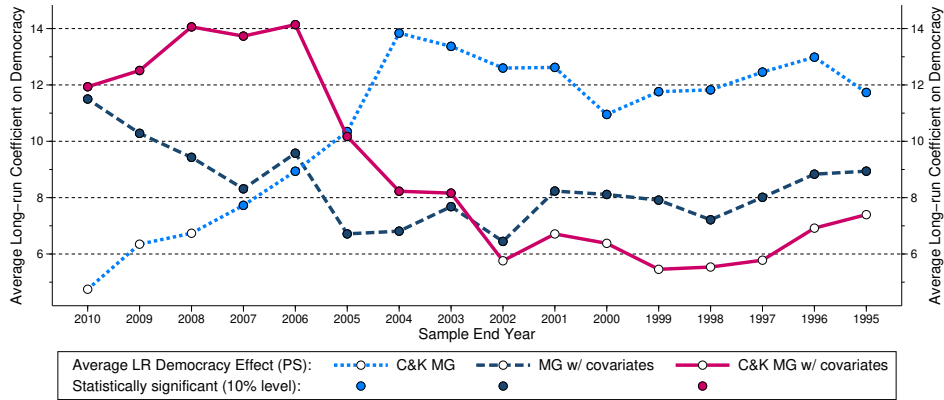
Figure D-2: Sample Reductions — end year



(a) Boix et al (2013)



(b) Cheibub et al (2010)



(c) Papaioannou and Siourounis (2008)

Notes: This figure provides sample reduction results for the static and dynamic Diff-in-Diff estimators using the alternative definition for democracy as indicated. This figure needs to be contrasted with Panel (b) of Figure 3 for a comparison with the results for the ANRR definition of democracy.

E Sample Reduction Exercises – ANRR and Madsen et al (2015)

E.1 ANRR

In this section I discuss results from two sample reduction exercises presented in Figure E-2. Table 4 summarizes the estimates and sample makeup of five *ad hoc* ‘thresholds’ in the long-run estimates for democracy: in Panel A for the full ANRR sample, in B the sample which yields an insignificant estimate, in C when the estimate falls below 5% in magnitude (less than one quarter of the full sample result), in D when the reduced sample estimate is outside the confidence interval of the full sample one, and in E the balanced panel estimate. Columns [1]-[4] and [5]-[8] are for the respective sample reduction strategies. Using results in Figure ?? I speculate about one potential source of the patterns observed.

ANRR adopt a variety of empirical implementations for an empirical model which captures country-specific fixed effects and the dynamics of per capita GDP:⁴²

$$y_{it} = \alpha_i + \gamma_t + \beta \text{Democracy}_{it} + \sum_{\ell=1}^p \rho_{\ell} y_{i,t-\ell} + \varepsilon_{it}, \quad (6)$$

where y is log per capita GDP (multiplied by 100), Democracy is a dummy variable, α_i and γ_t are country and time dummies, respectively, and ε is the error term.⁴³ In order to allow for a causal interpretation of the results they devise an instrumentation strategy which builds on regional waves of democratisation and reversal. The findings from these 2SLS models are shown to be in line with results adopting country fixed effects (2FE), the Arellano and Bond (1991, AB) and the Hahn, Hausman, and Kuersteiner (2001, HHK) estimators.

Sample reduction by minimum observation count I begin with the strategy which drops countries by their sample observation count. A major concern for this non-random sample reduction strategy is that even though the ‘small- T ’ countries may only account for a very small share of overall observations they may represent a disproportionate share of the democratisation and reversal events. If this were the case then the sample reduction strategy *by construction* makes it harder and harder for the estimators to identify a democracy effect. The histogram in Panel (a) of Figure E-1 speaks to this concern — this plot is based on the AB/HHK sample (the 2SLS sample typically has one additional observation per country), detailed information about the countries dropped in these sample reduction exercises are contained in an Appendix. Along the x -axis we can see the minimum observation count for inclusion in the sample; the thin gray bars indicate the total observation count (left scale, in logarithms). This highlights that over 60% of the full sample (around 4,000 observations) have data for all years, and for reference I report the results for this ‘balanced panel’ below. The coloured bars indicate the distribution of democratisation and reversal events by minimum observation count: again roughly 60%

⁴²My presentation is limited to the parametric results. The semi-parametric results for sample reduction strategy (i) yield confidence intervals which always include zero when around 20% of observations are omitted; for strategy (ii) results appear much less affected, if anything confidence intervals become *tighter* as respective end years are omitted. The source of this robustness relative to all other ANRR results is beyond the scope of this note, results are relegated to the Online Appendix.

⁴³ANRR test a variety of lag structures (p) but favour the specification with four lags.

of these events occur in the balanced panel sample, while the remainder are sprinkled thinly across other minimum observation samples.

Panel (b) presents the full and reduced sample results for the FE, AB, HHK and 2SLS estimators — all results are for the specification with four lags of GDP, which is preferred by ANRR.⁴⁴ In this and the equivalent plot in Panel (b) of Figure ?? a filled coloured (white) circle indicates statistical (in)significance at the 10% level. The left-most estimates correspond to the full sample results reported in the ANRR paper, the right-most to the estimates for a balanced panel. The x -axis is identical to the plot in panel (a), the y -axis indicates the long-run effect (in percent) of democracy on per capita GDP. For the 2FE estimator this sample reduction exercise has virtually no impact on the long-run democracy estimate: as we move to the right countries with fewer observations than the minimum number indicated on the x -axis are omitted from the regression sample, but the 2FE long-run estimate for democracy is virtually unchanged. The exception is the balanced panel result which is statistically insignificant, though at 15.6% still reasonably close to the full sample estimate of 21.2%.⁴⁵

The patterns for the AB and HHK estimates are very different: both decline and turn statistically insignificant when the minimum observation count is 17 and thereafter fall (more or less monotonically) towards and beyond zero. Results in Columns [2] and [3] of Table 4 indicate that the AB and HHK estimates are statistically insignificant and reduced by a quarter and two-thirds, respectively, once 5% of the full sample observations are dropped. The balanced panel results for these two estimators (-5.3 and -12.4) are derived from a sample where just over 40% of observations are dropped.

Democracy estimates based on the 2SLS estimator initially maintain a high and stable level in excess of 30%, but turn insignificant once countries with fewer than 21 observations are omitted (7% of the full sample of 6,300 observations). The magnitude of 2SLS estimates drops quite rapidly, such that it falls below 5% in magnitude and also outside the full sample 90% confidence interval once 18% of observations are dropped. In contrast to the patterns for the AB and HHK estimators the 2SLS estimates increase again if further countries are dropped.

Two aspects are worth emphasising comparing these findings to the results in ANRR: first, the parity between results for the within estimator on the one hand, and the AB, HHK and 2SLS estimators on the other, as presented in Tables 2 and 6 of ANRR, is not given in my sample reduction exercises: the within estimates clearly deviate from all others and the “triangulation of evidence” (ANRR: 8) is thus not given; second, all of the estimators intended to address endogeneity concerns show rapidly declining, at times even negative, long-run growth implications of democracy as the sample is reduced.⁴⁶

⁴⁴Results for one and two lags are presented in an Appendix, where I also provide 2SLS estimates for the alternative construction of the long-run estimate with qualitatively identical results.

⁴⁵Note that many researchers have serious reservations about the fixed effects estimator for causal inference in panel data (e.g. Gibbons, Suarez-Serratoz and Urbancic, 2019; Imai and Kim, 2019). A recent paper by Chen, Chernozhukov and Fernandez-Val (2019, CCF-V) builds on ANRR and employs AB and FE estimators but with bias-correction for the many instruments and incidental parameter problems, respectively, confirming the AB/FE results of ANRR. Note however that CCF-V’s sample choice (balanced panel from 1987-2009) leads to long-run estimate for democracy of 179.4 ($t=1.57$) if I adopt the ANRR 2SLS estimator!

⁴⁶ANRR note that the long-run estimates computed from their dynamic regressions are subject to small sample (attenuation) bias. Increasing the average time-series of sample countries by discarding countries with few observations should if anything *reduce* this bias and thus cannot account for the findings of my sample reduction exercise.

Sample reduction by sample end year Figure E-2 presents the results when observations are omitted by sample end year. The primary focus here is on the impact of the Global Financial Crisis in 2007/8 and its aftermath. Panel (b) of Figure E-1 charts the distribution of sample observations and democratisation/reversal events by year — here and in panel (b) of Figure E-2 the x -axis is in reverse chronological order. We can see that the annual sample observation count rises from the 1960s until peaking in the mid-2000s. The final three sample years 2008-10 account for around 8% of all observations (2010: 3%, 2009: 2%, 2008: 3%). The first 25 years of the sample indicate typically two to three democratisation/reversal events per annum, before a wave of events in the early 1990s following the collapse of the Soviet Union. The final three sample years 2008-10 indicate 14 events, around 9% of the total number of events over 1965-2010.⁴⁷

Panel (b) of Figure E-2 presents the sample reduction results, where the x -axis indicates the final year included in the sample, and the y -axis indicates the long-run effect (in percent) of democracy on per capita GDP — again all estimates are for the 4-lag specification preferred by ANRR. I only chart end years down to 1995, since omitting 1996-2010 amounts to around 40% of observations, similar to the 40% of observations omitted in the balanced panel of the ‘small T_i ’ exercise presented above.

As before the 2FE estimates are found to be fairly robust to sample reduction, only turning insignificant when 30% of observations are dropped. The AB/HHK estimates, in contrast, turn insignificant if the post-GFC years 2009 and 2010 are omitted, thereafter declining and eventually diverging, with HHK remaining positive (albeit insignificant throughout) while AB estimates turn negative (dto.). The 2SLS estimates are generally falling with earlier sample end years, but display curious patterns in the aftermath of the GFC: omitting only 2010 (3% of observations) yields a statistically insignificant long-run coefficient on democracy. Omitting both 2010 and 2009 (together 5% of observations) however restores the full sample coefficient in terms of magnitude and statistical significance, whereas the omission of further end years always yields statistically insignificant long-run democracy estimates. Table 4 provides all the details on estimates, standard errors and samples of the various ‘thresholds’ as defined above.

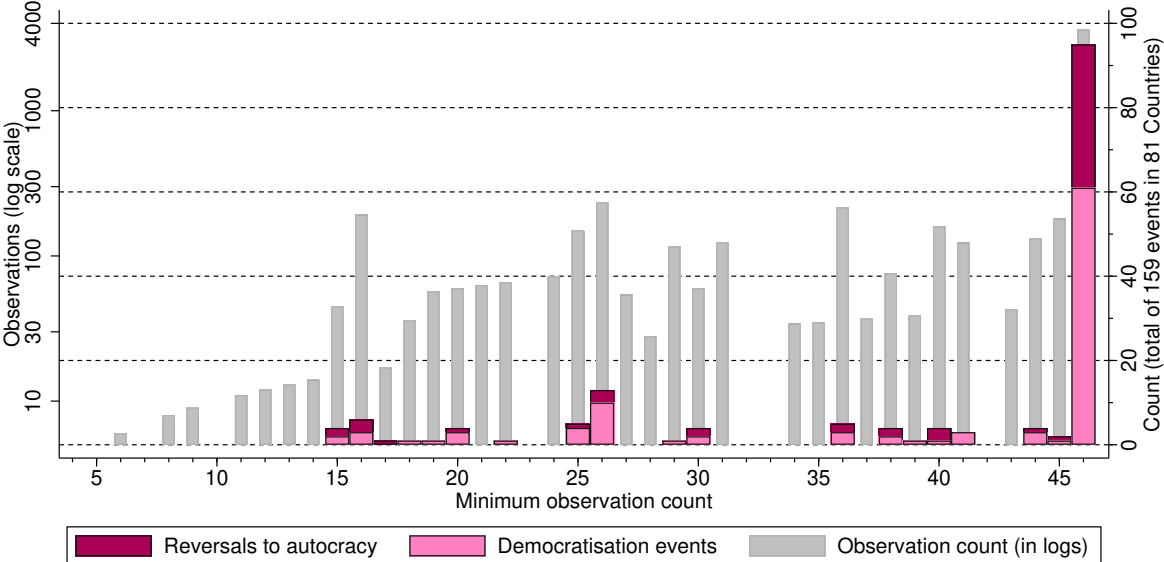
Sample reduction by trial and error The focus of the sample reduction exercises is primarily on the magnitudes of estimated coefficients, though statistical insignificance *can* indicate that underlying country estimates are heterogeneous and vary substantially across countries. If the focus of the exercise were more narrowly on the smallest sample reduction yielding a statistically insignificant long-run estimate for democracy, then the number of countries that would need to be dropped is very small: three for AB/HHK and four for 2SLS, amounting to fewer than 1% of observations in each case — see Table E-3.

Recent work by Young (2018) has highlighted the fragility of IV estimates, demonstrating that many findings of statistical significance are driven by few observations. Here, it should be emphasised that the results derive from a purposeful exercise in sample selection (by trial and error), and further dropping a small number of countries may similarly *restore* the statistical significance of the estimates. Nevertheless, in practical terms as well as conceptually, it is worrisome that empirical results of a supposedly ‘robust’ democracy-growth nexus can be made

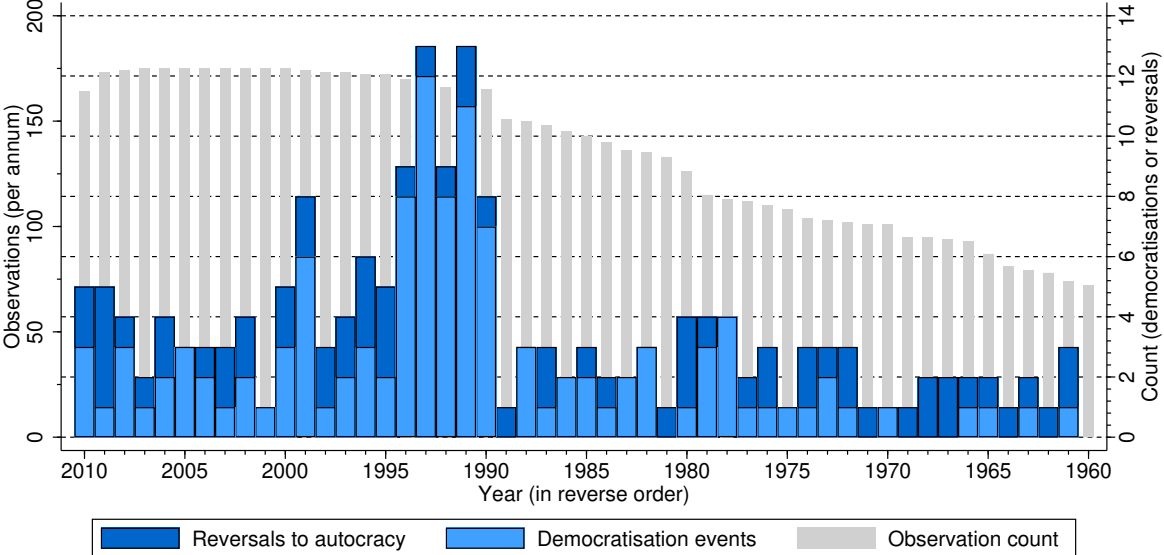
⁴⁷This is once again the AB/HHK sample for the four-lag specification, hence the 1965 start year.

to (statistically-speaking) disappear by the omission of three former Soviet Republics with 20 observations each, two of which (Turkmenistan and Uzbekistan) have no experience of democracy and the third (Ukraine) only has three sample years in autocracy.

Figure E-1: Sample and Event Distribution – ANRR



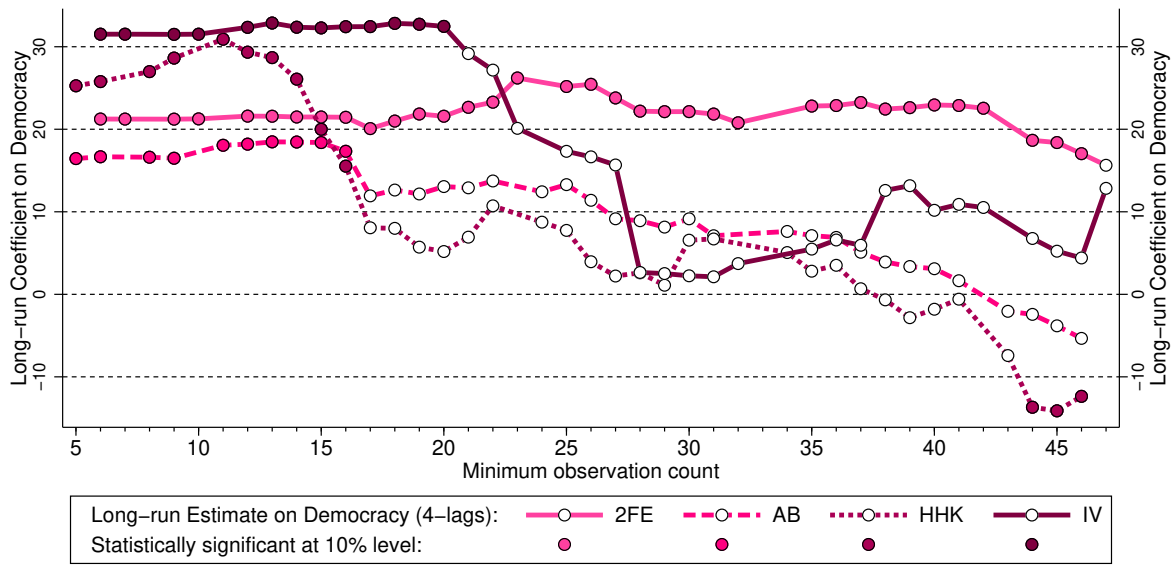
(a) Sample Reduction by T_i count



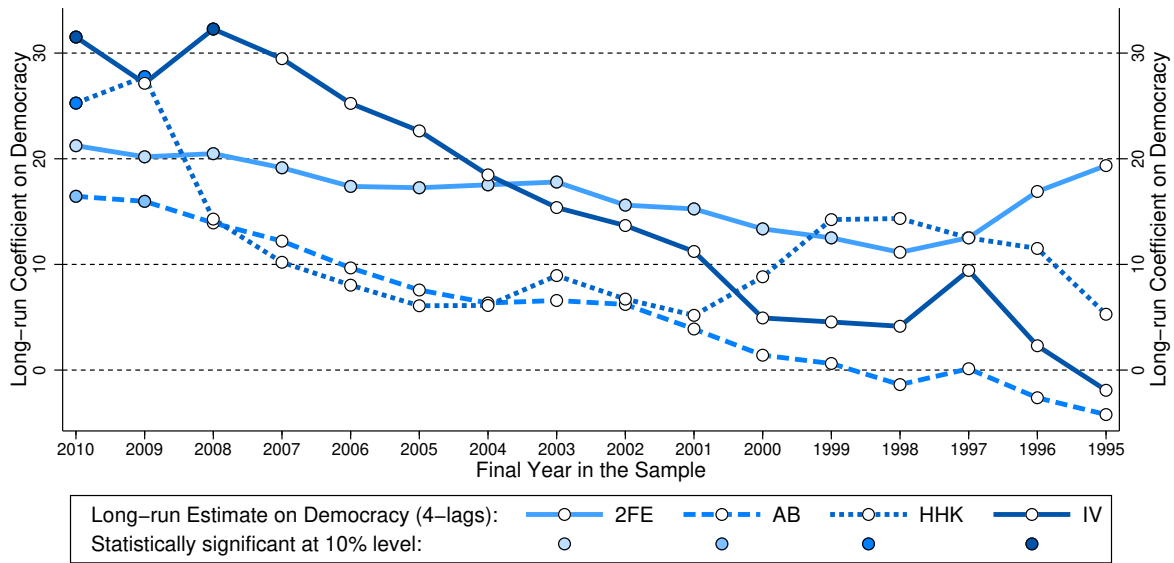
(b) Sample Reduction by end year

Notes: The figure presents the sample distribution for democratisation from varying empirical samples. The x -axis in panel (a) indicates the minimum number of observations required to be included in the sample, in panel (b) the sample end year (in reverse chronological order). The thin gray bars indicate the distribution of observations (log scale in panel (a), left axis) while the coloured bars indicate democratisation and reversal events (right scale). These distributions are for the AB/HHK samples.

Figure E-2: Sample Reductions — ANRR



(a) Sample reduction by T_i count



(b) Sample reduction by end year

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%.

Table E-1: Regression Sample — ANRR data (AB/HHK 4-lag specification)

obs	Transitioned into/out of democracy							Never a democracy			Always a democracy			
5								QAT						
6								LBY						
8								KWT						
9								IRQ						
11								MDV						
12								BIH						
13								KHM						
14								ERI						
15	DJI	HTI									PLW			
16	ARM	HRV	SLB					AZE	BLR	KAZ	CZE	LTU	MKD	
								YEM			POL SVN			
17	RUS													
18	LBN							TZA						
19	UKR							TKM UZB						
20	GIN	KGZ									NAM			
21								AGO	GNQ	TJK				
22	SVK							LAO	VNM					
24								BHR	UGA	WSM				
25	BTN	CPV	ETH	MNG				BRN	TON					
26	ALB	BGR	COM	EST	MDA	MOZ	ROM				CHE	LCA		
27											KNA	VUT		
28											NZL			
29	GRD										ATG	BLZ	DMA	
30	SUR										MUS			
31								JOR			CYP	KIR	VCT	
34								SYC						
35											PNG			
36	GNB							CUB	SWZ			DEU	IRL	MLT
37											BHS			
38	BGD							SAU						
39	MLI													
40	FJI	GMB						IRN			JAM			
41	GEO	HUN	LVA											
43											BRB			
44	GUY	LSO									BWA			
45	ZWE							OMN	SIN	TUN				
46	ARG	BDI	BEN	BFA	BOL	BRA	CAF	CHN	CMR	DZA	AUS	AUT	BEL	
	CHL	CIV	COG	DOM	ECU	ESP	GHA	EGY	GAB	MAR	CAN	COL	CRI	
	GRC	GTM	HND	IDN	KEN	KOR	LBR	MYS	RWA	SYR	DNK	FIN	FRA	
	MDG	MEX	MRT	MWI	NER	NGA	NIC	TCD	TGO	ZAR	GBR	IND	ISL	
	NPL	PAK	PAN	PER	PHL	PRT	PRY				ISR	ITA	JPN	
	SDN	SEN	SLE	SLV	THA	TUR	URY				LKA	LUX	NLD	
	VEN	ZAF	ZMB								NOR	SWE	TTO	
											USA			

Notes: The three samples contain 80, 46, and 49 countries, respectively. The analysis is based on the AB/HHK samples; for the 2SLS estimates the minimum observation count is typically increased by one observation.

Table E-2: Sample Reduction Estimates — ANRR

Estimator	Sample reduction by T_i count				Sample reduction by end year			
	[1] 2FE	[2] AB	[3] HHK	[4] 2SLS	[5] 2FE	[6] AB	[7] HHK	[8] 2SLS
Panel A: Full ANRR sample estimates								
Long-Run Democracy Effect	21.240 [7.215]***	16.448 [8.436]*	25.268 [10.869]**	31.521 [17.425]*	21.240 [7.215]***	16.448 [8.436]*	25.268 [10.869]**	31.521 [17.425]*
min T_i /End year	6	5	5	6	2010	2010	2010	2010
Countries	175	175	175	174	175	175	175	174
Observations	6,336	6,161	6,161	6,309	6,336	6,161	6,161	6,309
Share of ANRR 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 Democracy Effect	15.637 [9.867]	11.932 [8.071]	8.066 [7.047]	29.168 [17.733]	12.516 [7.386]	3.891 [8.131]	14.293 [11.504]	27.145 [17.309]
min T_i /End year	47	17	17	21	1999	2001	2008	2009
Countries	79	152	152	146	172	172	175	174
Observations	3,713	5,846	5,846	5,873	4,433	4,605	5,824	6,146
Share of ANRR sample	0.59	0.95	0.95	0.93	0.70	0.75	0.95	0.97
Panel C: Estimate below 5% in magnitude								
Long-Run Democracy Effect	n/a	3.918 [7.622]	3.949 [5.670]	2.651 [16.519]	1.160 [6.157]	3.891 [8.131]	-22.917 [28.970]	4.936 [17.275]
min T_i /End year		38	26	28	1991	2001	1994	2000
Countries		97	128	119	149	172	152	171
Observations		4,387	5,325	5,202	3,119	4,605	3,422	4,588
Share of ANRR sample		0.71	0.86	0.82	0.49	0.75	0.56	0.73
Panel D: Estimate outside 90% CI of full sample estimate								
Long-Run Democracy Effect	n/a	1.650 [8.722]	5.718 [6.287]	2.651 [16.519]	1.160 [6.157]	1.411 [8.409]	6.091 [8.090]	2.305 [23.466]
min T_i /End year		41	19	28	1991	2000	2005	1996
Countries		90	149	119	149	172	175	166
Observations		4,112	5,793	5,202	3,119	4,433	5,300	3,908
Share of ANRR sample		0.67	0.94	0.82	0.49	0.72	0.86	0.62
Panel E: Estimate for balanced panel								
Long-Run Democracy Effect	15.637 [9.867]	-5.337 [8.484]	-12.358 [6.899]*	12.843 [23.009]	n/a	n/a	n/a	n/a
min T_i	47	46	46	47				
Countries	79	79	79	78				
Observations	3,713	3,634	3,634	3,666				
Share of ANRR sample	0.59	0.59	0.59	0.58				

Notes: The table presents estimates for the two sample reduction exercises in columns [1]-[4] and [5]-[8], respectively (estimator as indicated). All estimates are based on specifications with four lags of per capita GDP and in case of the 2SLS using four lags of the instrument — these are the preferred specifications by ANRR. Long-run estimates are 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 computed via the Delta method). Results in Panel A are identical to those in ANRR Tables 2 (2FE, AB, HHK) and 6 (2SLS). The 2FE estimate in column [1] never drops below 5% in magnitude or outside the 90% confidence interval of the full sample estimate. 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.

Table E-3: Minimal Sample Reduction

ANRR Reference	Sample reduction by T_i count							
	[1] 2FE Table 2(3)		[2] AB Table 2(7)		[3] HHK Table 2(11)		[4] 2SLS Table 6(2)A	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Long-run effect of democracy	21.240 [7.215]***	15.637 [9.867]	16.448 [8.436]*	12.846 [8.023]	25.032 [10.581]***	9.221 [5.830]	31.521 [17.425]*	28.574 [17.394]
Observations	6,336	3,713	6,161	6,113	6,161	6,100	6,309	6,249
Obs dropped	none	2,623	none	48	none	61	none	60
dto. (in %)	0%	41.4%	0%	0.78%	0%	0.99%	0%	0.95%
Countries	175	79	175	172	175	171	174	171
Countries dropped	none	96	none	ARM, AZE, SLB	none	AZE, BLR, ERI, HTI	none	TKM, UKR, UZB
dto. (in %)	0%	54.9%	0%	1.7%	0%	2.3%	0%	1.7%

Notes: The table presents full sample estimates in columns marked (a) and reduced sample estimates in columns marked (b) for the 2FE, AB, HHK and 2SLS estimators. In a purposeful exercise I determine (via trial and error) the minimum set of countries that need to be dropped from the sample for the long-run democracy estimate to turn statistically insignificant (AB, HHK and 2SLS only). The countries dropped are indicated in the bottom of the table — for instance, the 2SLS estimate turns insignificant if Turkmenistan (TKM; 20 sample years in autocracy, none in democracy), the Ukraine (UKR; 3, 17), and Uzbekistan (UZB; 20, 0) are dropped from the sample. Statistical significance at the 10%, 5% and 1% level are indicated as *, **, and ***, respectively.

E.2 Madsen, Raschky and Skali (2015)

Their dataset of decadal observations for up to 141 countries covers 1820-2000. The empirical specification mirrors that of ANRR, though given the decadal data the dynamics are simpler (just a single lag for GDP, as opposed to four lags in ANRR), which seems intuitive. For country i and decade t (values are said to be averaged ‘within each interval’ but the range of these intervals is not entirely clear, most likely 1991-2000, 1981-1990, etc.):

$$y_{it} = \alpha_i + \gamma_t + \beta \text{Democracy}_{i,t-1} + \delta \text{HC}_{i,t-1} + \rho y_{i,t-1} + \varepsilon_{it}, \quad (7)$$

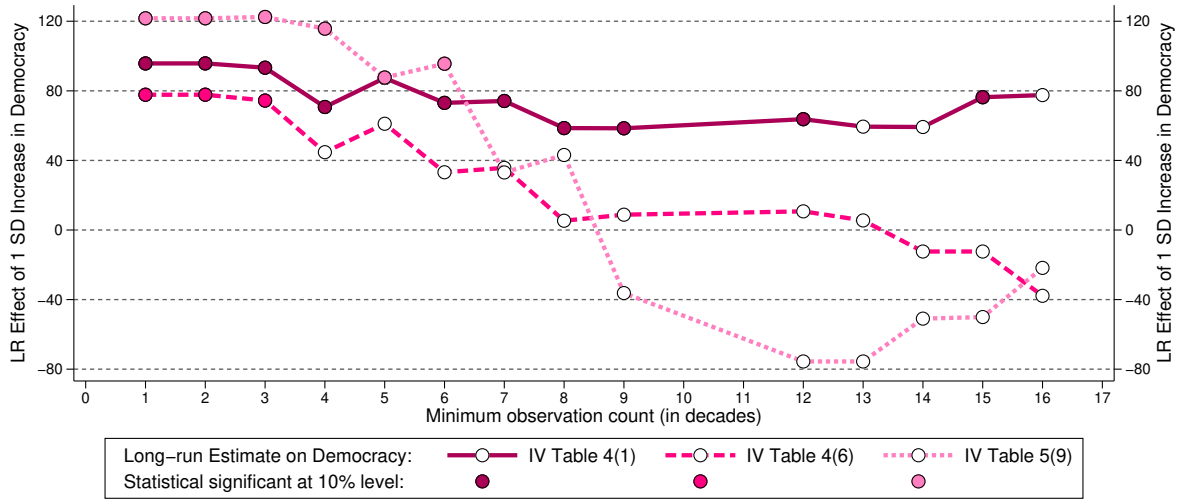
where y is the log of real per capita GDP (in PPP values), Democracy is a the (continuous) polity2 variable, α_i and γ_t are country and time dummies, respectively, and HC is human capital proxied by literacy. Democracy is instrumented using the linguistic distance-weighted average of ‘foreign’ democracy, if HC is included in the model, then it is instrumented using the interaction of minimal working age legislation (a dummy) with the number of compulsory schooling years. Alternative instruments are used in additional robustness checks.

I focus on three specifications, namely (i) a benchmark specification which excludes HC in Table 4, column 1 of the paper, (ii) the specification as presented in equation (7) in Table 4, column 6, and (iii) the same as the benchmark specification but with contemporaneous instead of lagged democracy in Table 5, column 9.

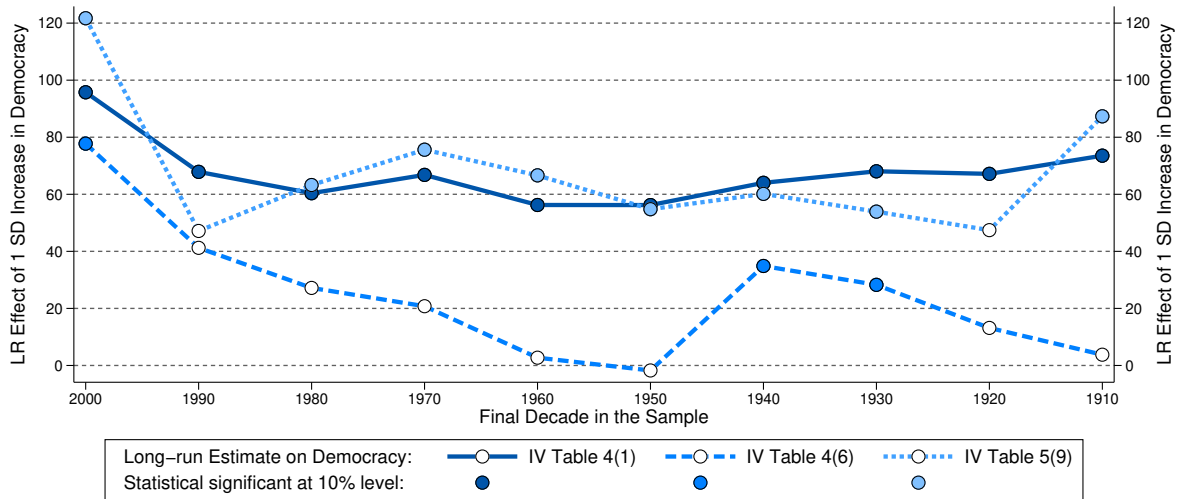
Sample reduction by minimum observation count Panel (a) of Figure E-3 provides the decade-by-decade results for the sample reduction by minimum observation count, columns [1] to [3] of Table E-4 the results for the full sample, for the sample when the democracy estimate turns statistically insignificant and the sample when the estimate falls outside the 90% confidence interval of the full sample result. While the more elaborate specification with human capital (itself also instrumented) as well as the model using the contemporaneous value of democracy drop substantially and turn insignificant when 13% and 26% of observations are omitted, respectively, the benchmark specification holds up much better, only turning insignificant when over one-third of observations are omitted, while its coefficient magnitude is also comparatively stable.

Sample reduction by sample end year In panel (b) of the same figure I present results for the second sample reduction exercise where the benchmark results (solid blue line) once again perform best: these are statistically significant through, even when the sample is reduced to 1820-1910, whilst maintaining a remarkably stable democracy effect of 60-80% higher per capita GDP for a one standard deviation increase in the democracy index. Both the models with contemporaneous democracy and the additional HC covariate see the democracy coefficient turn insignificant when a single decade, 2000, is omitted, but while the former then remains statistically significant and fairly stable (similar in magnitude to the benchmark results) the latter drops substantially and is mostly statistically insignificant.

Figure E-3: Sample Reductions — Madsen et al (2015)



(a) Sample reduction by T_i count



(b) Sample reduction by end year

Notes: The plots present long-run estimates for democracy from various specifications, computed as $\hat{\beta}^{LR} = \hat{\beta} / (1 - \hat{\rho}_{i,t-1})$, where $\hat{\beta}$ is the estimate on the (lagged or contemporaneous) democracy dummy and $\hat{\rho}$ that for the lag of per capita GDP (standard errors are constructed via the Delta method). The model for which estimates are presented by the solid line plots is for Madsen et al (2015) Table 4, Column 1 (baseline); the dashed line plots are for Table 4, Column 6, which includes lagged literacy as additional covariate; the short-dashed line plots are for Table 5, Column 9, which uses the contemporaneous term of democracy instead of its lag as in the above two specifications. The x -axis in panel (a) indicates the minimum observation count for countries to be included in the sample, in (b) the end year/decade of the sample. A filled (white) marker indicates that the coefficient on democracy is statistically (in)significant at the 10% level. In panel (a) the estimates in the three models presented turn insignificant when 35%, 10%, and 24% of observations are excluded in the models in Table 4(1), Table 4(6) and Table 5(9), respectively. In panel (b) the equivalent figures are 12% for both the latter two, while the baseline Table 4(1) model is always significant in the time frame considered here.

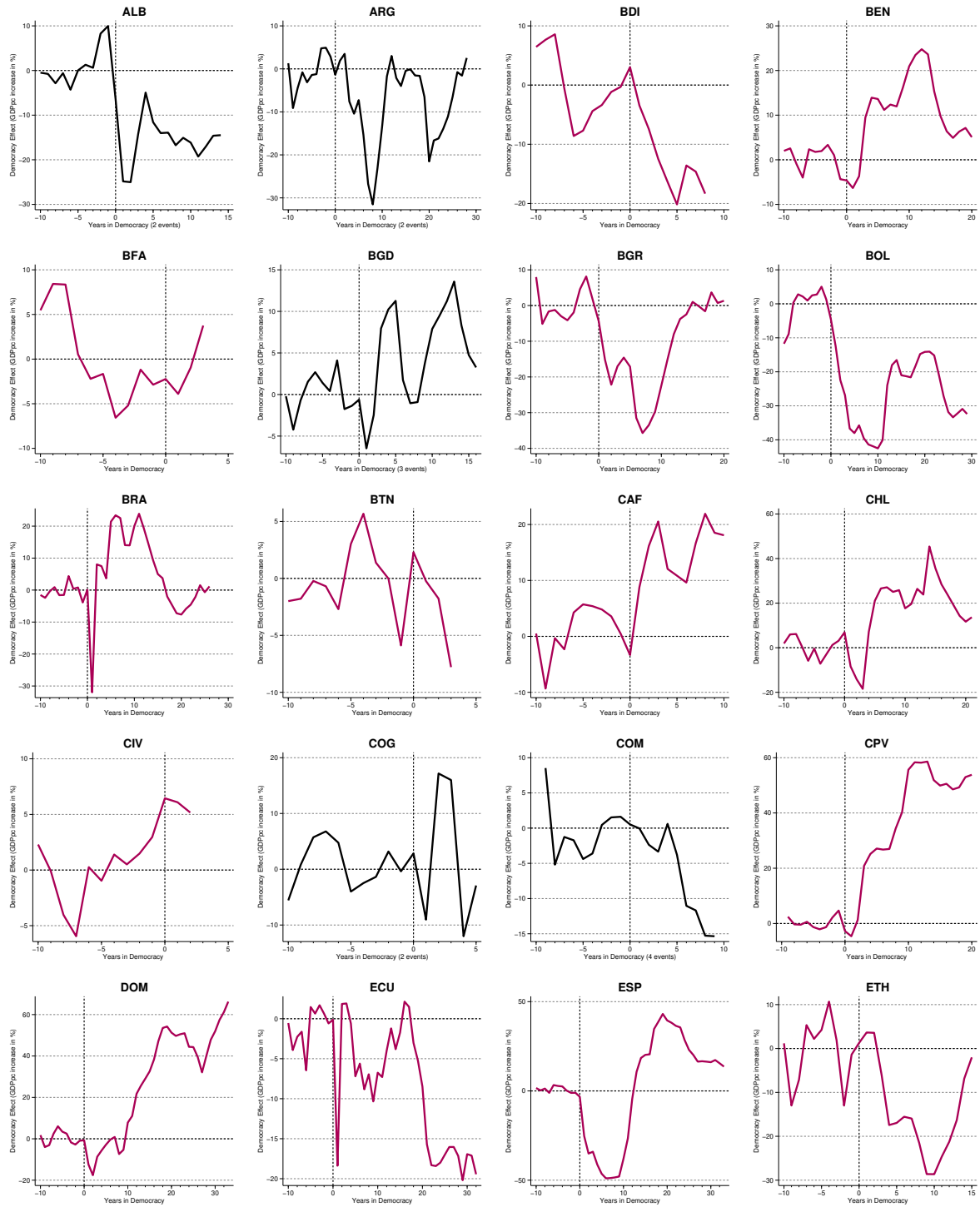
Table E-4: Sample Reduction Estimates — Madsen et al (2015)

	Sample reduction by T_i count			Sample reduction by end year		
	[1]	[2]	[3]	[4]	[5]	[6]
Estimator	IV	IV	IV	IV	IV	IV
Specification	Benchmark	Add Lit $_{t-1}$	Dem $_t$	Benchmark	Add Lit $_{t-1}$	Dem $_t$
Reference	Table 4(1)	Table 4(6)	Table 5(9)	Table 4(1)	Table 4(6)	Table 5(9)
Panel A: Full Madsen et al sample estimates						
Long-Run Coefficient	95.758 [25.745]***	77.763 [30.512]**	121.708 [35.549]***	95.758 [25.745]***	77.763 [30.512]**	121.708 [35.549]***
Countries	141	141	141	141	141	141
min T_i /End year	1	1	1	2000	2000	2000
Observations	1,143	869	1,276	1,143	869	1,276
Share of full sample	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Estimate insignificant (10% significance level)						
Long-Run Coefficient	59.417 [40.168]	44.752 [28.707]	33.093 [61.963]	n/a	41.243 [26.932]	47.151 [35.022]
Countries	45	95	62		124	139
min T_i /End year	13	4	7		1990	1990
Observations	749	755	945		725	1,138
Share of full sample	0.66	0.87	0.74		0.83	0.89
Panel C: Estimate outside 90% CI of full sample estimate						
Long-Run Coefficient	n/a	5.328 [8.722]	33.093 [61.963]	n/a	27.189 [24.587]	47.151 [35.022]
Countries		50	62		102	139
min T_i /End year		8	7		1980	1990
Observations		538	945		589	1,138
Share of full sample		0.62	0.74		0.68	0.89

Notes: The table presents estimates for the two sample reduction exercises in columns (1)-(3) and (4)-(6), respectively (estimator as indicated). Statistical significance at the 10%, 5% and 1% level are indicated as *, **, and ***, respectively. All models use the decadal data from 1820-2000. The models in (1) and (4) includes only Dem $_{t-1}$, which is instrumented using linguistic distance-weighted 'foreign' democracy; in (2) and (5) literacy in the previous decade is included as additional covariate; (3) and (6) are like the benchmark in (1) and (4) but use contemporaneous democracy. Min T_i here refers to the minimal number of decadal observations included in the regression.

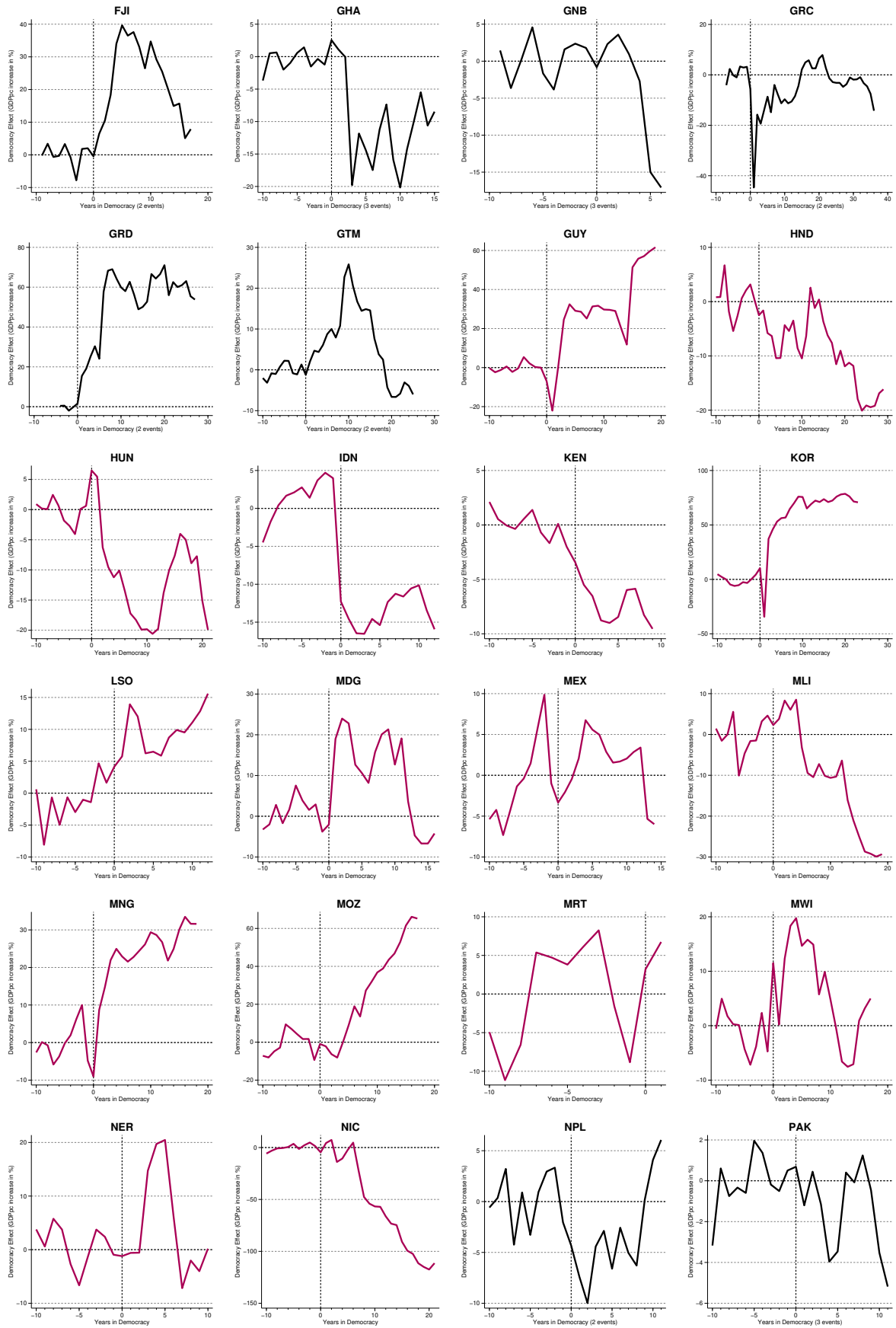
F Generalised Synthetic Control — Country Results

Figure F-1: Country results — Generalised Synthetic Control method



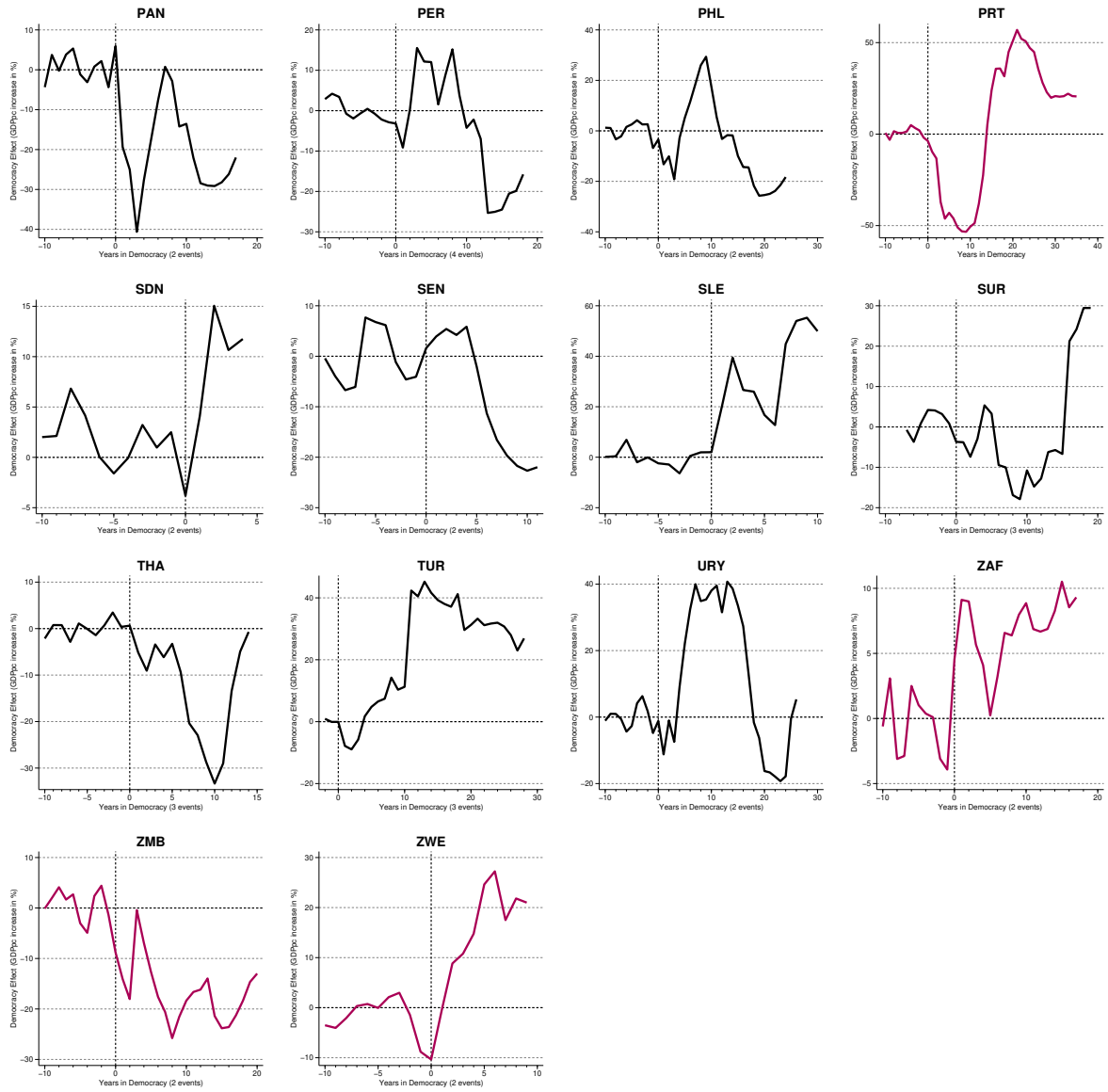
Continued Overleaf

Figure F-1: Country results — Generalised Synthetic Control method (Continued)



Continued Overleaf

Figure F-1: Country results — Generalised Synthetic Control method (Continued)



Notes: These plots show the country-specific results from a generalised synthetic control approach following [Xu \(2017\)](#) with the same covariates as in my [Chan & Kwok \(2021\)](#) approach in the main text and four common factors. The period prior to regime change is cut at a maximum of 10 years. The *x*-axis indicates the years before and after the democratic regime change following the ANRR definition. No additional allowances are made for countries with repeated democratisation events, though country results where just a single democratic regime change occurred are highlighted in dark pink.