DISCUSSION PAPER SERIES

DP16719

Democracy, Growth, Heterogeneity, and Robustness

Markus Eberhardt

MACROECONOMICS AND GROWTH

POLITICAL ECONOMY

ISSN 0265-8003

Democracy, Growth, Heterogeneity, and Robustness

Markus Eberhardt

Discussion Paper DP16719
Published 13 November 2021
Submitted 10 November 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Democracy, Growth, Heterogeneity, and Robustness

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 this assumption to confirm that in the long-run democracy has a positive average effect on per capita income of around 10%, adopting a range of alternative definitions for regime change in the form of binary indicators. Guided by existing hypotheses, additional analysis probes the patterns of the heterogeneous 'democratic dividend' across countries. A second common feature of this literature as well as cross-country growth empirics more generally is the absence of concerns for sample selection or influential observations. I carry out two rule-based robustness exercises to demonstrate that my empirical findings are highly robust to substantial changes to the sample.

JEL Classification: O10, P16

Keywords: democracy, growth, Political development, Difference-in-Difference Estimator, Interactive Fixed Effects

Markus Eberhardt - markus.eberhardt@nottingham.ac.uk
School of Economics, University of Nottingham and CEPR

Acknowledgements
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. The usual disclaimers apply.

Democracy, Growth, Heterogeneity, and Robustness

Markus Eberhardt*

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

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 this assumption to confirm that in the long-run democracy has a positive average effect on per capita income of around 10%, adopting a range of alternative definitions for regime change in the form of binary indicators. Guided by existing hypotheses, additional analysis probes the patterns of the heterogeneous 'democratic dividend' across countries. A second common feature of this literature as well as cross-country growth empirics more generally is the absence of concerns for sample selection or influential observations. I carry out two rule-based robustness exercises to demonstrate that my empirical findings are highly robust to substantial changes to the sample.

JEL Classification: O10, P16

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

*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 96% increase in income per capita for a one standard deviation improvement in democracy. Adopting a new binary measure for democratic regime change 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.[1] These results are novel and important since previous contributions to the literature frequently[2] failed to establish a significant positive relationship (Helliwell 1994, Barro 1996, Minier 1998, Baum & Lake 2003 and Murtin & Wacziarg 2014 for continuous democracy indices and Giavazzi & Tabellini 2005, Rodrik & Wacziarg 2005 for binary measures).[3] For these and other contributions see the schematic review in Appendix Table B-1.[4]

All of the above studies adopt empirical implementations which treat the democracy-growth relationship as *common* across countries.[5] Conceptually, this is a curious choice, given the range of studies in economics and political science which (implicitly or explicitly) develop arguments in favour of heterogeneous economic effects of democratisation (e.g. Cervellati & Sunde 2014, Albertus & Menaldo 2018, Treisman 2020). Other arguments, drawn 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 knowledge diffusion (Comin & Hobijn 2004), further provide motivation for the use of country-specific empirical models. Econometrically, it has been known for some time that heterogeneity misspecification can result in serious bias in static (Sul 2016) or dynamic specifications (Pesaran & Smith 1995). This is most significantly the case in instrumental variable regressions, favoured in the aforementioned recent contributions to this literature, given that any informative instrument is automatically invalid if an underlying heterogeneous relationship is misspecified as homogeneous.[6]

[1] These authors do not necessarily advertise the 30% figure, but a more modest 20% from the 2FE regressions (e.g. James Robinson at 'The Case for Democracy Week' `https://www.youtube.com/watch?v=ARBBiMgJT7A`).

[2] Exceptions 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 great scepticism in the context of cross-country regressions.

[3] A separate strand of the empirical 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).

[4] A broader set of papers is reviewed in Dodsworth & Ramshaw (2021).

[5] Some of the papers cited here have employed interaction terms to highlight the differential growth impact of this or other characteristic studied. None however allow for the democracy-growth relationship to be estimated entirely flexibly (within the confines of the parametric model) as is done in this paper. Pooling country data is not necessarily intended to speak to a homogeneous equilibrium relationship, since results from a pooled estimator might capture an *average* relationship across countries. I engage with these arguments in detail in Section 2.

[6] Briefly, this misspecification introduces $(\beta_i - \beta)x_{it}$ into the error term, which is clearly correlated with any

The above arguments raise the question whether, *conceptually*, all countries benefit from democratic regime change, and if so, to the same extent? And in the same vein, whether, *empirically*, a much more demanding 'heterogeneous' empirical specification to capture this variety *on average* still yields statistically significant and large growth effects? Can these empirical estimates provide some new evidence to speak to existing arguments about how democracies differ and the economic implications of this heterogeneity? These are the questions I seek to answer in this paper.

I study the causal link between democracy and growth when the long-run equilibrium relationship is allowed to differ across countries.[7] I adopt a factor-augmented difference-in-difference implementation which accommodates heterogeneous trends prior to regime change and the endogenous selection of countries into democracy (Chan & Kwok 2018). This is achieved by means of a country-specific estimation equation for 'treated' countries which is augmented with common factor proxies estimated from the set of control countries which never transitioned into democracy. Like time-*invariant* country fixed effects in a pooled model, these time-*varying* 'interactive fixed effects' (Bai 2009) and their country-specific parameters can be correlated with the other regressors, most notably the democracy dummy, which accommodates challenges to identification such as selection into democracy.

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 — or the specific time period studied. While it is tacitly acknowledged that the countries *included in the regression sample* are typically not a representative sample of the population of countries in the world, these concerns are brushed aside when the sample contains a large number of countries or the majority of a group of countries (e.g. OECD countries for the analysis of advanced economies). 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 no 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,

candidate instrument z for an endogenous variable x, unless the instrument is uniformative.

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

dropping countries by the length of their time series or end years of the sample.

There are four main findings: first, I find statistically significant effects of around 10% higher income per capita in the long-run using the democracy indicator introduced by ANRR during the 1960 to 2010 time period. This is around one-half to one-third of the effect suggested by these authors, depending on their implementation. Second, I obtain statistically significant long-run effects of democracy at 12%, 7% and 11% when adopting alternative definitions of democracy (all binary) by Papaioannou & Siourounis (2008, PS), Cheibub et al. (2010, CGV), and Boix et al. (2013, BMR) — it would seem that the specific binary measure for democracy employed is less important for the empirical results than assumptions about parameter homogeneity and the mode of causal inference. Third, my analysis of the patterns of cross-country heterogeneity in the democracy-growth nexus finds no evidence for the relevance of a democratic legacy for a positive democracy-growth nexus; that only rich countries can make democracy 'work' for growth; or that 'elite-biased' democratisations 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 some indications that the democracy effect is a one-off levels effect (in line with ANRR) rather than a permanent growth effect (in line with PS). Somewhat surprisingly, democratisation 'by mistake' (Treisman 2020) appears to yield *higher* long-run growth. Fourth, when I conduct a set of stringent, rule-based sample reduction exercises these highlight the robustness of my empirical findings to substantial sample variations.[8]

The remainder of this paper is organised as follows: in the next section I I draw on the existing economics and political science literatures to motivate why democracy may not have uniform economic effects across countries. Section 3 briefly discusses the data sources. In Section 4 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 5 focus on whether my findings hold when I reduce the cross-section or time-series dimension of the panel. A conclusion follows.

2 Motivation

In this section I review recent literature on the potentially heterogeneous growth effects of democracy. For illustration I draw on various data presented in Figure 1 — unless indicated

[8] I compare and contrast this outcome with the robustness of (a) the seminal ANRR study, where the 2SLS estimates as well as their 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; and (b) select specifications of Madsen et al. (2015), where 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.

the definition for democracy and the sample period (1960-2010) is that of ANRR.

A first set of arguments broadly suggests that not all democracies have the same 'quality' or 'soundness' of political institutions. This is, of course, inevitable with respect to the crude, binary democracy indicators used in this empirical literature, but the following aims to provide some conceptual arguments for this heterogeneity. Recent work on elite-biased democracy (Albertus & Menaldo 2018, Boucekkine et al. 2020) suggests 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). The initial heterogeneity at the point of democratisation aside, if shaking off such historical shackles takes different lenghts of time in different countries, then the 'democratic dividend' will materialise much slower in some than in other new democracies, since elite-biased regimes continue to exploit the many for the few. Panel (a) in Figure 1 contrasts the subsequent economic growth performance of countries which experienced elite-biased democratisation with those which experienced 'popular' democratisation:[9] while the former on average have lower real GDP per capita growth, the estimates are very imprecise and hence not statistically significantly different between the two groups.[10]

A meticulous case study approach by Treisman (2020) posits that over two-thirds of democratisations since 1800 did not occur because the incumbents chose this course of action, but because they slipped up when trying to prevent it. If autocracies frequently turn into democracies due to miscalculation, then structural characteristics (possibly necessary conditions for democratisation, such as a certain level of economic development or human capital) can no longer play the primary role in *determining regime change*, yet they continue to dominate *economic performance after regime change* (e.g. poorer economies and those with higher human capital should grow faster). Democratising 'by mistake' also aligns with other work pointing to the growth-enhancing or -suppressing characteristics of different *modes* of democratic transition (e.g. a peaceful or violent revolution, as in Cervellati & Sunde 2014). Panel (b) in Figure 1 shows[11] that on average subsequent growth performance in countries which democratised 'by mistake' has statistically significantly *outperformed* that in other new democracies.[12]

[9]This analysis adopts standard OLS, median and robust regression models of the annual real GDP per capita growth rate for new democracies over the period they were democracies, regressed on a constant and a dummy for elite-bias (i.e. a new democracy's constitution was drawn up by the incumbent autocratic regime), taken from Table 3.2 of Albertus & Menaldo (2018). If the constitution is subsequently annulled I do not group the democratisation as elite-biased. Countries (can) have several episodes of democracy. The definition of democracy (and reversal) is that of ANRR, but adjusted in timing to Albertus & Menaldo's (2018) choice of CGV.

[10]A comparison of means test (t-test) cannot reject the null of zero difference or one group mean being larger than the other, regardless of whether equal or unequal variances are assumed.

[11]I construct a dummy for Treisman's value of 5 for 'democratisation by mistake' as reported in the data appendix of his paper. Results using a dummy for either value 4 or 5 are qualitatively identical *although* in this case there are merely 19% of sample years following 'normal' democratisations — in the case presented here there are almost 40%.

[12]A comparison of means test borderline-rejects the null of no difference in favour of non-zero difference (two-

Other arguments for 'heterogeneous democracies' include the recent work on populist leaders and their negative implications for economic performance by Funke et al. (2020).[13]

Within advanced economies, democracy matters for growth via a Schumpeterian argument whereby democracy "facilitates creative destruction and thereby encourages innovation" (Aghion et al. 2014, 546; see also Acemoglu & Robinson 2006, Aghion et al. 2007). Given that most of the countries undergoing democratic regime change in the post-WWII period were poor,[14] I draw on broad arguments about the fundamental drivers and patterns of 'modern' economic development more generally to provide additional support for heterogeneous economic performance: neoclassical (Solow 1956) and second generation endogenous growth theory (Howitt 2000) posit/accommodate the notion of cross-country income convergence. Overall, empirical evidence for income convergence is weak (Johnson & Papageorgiou 2020), and panel (c) of Figure 1 investigates this issue within democratising economies: I plot the average per capita GDP growth 'during democracy' against the log GDP pc level in the year of democratic regime change. Regardless of whether 'failed' democratisation episodes are excluded or not, this evidence at best points to non-convergence — both fitted regression lines are positive but with statistically insignificant slopes. One explanation for this could be the different ways in which countries benefit from existing knowledge: with modern technologies diffusing from advanced economies to 'laggards,' the differential intensity of use of adopted technologies is suggested to explain much of the lack of income convergence in the twentieth century (Comin & Hobijn 2004, Comin & Mestieri 2018). De Visscher et al. (2020) demonstrate that countries' heterogeneous and time-varying 'absorptive capacity' (the ability to adapt and adopt existing knowledge, encompassing factors such financial development, human capital, competition policy, and knowledge stock) can capture the observed TFP evolution over the post-WWII period in a sample of advanced economies. As 'miracle' economies like South Korea or Taiwan exemplify, the ability to substantially *increase* absorptive capacity can lead to impressive catch-up growth, and the same principles and potential outcomes can apply to countries further down the income scale. If democratisation does not *uniformly* (re)move the barriers to technology diffusion, knowledge creation and absorption, then this would clearly lead to heterogeneous growth under democracy.[15]

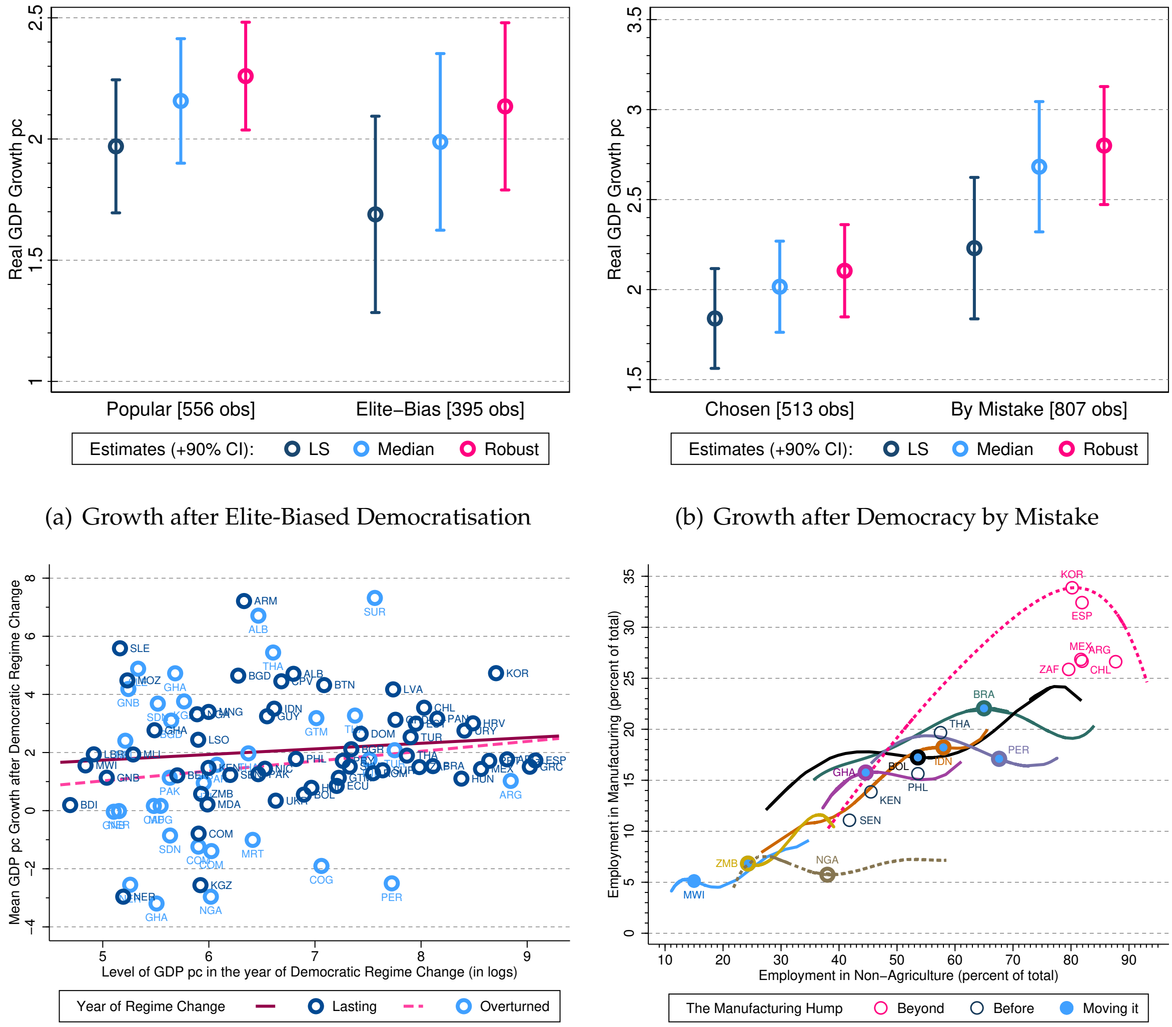
sided, $p = .10$) and rejects in favour of a larger mean for the 'mistake' sample (one-sided, $p = .05$).

[13]Within my 1960-2010 sample of democratising economies Funke et al. (2020) list leaders in the following democracies as populist: ARG (Menem, Kirchner, and Fernandez), BOL (Morales), BRA (Collor), BGR (Borisov), ECU (Bucaram, Correa), GRC (Papandreou), PER (Garcia, Fujimori), PHL (Estrada), POL (Kaczynski brothers), THA (Shinawatra), TUR (Erdogan), TWN (Chen), and ZAF (Zuma).

[14]Of the 103 democratisation events studied in ANRR, 60% are in countries where per capita GDP was *less than half* the sample median for that year (note that ANRR's sample covers over 6,000 observations for 175 countries, including for countries which were democracies throughout the 1960-2010 period analysed).

[15]The cross-country literature on the link between democracy and innovation is surprisingly thin, possibly due to the difficulties in collating meaningful cross-country patent statistics as well as the limited use of patenting (and the absence of obvious alternative proxies for innovation) in less-developed countries.

Figure 1: Democracy and Heterogeneous Development — Descriptive Analysis



(a) Growth after Elite-Biased Democratisation

(b) Growth after Democracy by Mistake

(c) (No) Income Convergence after Regime Change

(d) Economic Structure and Democratisation

Notes: The figure presents arguments for heterogeneous growth performance following democratic regime change. Panel (a) compares the average growth performance after popular vs elite-biased democratisation (Albertus & Menaldo 2018) using least squares, median and robust regression. Panel (b) does the same for Treisman's (2020) democracy by mistake. Panel (c) plots the mean growth rate after democratic regime change (following ANRR) against income per capita in the year of regime change (in logs); dark blue hoops are for the 68 countries which 'permanently' moved to democracy, light blue hoops for the 46 additional democratisation events which were subsequently overturned (reversal into autocracy). Panel (d) uses data from Herrendorf et al. (2014)/GGDC (see Data Appendix for more details) to chart the smoothed employment share in manufacturing against that in 'non-agriculture' (following Huneeus & Rogerson 2020) for six countries along with a marker for the snapshot of the year of regime change in these six and a further eleven countries. Countries in the NE corner of the plot democratised near (like KOR) or 'beyond' the peak of the manufacturing hump, the four 'before the hump' countries are all following upward-sloping trajectories.

A second explanation for the lack of *aggregate* income convergence relates to structural transformation, the movement of 'surplus' labour in agriculture to the manufacturing sector, which has been argued to act as an 'elevator' for economic growth, even leading to unconditional convergence in *sectoral* labour productivity (Rodrik 2013). Given the integral function of restrictions on labour movement between sectors in theoretical models of autocracy (e.g. Acemoglu & Robinson 2006, Acemoglu et al. 2015), it should be expected that democratisation can provide a renewed impetus to structural change and industrialisation when such frictions are removed. Panel (d) of Figure 1 uses long time series data on sectoral employment share (data cover 1950s-2011, from Herrendorf et al. 2014 and Vries et al. 2014) and for a subset of the eight countries presented in this graph a smoothing procedure adopted from Huneeus & Rogerson (2020)[16] to indicate the relationship between structural change on the one hand (share of manufacturing employment on the y-axis and the share of employment outside agriculture on the x-axis) and the timing of democratic regime change on the other.[17] The existing literature suggests successful industrialisation follows a hump-shaped pattern for manufacturing employment, as exemplified by the South Korean profile plotted in pink dashes. Transformation of an agrian-based economy into a modern industrialised economy, at least as far as this growth paradigm is concerned, depends on a relatively high share of manufacturing employment and low share of agricultural employment at the peak of this process: all the countries in the North-Eastern corner of this plot ('beyond the hump') experienced democratic regime change near or beyond the peak of this hump — hence, their growth dividend from industrialisation is largely exhausted. An additional four countries ('before the hump') are highlighted as undergoing democratic regime change at a time when the manufacturing share of employment is rising steadily and none of these countries have reached their peak yet.

An important paper by Rodrik (2016) established that, more recently, many low-income countries appear to have *de*-industrialised prematurely, i.e. the peaks of their manufacturing humps are substantially further to the west and south of the shining Korean example in the graph in panel (d). In this context, the seven economies in a third category for which the smoothed evolution paths are provided here (solid lines) are particularly interesting: (i) all of these appear to have profiles which are not *obviously* single-peaked, and (ii) where the timing of democratisation could motivate the hypothesis that democratic regime change *enabled these countries* to 'move the manufacturing hump'. Take Bolovia (BOL), where the 1982 regime

[16]The manufacturing employment share h_{mt} is regressed on a fifth-order polynomial of the 'non-agriculture' employment share $1 - h_{at}$ and the results are used to predict the smoothed $\hat{h}_{mt}$ used in the country-plots. This is carried out separately for each country. It is important to emphasise that this procedure only produces sensible results when long time series data are available, like the Vries et al. (2014) data. I do not present other evolution paths for 'beyond the hump' and 'before the hump' countries since they follow the patterns described below and would distract from the interesting pattern of the seven countries highlighted.

[17]For countries like Ghana or Peru with several democratisation events I adopt the year of a *permanent* shift to democracy in this illustration.

change occurred some time after the manufacturing employment share seemed to have peaked at around 18%. Yet, *after democratisation*, further structural transformation led to a second peak at 24%, substantially higher than the pre-democracy equivalent. The other countries selected for illustration here (solid lines) similarly follow S-shaped profiles, with democratic regime change taking place some time before an (eventual) uptick in industrialisation. This renewed shift towards manufacturing is weaker in some (Brazil) than in other countries (Malawi) and of course these patterns are merely descriptive findings. This aside, many of the democratising countries for which long time series for sectoral employment are *not* available may have hardly industrialised (moved north) at all, or may follow profiles like that exemplified by Nigeria (olive dashed line), where the years after regime change in 1999 saw virtually no change in manufacturing employment share.[18] Nevertheless, to the best of my knowledge, the S-shaped patterns along with the relative timing of democratic regime change have not been highlighted in the literature before. Again, if democracy does not uniformly relax existing 'binding constraints' in an economy (here: the movement of labour across sectors as formally modelled in Acemoglu et al. 2015),[19] then this may explain heterogeneous growth outcomes within a democratising sample of (developing) countries.

In addition to these economic arguments in favour of heterogeneous growth effects of democratisation, the literature investigating sufficient conditions for democratisation has studied elevated income levels ('you have to be rich') or human capital thresholds ('you have to be educated') as possible necessary conditions for democracy to be a lasting (economic) success (e.g. Przeworski et al. 2000, Madsen et al. 2015, ANRR). Treisman (2020) notes that most scholars would agree that *by themselves* these qualities are not sufficient to explain democratisation.

While there is hence a wealth of arguments why the economic implications of democracy may differ across countries, the empirical practice to estimate pooled models is largely unchallenged. Econometrically, the distinction between a common estimate derived from a pooled model, say $\hat{\beta}$, and an average ('Mean Group') estimate from a heterogeneous parameter model, say $\hat{\beta}^{MG} = \sum_i \hat{\beta}_i$, may seem innocuous, given that 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*.[20] Yet the distinction matters greatly for identification: as has been known for a long time (Pesaran & Smith 1995), misspecifi-

[18] The *most recent* years for Nigeria are clustered around the year of 'permanent' democratisation. In all other country plots the evolution of structural change largely follows a 'west to east' chronology (early to recent years).

[19] This might be the case if land reform, a known contributing factor to sectoral transformation (Galor et al. 2009, Bhattacharya et al. 2019), has been botched or kicked into the long grass by the previous regime (Albertus 2015).

[20] It is indeed 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).

cation of this form implies that the basic assumptions of 2SLS estimators are violated. If the true coefficient on the variable of interest x_{it} is β_i, yet the implementation instead assumes/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, $E[z_{it}\epsilon_{it}] = 0$. While the ultimate driving force or forces for a heterogeneous democracy-growth nexus may still be the subject of lively debate, the econometric argument highlights the serious implications for any claims of 'causal inference' when heterogeneity is ignored.

3 Data

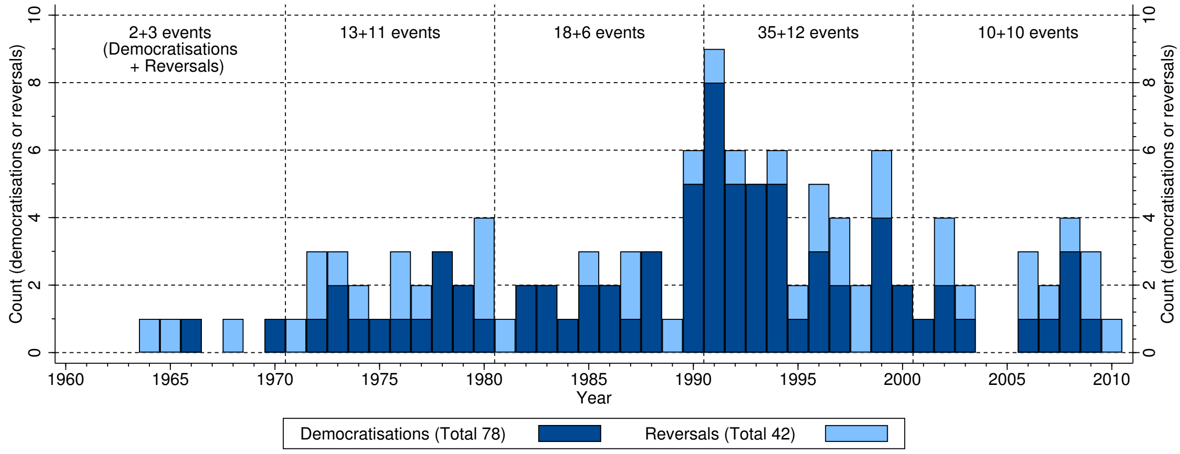
Given the prominence of their work as well as the robustness of the empirical results across different (pooled) implementations, I adopt the dataset and sample from ANRR. Most importantly, this includes a new democracy indicator which combines information from two separate sources ("to purge spurious changes", ANRR: 48) 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).[21] 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 limited to permanent democratisations). A graphical representation (Figure A-1 in the Appendix) 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.

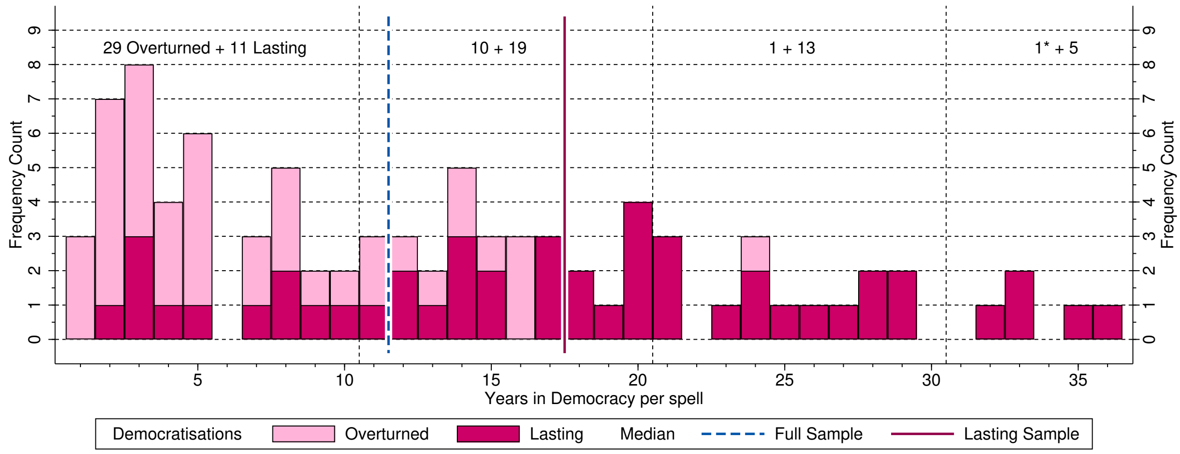
Figure 2 Panel (a) plots the distribution of democratisation events as well as reversals to autocracy (following the ANRR definition), restricted to my dynamic regression sample of 61 countries (treatment sample). The average country had 1.3 democratisation events and 0.7 reversals; 16 countries experienced 2 or more democratisations (THA had 4), 8 countries had two or more reversals — in additional analysis below, Section 4.3, I will separately report the

[21] For more details see Appendix A1 of ANRR.

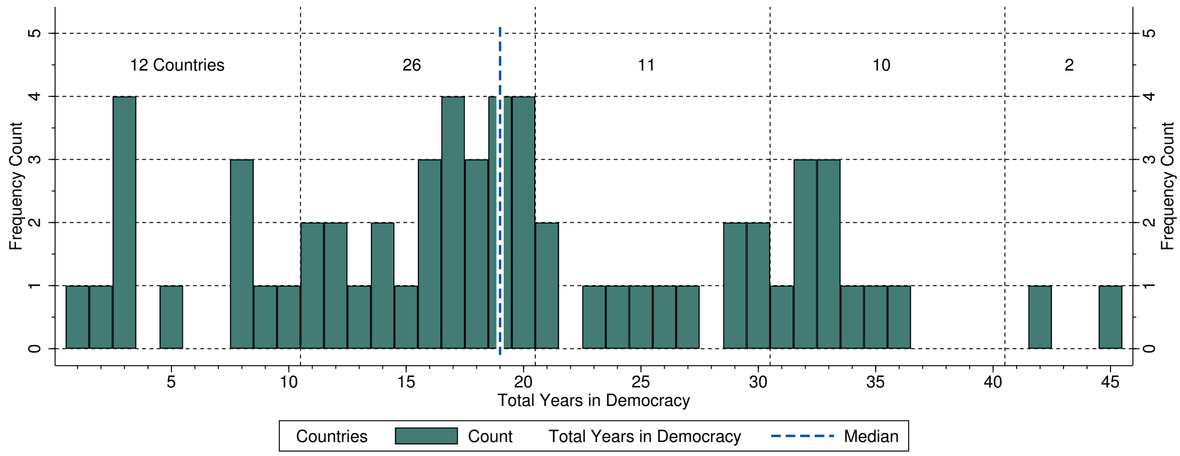
Figure 2: Democratisation, Reversal, and Time in Democracy — 1960-2010



(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.

average growth 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 episodes/spells and distinguishes lasting democratisation from democratic episodes which were subsequently overturned/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'.[22] 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, 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 for 45 countries. My control sample is for all those countries which never transitioned into democracy, for which there are 1,194 observations in 38 economies (accounting for data availability) — 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.

Once data availability for the controls is accounted for the remainder is a potential sample of 3,017 observations in 84 countries. However, estimating dynamic heterogeneous panel models is demanding and hence I have to impose a minimum time-series observation count of 24 — given that the focus of the analysis is on the long-run, I do not feel that this introduces undue distortions.[23] 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 Ta-

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

[23]Compared with the ANRR dynamic IV regressions this omits 16 countries with 19 democratisation events and 6 reversals (given additional data availability for gross investment and trade): *ARM (18/20 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 results.

ble A-2 for details on each country in the 'treated' sample.[24] The samples for the alternative definitions of democracy are smaller and capture fewer events.

4 Heterogeneity

4.1 Methodology: Heterogeneous treatment effects

I this paper I study the impact of observable and unobservable heterogeneity on empirical estimates of the democracy-growth nexus. My model builds on the panel time series econometric literature which has emphasised heterogeneous parameters across panel members (Pesaran & 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. Strong correlation across panel members 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). This literature has taken to specifying a multi-factor error structure, $\boldsymbol{\lambda}_i'\boldsymbol{F}_t$, where $\boldsymbol{F}$ is a set of common factors with associated heterogeneous factor loadings $\boldsymbol{\lambda}$, to capture this strong dependence.[25]

The most recent contributions to this econometric literature have been able to build bridges to the literature on policy evaluation using synthetic control methodology (Xu 2017) or the difference-in-difference specifications (Gobillon & Magnac 2016, Chan & Kwok 2018) most suited to the present empirical setup. What distinguishes these latest approaches from their canonical predecessors is the adoption of a multi-factor error structure in order to address three challenges to identification in these popular methods: (i) the presence of uncommon trends prior to the policy change evaluated, (ii) endogeneous selection into 'treatment', and (iii) the possibility that, following the policy change, 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).

Previous work analysing the democracy-growth nexus using difference-in-difference specifications includes Giavazzi & Tabellini (2005), Papaioannou & Siourounis (2008) and Cervellati & Sunde (2014). My implementation follows the spirit of Chan & Kwok's (2018) estimator but adopts cross-section averages à la Pesaran (2006) instead of estimated factors à la Bai (2009) due to the strongly unbalanced nature of the panel data at hand — more details below. Crucially, this setup allows for correlation between the unobserved determinants of growth (culture, ab-

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

[25]In other work I have provided detailed discussions of how to motivate and implement observed and unobserved heterogeneity in the context of the cross-country production function which underlies the empirical growth literature (e.g. Eberhardt & Presbitero 2015, Eberhardt & Teal 2020) — since this paper is focused on estimating 'treatment effects' I refer interested readers to these articles.

sorptive capacity, etc.) and selection into democratic transition. Since it may be suggested that 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 (2018) estimator, relegating the static results to an appendix.

In the potential outcomes framework, democratic regime change (the 'treatment') for country i at time T_0 can be written as

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

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

The two indicator variables $\mathbf{1}_{\{\cdot\}}$ refer to the country and the year treated, Θ_{it} is the time-varying heterogeneous treatment effect, $\boldsymbol{X}$ is a vector of observed covariates with associated country-specific parameters $\boldsymbol{\beta}_i$, $\boldsymbol{\lambda}_i'\boldsymbol{F}_t$ represents a set of unobserved common factors $\boldsymbol{F}_t$ with country-specific factor loadings $\boldsymbol{\lambda}_i$,[26] and $\widetilde{\epsilon}_{it}$ is the error term.

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

$$y_{it} = \overline{\Theta}_i\mathbf{1}_{\{i\in E\}}\mathbf{1}_{\{t>T_0\}} + \varsigma_i + \boldsymbol{\beta}_i'\boldsymbol{X}_{it} + \boldsymbol{\lambda}_i'\boldsymbol{F}_t + \epsilon_{it}. \tag{3}$$

The composite error term, $\epsilon_{it} = \widetilde{\epsilon}_{it} + \widetilde{\Theta}_{it}\mathbf{1}_{\{i\in E\}}\mathbf{1}_{\{t>T_0\}}$, has zero mean due to the decomposition assumption made above, but it can be heteroskedastic (perhaps due to spatial correlation) and/or serially correlated.

The basic intuition for the Chan & Kwok (2018) implementation is the same as that in the microeconometric literature on production function estimation in the 'control function estimators' of Olley & Pakes (1996) and Levinsohn & Petrin (2003): 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 $\boldsymbol{X}$. This implies that *democratisation can be endogenous to observed variables* (gross investment/GDP and trade openness) *and unobserved common factors*: time-varying latent driving forces of economic development such as culture or

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

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. (2018), among others; and (ii) that the composite error term ϵ is orthogonal to $\boldsymbol{X}$, $\boldsymbol{F}$, parameters and factor loadings, as well as the treatment dummies: any selection into democracy is 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)[27] I specify the following static regression model

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

where y is per capita GDP (in logs and multiplied by 100), Dem is the democracy dummy, and $\boldsymbol{X}$ is a set of additional controls (gross investment share of GDP and trade openness). $\overline{y}$ and $\overline{\boldsymbol{X}}$ are the cross-section averages of the observed variables *but for those countries which never experienced democracy during the sample period* (the control group).[28] 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 and I prefer it here to an approach estimating the unobserved common factors because of the unbalanced nature of the panel.[29]

The dynamic variant of equation (4) is:

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

where the two terms involving sums in the first line capture the short-run effects, while θ_i^* and $\boldsymbol{\beta}_i^*$ represent the long-run coefficients for the respective effects of democracy and additional controls on income per capita — 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

[27] With the exception of PS it is common in this literature to lump together single and multiple regime switchers, including countries which only reversed to democracy during the sample period — in Section 4.3 I will have a closer look at the implications of this convention for my results.

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

[29] This would require a complex expectation maximisation procedure to allow for the estimation of the principal components when observations are missing.

Table 1: Main Results – Dynamic Specification (long-run estimates)

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

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

'CS-DL' version of the Chan & Kwok (2018) 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.[30] Following suggestions in Chudik et al. (2016) I adopt $p_{\bar{y}} = 0$ and $p = p_{\overline{X}} = int(T^{1/3}) = 3$, where T is the time dimension of the panel. My presentation below will focus on average estimates of $\hat{\theta}^*$ in the dynamic case (which can be interpreted as ATET estimates); in line with the literature I adopt robust regression (Hamilton 1992) to compute outlier-robust means. In the sources of heterogeneity analysis in Section 4.4 I employ the country-specific ITET estimates $\hat{\theta}_i^*$. Inference for the robust 'Mean Group' estimate is based on standard errors computed non-parametrically, following Pesaran & Smith (1995). Observed covariates $\boldsymbol{X}$ are not included in what I refer to as the 'plain vanilla' Chan & Kwok (2018) implementation — the covariate cross-section averages from the control sample, $\bar{y}$ and $\overline{X}$, are however always included.[31] For comparison, I also estimate simple Mean Group models (Pesaran & Smith 1995) which exclude the cross-section averages in equation (5).

4.2 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, i.e. the second line of equation (5), whereas C&K MG is the Chan & Kwok (2018) 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. The dynamic C&K MG implementation requires estimation of 21 parameters plus an intercept, however due to missing observations the minimum requirement for the ANRR definition of democracy is 24 observations, although the first country in the sample had a minimum of 26. In order to make the estimates across different implementations (i.e. across columns) directly comparable I fix the sample at the C&K MG minimum for T_i.

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

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

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

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

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 (2018) 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 (2018) 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.[32] Panel (d) adopts the PS definition which is limited to 41 permanent democratisations and excludes any autocratic reversals, resulting in two more years spent in democracy for the average country and a long-run democracy effect of 12% in the preferred C&K MG model. It is notable that the standard MG model without adjustment for selection into democracy in column (1) arrives at an average democracy effect of 28%.

Hence, the average long-run effect of democratic transition in the preferred implementation ranges from 7 to 12% across these four different specifications. This translates into one-half to one-third of the long-run effects found in ANRR, depending on their implementation. 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.

4.3 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 (2018) 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

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

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'.

4.4 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? In 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).

First, I return to the competing explanations in political science whereby democratisations can be distiguished as 'elite-biased' or 'popular' (Albertus & Menaldo 2018), and to have occurred 'by mistake' rather than intention (Treisman 2020). Panels (a) and (b) of Figure 3 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.[33] Like in the descriptive analysis of the per capita GDP growth data during democracy in Section 2, the estimated long-run coefficients do not provide statistically significant evidence for lower long-run

[33]The same definitions as laid out in Section 2 are employed here.

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	16.624	7.692	7.712	10.074
(ANRR Definition)	(4.630)***	(2.854)***	(3.647)**	(3.651)***
Observations	2443	2443	2443	2443
Countries (N)	61	61	61	61
Democratisations	78	78	78	78
Reversals	42	42	42	42
Avg Years in Dem	19.6	19.6	19.6	19.6
RMSE	18.861	7.942	8.515	4.115
(b) Full Sample excluding four	18.646	8.231	8.746	11.970
Reversal-only Countries	(4.837)***	(2.895)***	(4.048)**	(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	27.393	4.655	9.702	12.721
without Reversal	(6.456)***	(4.544)	(5.972)	(5.758)**
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	19.548	5.192	8.448	12.270
Democratisation	(6.046)***	(3.521)	(5.451)	(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	16.557	13.654	9.088	12.965
Democratisations	(8.079)**	(4.511)***	(5.851)	(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).

growth in 'elite-biased' democracies,[34] although the mean estimates do point in that direction. The analysis of 'democracy by mistake' confirms the findings from the raw data, namely 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 sub-samples (Sub-Saharan Africa, Latin America and the Caribbean, and Other regions) — for the full-country plot I omit the countries with the largest and the smallest base-year GDP, respectively. Panel (c) of Figure 3 shows the resulting plot for all countries in this sample using a dashed blue line and a shaded blue 90% confidence interval: although the regression line has a minimal hump, the wide confidence interval suggest that no matter whether countries were initially rich or poor, on average the long-run democracy coefficient is around 10%. But looking at three distinct geographical regions (Africa in blue, Latin America and the Caribbean (LAC) in red, and 'Other regions' in orange) yields very different patterns for countries with similar base year GDP (log values 6.6 to 8): low and declining for Africa, high and U-shaped for LAC, and high and stable/rising somewhat for the 'Other' regions. Note that the median share of years in democracy for these regions are 33% for Africa and almost exactly twice that for both LAC and 'others', so that at least for the latter two my simple analysis is not distorted by democratic experience. This analysis would not seem to support the notion that a democracy-growth nexus hinges on a certain minimum-level of income. The regional analysis points to substantial heterogeneity but should not be (mis)read as advocating geographic determinism.

Third, I assume that the effect of democracy on growth does *not* differ across countries, but that different length of 'treatment' (years spent in democracy) results in heterogeneous long-run estimates across countries. In panel (d) of Figure 3 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

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

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.[35] 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 in democracy in 1975 — this cut-off maximises the data availability in the Polity IV dataset.[36] 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 3 shows a fitted linear regression line[37] for the relationship between the long-run democracy coefficient and democratic legacy in years (blue line, shaded 90% CI), together with a histogram for the latter variable. There is no clear advantage or disadvantage of democratic legacy for the 'treated' countries which transitioned into or out of democracy during the sample period. Having said that, as the histogram indicates there is a mere sprinkling of countries with legacies in excess of twenty years. 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.[38]

In panel (f) I study the relationship between human capital endowment and the long-run democracy coefficient: I adopt the literacy data from Madsen et al. (2015) for 1970, which provides largest coverage for my treatment sample while still representing a reasonably early

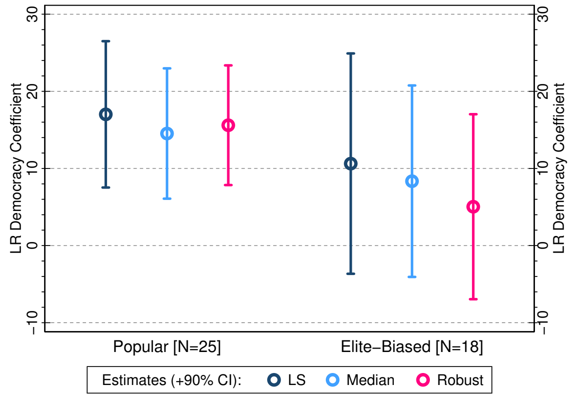
[35] This is constructed from the robust mean estimates in each band following Pesaran & Smith (1995) as in the main results in Section 4.2 above.

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

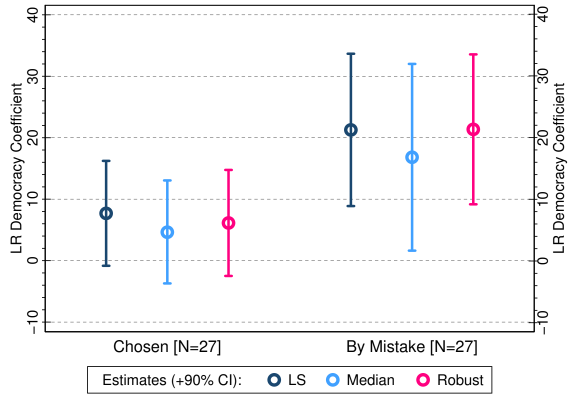
[37] A fractional polynomial points to this linear relationship but has a wider confidence interval in the right (but not the left) tail. I adopt the linear fit for ease of presentation.

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

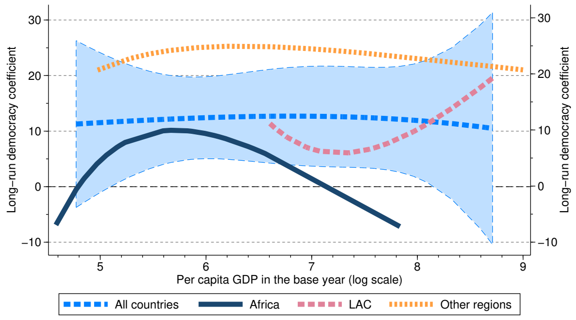
Figure 3: Analysing Heterogeneity (Long-run Coefficients)



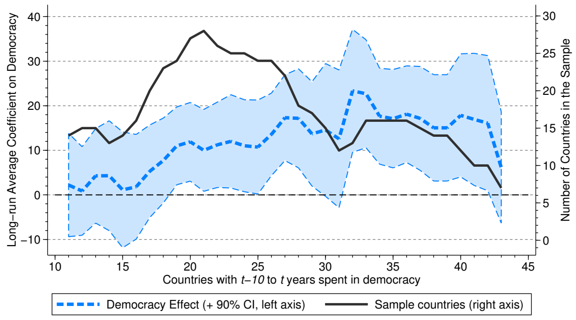
(a) Elite-Biased Democratisation



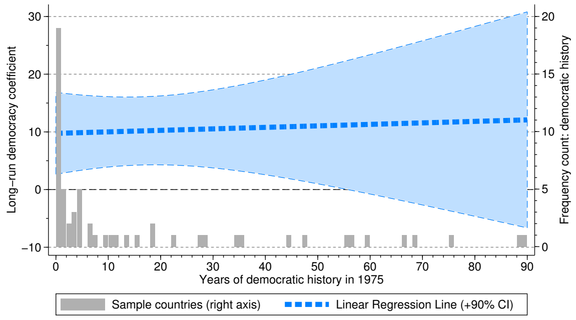
(b) Democracy by Mistake



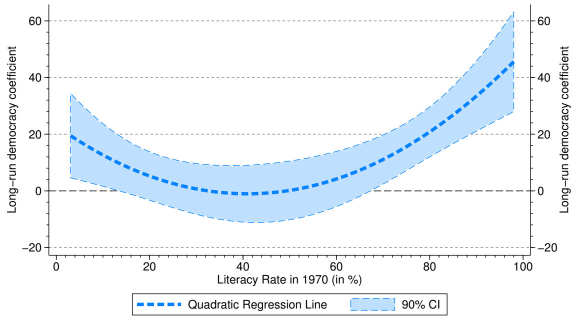
(c) Income Levels and Regions



(d) Sample Years in Democracy



(e) Democractic Legacy



(f) Human Capital

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.

'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.

In my motivation I provided some interesting patterns in selected raw data which pointed to a potential boost to industrialisation from democratisation in countries which were still primarily agrarian. As was pointed out, this was unlikely to be the universal picture, and I employed this descriptive evidence *primarily* to argue for heterogeneous growth effects under democracy. The hypothesis that democracy aids structural change is less straightforward to confirm empirically than the other arguments I analysed above. Productivity-led endogenous structural change would focus on TFP growth in agriculture as a means to release labour from the sector (Huneeus & Rogerson 2020). At the same time, in addition to barriers to labour mobility across sectors, the subsistence/smallholder agriculture-bias of autocratic regimes is associated with underinvestment in agricultural capital stock and other factor inputs (e.g. fertilizer, improved seed varieties). Democratic regime change could bring in more investment and address other productive inefficiencies — although subject to diminishing returns — to boost agricultural output and hence contribute to increased structural change. If manufacturing represents the failsafe 'growth escalator' implied by the unconditional convergence result (Rodrik 2013), then an observed shift of labour from agriculture to manufacturing in response to democratisation could provide the empirical support for theoretical arguments (e.g. Acemoglu et al. 2015). Yet, the recent data from developing countries have shown a shift *directly into services* rather than manufacturing (Rodrik 2016), in the process leading to widespread 'urbanisation without industrialisation' (Gollin et al. 2016). Trade and globalisation are intimately linked to these transformations, with the nagging question lurking in the background whether the manufacturing-led economic development paradigm of Korea and Taiwan was even *feasible* for many economies during the last two decades following the rise and dominance of China in labour-intensive manufacturing for export. These thoughts require careful consideration and are deserving of detailed treatment, which I leave for future research.

5 Robustness

In this section I investigate how robust my main findings are to changes in the sample makeup. 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*,[39] 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.[40] 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 over/-estimated (if more elaborate dynamics translate into less precise estimates given the limited time series data). Studying the evolution of my estimates as the sample is restricted to countries with larger and larger minimum observation counts should go some way to address these concerns.

I then shift my attention to restricting the sample by moving the end year of analysis: my data, taken from ANRR, covers 1960 to 2010, and hence includes the Global Financial Crisis (GFC) — the most significant global macroeconomic shock since the 1930s — as part of the final sample years. Although the impact of the GFC was substantial, it was by no means uniform across countries. The same could be said for the post-crisis recovery. Recent work by Young (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.[41] 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. Coversely, 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.

The focus of these exercises is on my results for the heterogeneous panel models analysed above. In order to provide a benchmark for comparison I carry out the same exercises for the seminal ANRR and Madsen et al. (2015) contributions and briefly discuss the finding.

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

[40]This analysis is based on the spell data presented in Figure 2 Panel (b).

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

5.1 Sample reduction by minimum observation count

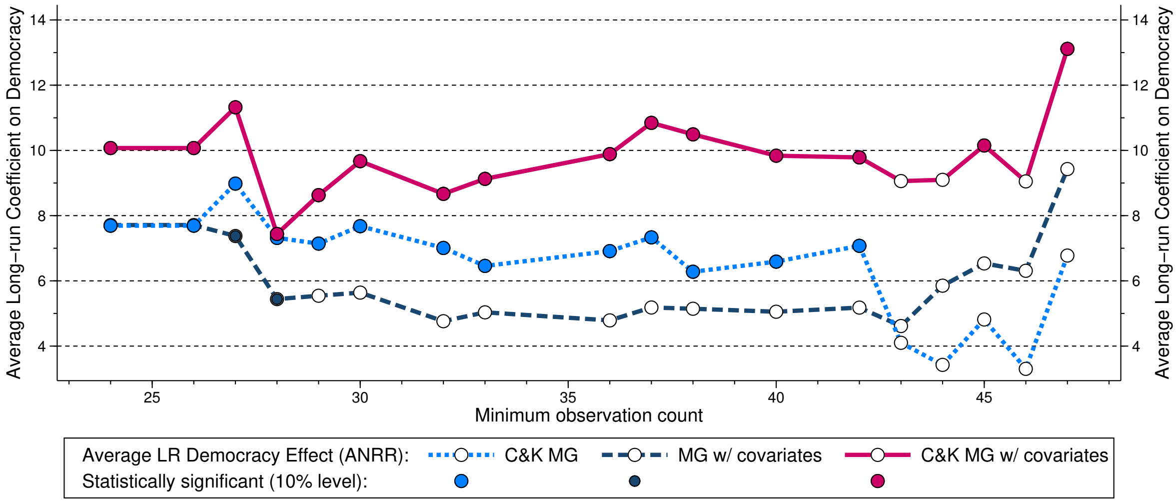
Panel (a) of Figure 4 presents the results from dynamic specifications of three heterogeneous parameter models for the first sample reduction exercise by country observation count, adopting the ANRR definition of democracy — the plots for the BMR, CGV and PS alternatives are provided in an appendix. 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 3 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.

The estimates from the empirical model ignoring any potential factor structure and thus selection, uncommon trends and/or common shocks with heterogenous 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 (2018) model accounting for these distortions (in short blue dashes) yields more stable long-run estimates at around 6-8% until the coefficient turns insignificant when $T_i = 43$. The model including additional covariates (gross investment ratio and trade openness) lifts the estimate somewhat to around 10% throughout the sample reduction exercise. Statistical insignificance first occurs again when $T_i = 43$, where almost 40% of sample observations have been dropped, though the coefficient magnitude is still stable.

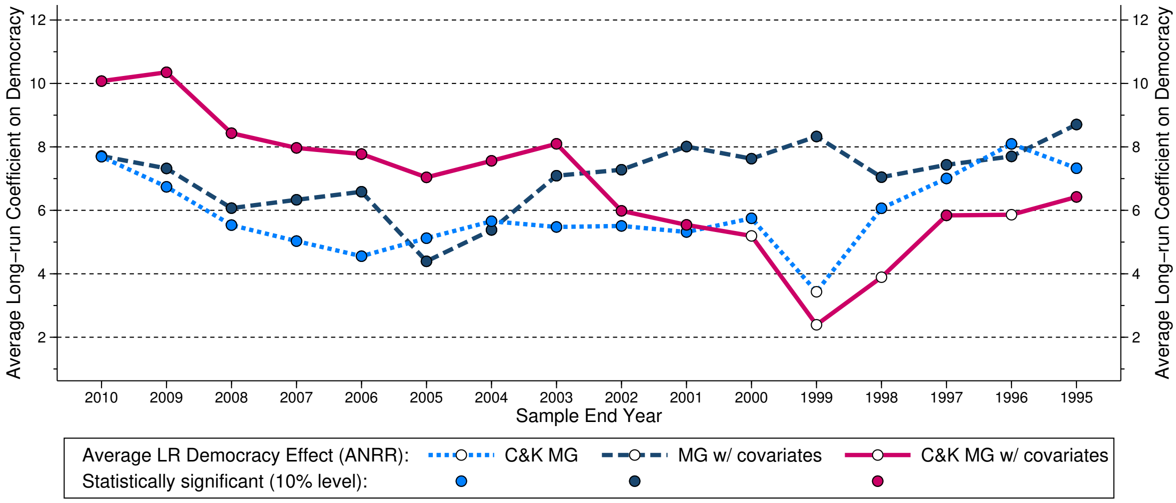
Columns (2)-(4) in Panel (B) of Table 3 report a similar level of robustness for the results when adopting the alternative definitions of BMR and PS for democracy: for BMR the democracy coefficient turns insignificant when 38% of observations are dropped, for PS the figure is 26%. The CGV results turn insignificant when fewer than 1% of observations are omitted — the Appendix figure charting the step-by-step results for this definition however 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 analysis.

In Appendix E I report the findings when I carry out the same sample reduction exercises for a number of implementations in ANRR and Madsen et al. (2015), including the preferred IV specifications. The long-run democracy estimate in ANRR turns statistically insignificant when fewer than 7% of observations are omitted (the coefficient merely drops from 31.5% to 29%) — an accompanying histogram indicates that this does not affect the number of democratisation

Figure 4: Sample Reduction Exercises — Full results



(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.

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.[42] 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.

5.2 Sample reduction by sample end year

Panel (b) of Figure 4 presents the results when observations are omitted by sample end year. In this graph for the ANRR democracy dummy and in the equivalent graphs in the Appendix for BMR, CGV and PS definitions the x-axis is in *reverse* chronological order. Here, the standard MG estimate of the long-run democracy effect (in dark blue dashes) is remarkably stable and remains statistically significant throughout the years displayed. With one exception in 1999 the same is true for the 'plain vanilla' Chan & Kwok (2018) 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 (2018) 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 insights to the findings for the same exercise in the ANRR and Madsen et al. (2015) models presented in Appendix E highlights the 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

[42]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. Note that the ANRR persistence parameter stays within the 90% confidence interval of that full sample estimate, even when almost 42% of observations have been dropped in the balanced panel sample, while at the same time the coefficient on the democracy dummy (from which the long-run coefficient is computed) has bounced about between 1.17 and 0.07.

Table 3: Sample Reduction Exercises

	Sample reduction by T_i count				Sample reduction by end year			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Definition	ANRR	BMR	CGV	PS	ANRR	BMR	CGV	PS
Panel A: Full sample estimate								
Long-Run Effect	10.074	11.118	6.934	11.936	10.074	11.118	6.934	11.936
	(3.651)***	(3.612)***	(4.179)*	(5.072)**	(3.651)***	(3.612)***	(4.179)*	(5.072)**
min T_i/End year	24	23	24	24	2010	2007	2008	2010
Countries	61	55	50	41	61	55	50	41
Observations	2,443	2,051	1,922	1,670	2,443	2,051	1,922	1,670
Share of full sample	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Estimate insignificant (10% significance level)								
Long-Run Effect	9.060	8.589	6.084	9.222	5.194	4.715	5.647	5.760
	(5.551)	(5.485)	(4.014)	(6.147)	(3.220)	(3.135)	(3.520)	(5.124)
min T_i/End year	43	40	25	40	2000	1999	2007	2002
Countries	32	29	49	27	48	42	50	33
Observations	1,488	1,262	1,898	1,239	1,638	1,402	1,875	1,205
Share of full sample	0.61	0.62	0.99	0.74	0.67	0.68	0.98	0.72
Panel C: Balanced Panel Estimate								
Long-Run Effect	13.113	10.488	10.420	9.222	n/a	n/a	n/a	n/a
	(6.088)**	(6.429)	(5.795)*	(7.719)				
min T_i/End year	47	44	45	47				
Countries	24	23	22	19				
Observations	1,128	1,012	990	893				
Share of full sample	0.46	0.49	0.52	0.53				

Notes: The table presents estimates for the two sample reduction exercises in columns (1)-(4) and (5)-(8), respectivly (definition of democracy as indicated). All estimates the robust long-run coefficient for democracy from the Chan & Kwok (2018) 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 4 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.

declining coefficient which is always insignificant.[43] Madsen et al.'s (2015) benchmark specification is again remarkably stable and does not turn statistically insignficant in the periods considered in Figure E-4, 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).

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 (my focus above is on the IV estimates, but AB and HKK estimates are found to be similarly fragile), while the analysis of decadal data over a much longer time horizon in Madsen et al. (2015) holds up quite well.

6 Conclusion

In the introduction to their article, PS draw on a quote from Robert Dahl's (2000) *On Democracy*, which laments that democracy "has meant different things to different people at different times and places" (3). While Dahl was really speaking about the lengthy history of the concept of democracy, PS, and subsequently CGV, BMR, and ANRR, have taken this statement as an invitation to develop their own dichotomous classifications for democracy in the modern, post-WWII era, with the most recent iteration by ANRR designed to "purge spurious changes" (48) in each of its constituent elements — mainly the PolityIV data and the Freedom House Index, but also the other three aforementioned indicators. Since there is so little agreement on how a dichotomous democracy indicator should be defined — see my graphical representation in Figure A-1[44] — why not instead allow for the possibility that democracy may have had different implications for different people at different times and places?

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 results suggest a positive significant long-run growth effect of around 10%, which does not differ substantially across four alternative binary measures for democracy. Further analysis suggests that even countries which experienced several episodes of democratisation

[43] Again, the two-way fixed effects results excepted, the alternative implementations of ANRR do not necessarily 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%.

[44] Curiously, the 'spurious changes' in PS, CGV and BMR (note that 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 between 5 and 9% smaller than ANRR's, still yield next to identical results for their preferred IV specification, with only the CGV implementation yielding a somewhat higher 40% democratic dividend. Perhaps more curiously, the dynamics for all four models yield an *identical* persistence parameter of 0.964, with *identical* standard error of 0.005 (and hence a t-statistic of 193).

and reversal can achieve the same magnitude of effect, provided their years spent in democracy are substantial.

I investigate a number of existing hypotheses in the political science and economics literatures to explain the heterogeneity across countries in their long-run democracy effect. My analysis does not find any evidence of legacy effects for democracy, or that only rich countries can 'make democracy work'. While there is insufficient evidence to support the notion that democracies where regime change came about with the help of the elite, there is substantial evidence that unintended democratisation on behalf of the autocratic leadership seems to lead to better growth outcomes. Human capital appears to nonlinearly interact with the magnitude of the democracy effect, while I also tentatively conclude that the long-run democracy effect is a one-off levels effect rather than a perpetual growth effect — a finding which is arguably underscored by the above-mentioned result that failed attempts do not impinge on the 10% dividend, provided sufficient time in democracy elapsed to achieve this levels effect.

I subject my main results to two sample reduction exercises, which I motivate as challenging the robustness of the findings against the realities of data availability, influential outliers and the macroeconomic shock of a century in form of the Global Financial Crisis. For three of the four democracy indicators adopted the long-run democracy estimates in my preferred empirical model are stable and remain statistically significant until I drop around almost 40% of observations.

References

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.

Acemoglu, D. & Robinson, J. A. (2006), *Economic origins of dictatorship and democracy*, Cambridge University Press.

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.

Aghion, P., Alesina, A. & Trebbi, F. (2007), Democracy, technology, and growth, Working Paper 13180, National Bureau of Economic Research.

Albertus, M. (2015), *Autocracy and redistribution*, Cambridge University Press.

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', *The Review of Economic Studies* **58**(2), 277–297.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G. & Khosravi, K. (2018), Matrix completion methods for causal panel data models, NBER Working Paper 25132, National Bureau of Economic Research.

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

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.

Bhattacharya, P. S., Mitra, D. & Ulubaşoğlu, M. A. (2019), 'The political economy of land reform enactments: New cross-national evidence (1900–2010)', *Journal of Development Economics* **139**, 50–68.

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. (2020), 'A theory of elite-biased democracies', *Aix-Marseille School of Economics Working Paper Series No. 39, halshs-03044565* .

Broderick, T., Giordano, R. & Meager, R. (2020), 'An automatic finite-sample robustness metric: Can dropping a little data change conclusions?', *arXiv preprint arXiv:2011.14999* .

Cervellati, M. & Sunde, U. (2014), 'Civil conflict, democratization, and growth: Violent democratization as critical juncture', *The Scandinavian Journal of Economics* **116**(2), 482–505.

Chan, M. K. & Kwok, S. (2018), Difference-in-differences when trends are uncommon and stochastic, Technical report, SSRN Working Paper 3125890.

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.

Comin, D. & Mestieri, M. (2018), 'If technology has arrived everywhere, why has income diverged?', *American Economic Journal: Macroeconomics* **10**(3), 137–78.

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.

Eberhardt, M. & Presbitero, A. F. (2015), 'Public debt and growth: Heterogeneity and nonlinearity', *Journal of International Economics* **97**(1), 45–58.

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.

Eberhardt, M. & Teal, F. (2020), 'The magnitude of the task ahead: Macro implications of heterogeneous technology', *Review of Income and Wealth* **66**(2), 334–360.

Funke, M., Schularick, M. & Trebesch, C. (2020), 'Populist leaders and the economy', *CEPR Discussion Paper No. 15405* .

Galor, O., Moav, O. & Vollrath, D. (2009), 'Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence', *The Review of Economic Studies* **76**(1), 143–179.

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.

Gobillon, L. & Magnac, T. (2016), 'Regional policy evaluation: Interactive fixed effects and synthetic controls', *Review of Economics and Statistics* **98**(3), 535–551.

Gollin, D., Jedwab, R. & Vollrath, D. (2016), 'Urbanization with and without industrialization', *Journal of Economic Growth* **21**(1), 35–70.

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.

Herrendorf, B., Rogerson, R. & Valentinyi, A. (2014), Growth and structural transformation, *in* P. Aghion & S. N. Durlauf, eds, 'Handbook of Economic Growth', Vol. 2, Elsevier, pp. 855–941.

Howitt, P. (2000), 'Endogenous growth and cross-country income differences', *American Economic Review* **90**(4), 829–846.

Huneeus, F. & Rogerson, R. (2020), Heterogeneous paths of industrialization, Technical Report 13180, National Bureau of Economic Research.

Johnson, P. & Papageorgiou, C. (2020), 'What remains of cross-country convergence?', *Journal of Economic Literature* **58**(1), 129–75.

Knutsen, C. H. (2013), 'Democracy, state capacity, and economic growth', *World Development* **43**, 1–18.

Levinsohn, J. & Petrin, A. (2003), 'Estimating production functions using inputs to control for unobservables', *The Review of Economic Studies* **70**(2), 317–341.

Madsen, J. B., Raschky, P. A. & Skali, A. (2015), 'Does democracy drive income in the world, 1500–2000?', *European Economic Review* **78**, 175–195.

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* .

Minier, J. A. (1998), 'Democracy and growth: Alternative approaches', *Journal of Economic Growth* **3**(3), 241–266.

Murtin, F. & Wacziarg, R. (2014), 'The democratic transition', *Journal of Economic Growth* **19**(2), 141–181.

Olley, G. S. & Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica* **64**(6), 1263–1297.

Papaioannou, E. & Siourounis, G. (2008), 'Democratisation and growth', *The 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. (2013), 'Unconditional convergence in manufacturing', *The Quarterly Journal of Economics* **128**(1), 165–204.

Rodrik, D. (2016), 'Premature deindustrialization', *Journal of Economic Growth* **21**(1), 1–33.

Rodrik, D. & Wacziarg, R. (2005), 'Do democratic transitions produce bad economic outcomes?', *American Economic Review, Papers & Proceedings* **95**(2), 50–55.

Solow, R. M. (1956), 'A contribution to the theory of economic growth', *Quarterly Journal of Economics* **70**(1), 65–94.

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* p. forthcoming.

Vries, G. d., Timmer, M. & Vries, K. d. (2014), Patterns of structural change in developing countries, Technical report, Groningen Growth and Development Centre, University of Groningen.

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.

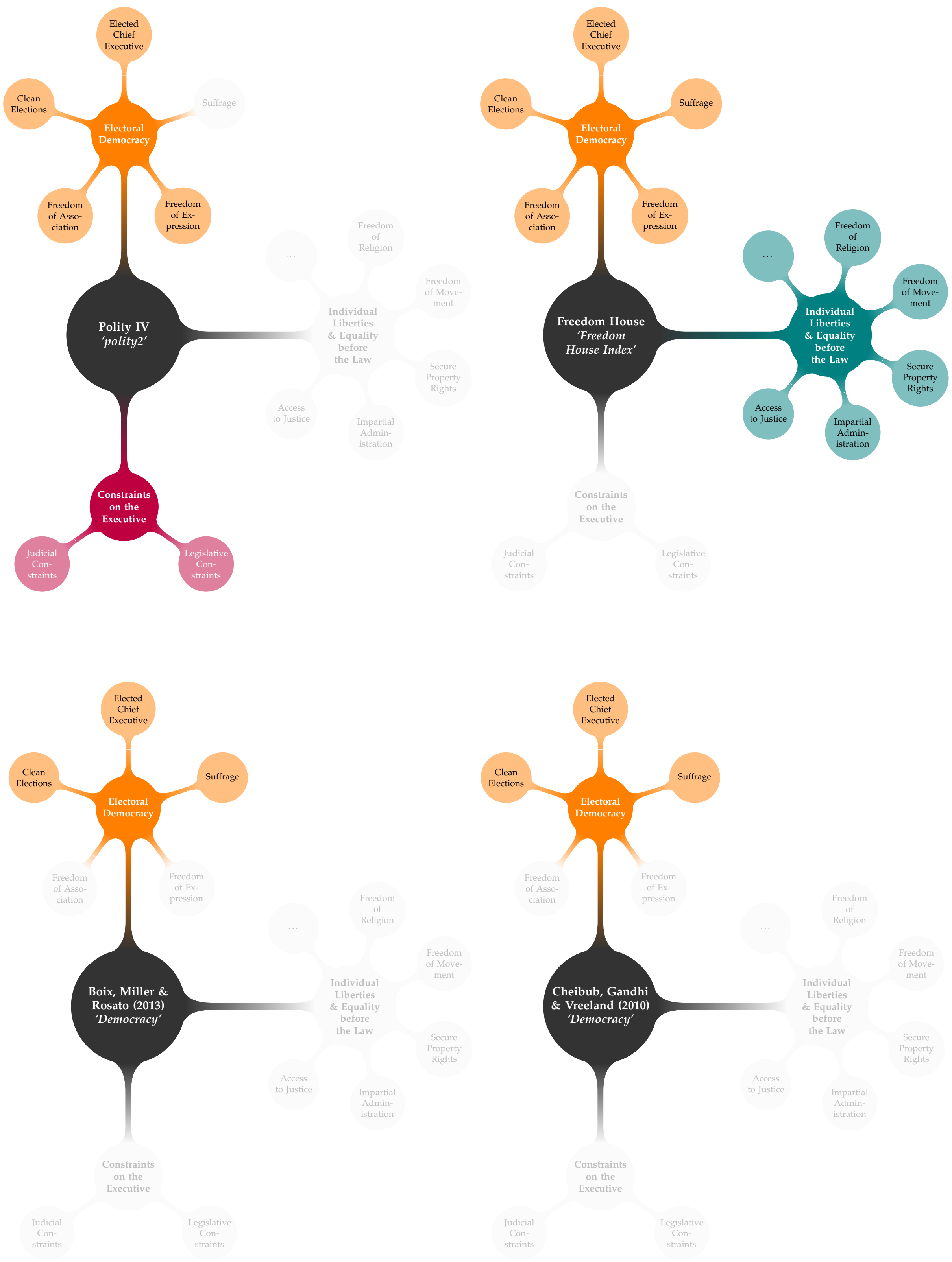
Democracy (ANRR) Data are combined from the Polity IV project, Freedom House and a number of alternative sources. The primary strategy for construction of the democracy dummy prescribes a positive Polity score in addition to a 'free' or 'partially free' label in Freedom House. Further strategies are 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	Indonesi	**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-2: Regression Sample — Dynamic 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
36	Sri Lanka	LKA						**LKA**	14	26	40	1	1	**LKA**	12	29	41	1	1	LKA					
37	Lesotho	**LSO**	25	17	42	2	1	**LSO**	33	6	39	1	0	LSO						**LSO**	24	18	42	1	0
38	Madagascar	**MDG**	30	16	46	1	1	**MDG**	29	15	44	1	0	**MDG**	29	16	45	1	0	**MDG**	29	17	46	1	0
39	Mexico	**MEX**	33	14	47	1	0	**MEX**	36	8	44	1	0	**MEX**	36	9	45	1	0	**MEX**	33	14	47	1	0
40	Mali	**MLI**	21	19	40	1	0	**MLI**	21	16	37	1	0	**MLI**	21	17	38	1	0	**MLI**	21	19	40	1	0
41	Mongolia	**MNG**	8	18	26	1	0	**MNG**	5	18	23	1	0	**MNG**	5	19	24	1	0	**MNG**	8	18	26	1	0
42	Mozambique	**MOZ**	10	17	27	1	0	**MOZ**	14	10	24	1	1	MOZ						**MOZ**	10	17	27	1	0
43	Mauritania	**MRT**	46	1	47	1	1	MRT						MRT						MRT					
44	Malawi	**MWI**	27	13	40	1	0	**MWI**	27	10	37	1	0	**MWI**	27	11	38	1	0	**MWI**	27	13	40	1	0
45	Niger	**NER**	30	12	42	2	1	**NER**	32	10	42	2	1	**NER**	33	9	42	2	1	NER					
46	Nicaragua	**NIC**	26	21	47	1	0	**NIC**	20	24	44	1	0	**NIC**	20	25	45	1	0	**NIC**	26	21	47	1	0
47	Nepal	**NPL**	26	16	42	2	1	**NPL**	28	11	39	1	1	**NPL**	27	13	40	2	1	NPL					
48	Pakistan	**PAK**	21	19	40	3	2	**PAK**	21	16	37	2	2	**PAK**	21	17	38	3	2	PAK					
49	Panama	**PAN**	10	17	27	1	0	**PAN**	7	17	24	1	0	**PAN**	5	20	25	1	0	**PAN**	10	17	27	1	0
50	Peru	**PER**	13	34	47	2	2	**PER**	23	21	44	2	2	**PER**	23	22	45	2	2	**PER**	16	31	47	1	0
51	Philippines	**PHL**	22	25	47	1	1	**PHL**	21	23	44	1	1	**PHL**	21	24	45	1	1	**PHL**	23	24	47	1	0
52	Portugal	**PRT**	3	35	38	1	0	**PRT**	3	32	35	1	0	**PRT**	3	33	36	1	0	**PRT**	3	35	38	1	0
53	Sudan	**SDN**	30	3	33	1	1	**SDN**	27	3	30	1	1	**SDN**	28	3	31	1	1	SDN					
54	Senegal	**SEN**	36	11	47	1	0	**SEN**	36	8	44	1	0	**SEN**	36	9	45	1	0	**SEN**	36	11	47	1	0
55	Sierra Leone	**SLE**	17	11	28	2	1	**SLE**	19	6	25	1	0	**SLE**	14	12	26	2	1	SLE					
56	Suriname	**SUR**	9	19	28	2	2	**SUR**	9	19	28	2	2	**SUR**	9	19	28	2	2	**SUR**	13	15	28	1	0
57	Thailand	**THA**	15	32	47	4	3	**THA**	21	23	44	3	3	**THA**	17	28	45	4	3	**THA**	28	19	47	1	0
58	Turkey	**TUR**	5	42	47	2	2	**TUR**	3	41	44	1	1	**TUR**	3	42	45	1	1	**TUR**	19	28	47	1	0
59	Uganda	**UGA‡**	26	3	29	0	1	**UGA‡**	23	3	26	0	1	**UGA‡**	24	3	27	0	1	UGA					
60	Uruguay	**URY**	13	32	45	1	1	**URY**	12	32	44	1	1	**URY**	12	33	45	1	1	**URY**	21	24	45	1	0
61	Venezuela	**VEN‡**	2	45	47	0	1	**VEN‡**	3	41	44	0	1	VEN						VEN					
62	South Africa	**ZAF**	30	17	47	1	0	**ZAF**	30	14	44	1	0	ZAF						**ZAF**	30	17	47	1	0
63	Zambia	**ZMB**	24	20	44	1	0	ZMB						ZMB						**ZMB**	24	20	44	1	0
64	Zimbabwe	**ZWE‡**	24	8	32	0	1	ZWE						ZWE						ZWE					
Totals		61	1,249	1,194	2,443	78	42	55	1,034	1,017	2,051	66	35	50	922	1,000	1,922	68	34	41	783	887	1,670	41	0

Notes: This table presents the sample make-up of the dynamic 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.

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	D	A		0	1	Obs	D	A		0	1	Obs	D	A		0	1	Obs	D	A
1	Albania	**ALB**	12	18	30	2	1	**ALB**	12	15	27	2	1	**ALB**	10	18	28	1	0	**ALB**	11	12	23	1	0
2	Argentina	**ARG**	19	31	50	2	1	**ARG**	15	32	47	3	3	**ARG**	15	33	48	3	3	**ARG**	22	21	43	1	0
3	Armenia	**ARM**	2	18	20	2	1	ARM						ARM						**ARM**	7	6	13	1	0
4	Antigua & Barbuda	ATG						**ATG**	22	4	26	1	0	ATG						ATG					
5	Azerbaidjan	**AZE**	19	1	20	1	1	AZE						AZE						AZE					
6	Burundi	**BDI**	41	8	49	1	0	**BDI**	43	3	46	1	0	**BDI**	40	7	47	2	1	BDI					
7	Benin	**BEN**	30	20	50	1	0	**BEN**	30	17	47	1	0	**BEN**	30	18	48	1	0	**BEN**	30	13	43	1	0
8	Burkina Faso	**BFA**	47	3	50	1	1	BFA						BFA						BFA					
9	Bangladesh	**BGD**	19	20	39	2	2	**BGD**	16	21	37	1	1	**BGD**	17	21	38	1	1	**BGD**	19	13	32	1	0
10	Bulgaria	**BGR**	10	20	30	1	0	**BGR**	9	18	27	1	0	**BGR**	9	19	28	1	0	**BGR**	10	13	23	1	0
11	Belarus	**BLR**	16	4	20	1	1	**BLR**	14	3	17	1	1	BLR						BLR					
12	Bolivia	**BOL**	11	29	40	1	0	**BOL**	10	27	37	2	1	**BOL**	10	28	38	2	1	**BOL**	11	22	33	1	0
13	Brazil	**BRA**	21	29	50	1	1	**BRA**	15	32	47	1	1	**BRA**	21	27	48	1	1	**BRA**	24	19	43	1	0
14	Bhutan	**BTN**	27	3	30	1	0	BTN						**BTN**	26	2	28	1	0	BTN					
15	Central African Rep	**CAF**	40	10	50	1	1	**CAF**	37	10	47	1	1	**CAF**	38	10	48	1	1	CAF					
16	Chile	**CHL**	17	33	50	1	1	**CHL**	17	30	47	1	1	**CHL**	17	31	48	1	1	**CHL**	29	14	43	1	0
17	Cote d'Ivoire	**CIV**	46	2	48	1	1	CIV						CIV						CIV					
18	Rep of Congo	**COG**	43	7	50	1	2	**COG‡**	45	2	47	0	1	**COG**	41	7	48	1	2	COG					
19	Comores	**COM**	13	16	29	3	2	**COM**	25	2	27	1	0	**COM**	18	10	28	2	1	COM					
20	Cape Verde	**CPV**	10	20	30	1	0	**CPV**	10	17	27	1	0	**CPV**	9	19	28	1	0	**CPV**	10	13	23	1	0
21	Cyprus	CYP						**CYP**	1	31	32	1	0	**CYP**	7	26	33	1	0	CYP					
22	Czech Rep	**CZE**	2	18	20	1	0	CZE						**CZE**	2	16	18	1	0	CZE					
23	Djibouti	**DJI**	8	9	17	1	0	DJI						DJI						**DJI**	8	5	13	1	0
24	Dominican Rep	**DOM**	17	33	50	1	0	**DOM**	5	42	47	1	0	**DOM**	5	43	48	1	0	**DOM**	17	26	43	1	0
25	Ecuador	**ECU**	18	32	50	1	1	**ECU**	19	28	47	2	2	**ECU**	18	30	48	2	2	**ECU**	18	25	43	1	0
26	Spain	**ESP**	8	33	41	1	0	**ESP**	7	31	38	1	0	**ESP**	7	32	39	1	0	**ESP**	8	26	34	1	0
27	Ethiopia	**ETH**	14	15	29	1	1	ETH						ETH						**ETH**	13	9	22	1	0
28	Fiji	**FJI**	6	33	39	1	2	**FJI‡**	21	17	38	0	1	**FJI**	31	8	39	1	1	FJI					
29	Georgia	**GEO**	7	16	23	1	0	**GEO**	16	4	20	1	0	**GEO**	16	5	21	1	0	**GEO**	4	9	13	1	0
30	Ghana	**GHA**	31	19	50	3	2	**GHA**	32	15	47	3	2	**GHA**	27	21	48	3	2	**GHA**	35	8	43	1	0

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

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
61	Malawi	**MWI**	30	16	46	1	0	**MWI**	30	13	43	1	0	**MWI**	30	14	44	1	0	**MWI**	30	9	39	1	0
62	Niger	**NER**	33	12	45	2	1	**NER**	35	10	45	2	1	**NER**	36	9	45	2	1	NER					
63	Nicaragua	**NIC**	29	21	50	1	0	**NIC**	23	24	47	1	0	**NIC**	23	25	48	1	0	**NIC**	29	14	43	1	0
64	Nepal	**NPL**	29	16	45	2	1	**NPL**	31	11	42	1	1	**NPL**	30	13	43	2	1	NPL					
65	Pakistan	**PAK**	24	19	43	3	2	**PAK**	24	16	40	2	2	**PAK**	24	17	41	3	2	PAK					
66	Panama	**PAN**	13	17	30	1	0	**PAN**	10	17	27	1	0	**PAN**	8	20	28	1	0	**PAN**	13	10	23	1	0
67	Peru	**PER**	14	36	50	3	3	**PER**	24	23	47	3	3	**PER**	24	24	48	3	3	**PER**	19	24	43	1	0
68	Philippines	**PHL**	22	28	50	1	1	**PHL**	21	26	47	1	1	**PHL**	21	27	48	1	1	**PHL**	26	17	43	1	0
69	Portugal	**PRT**	6	35	41	1	0	**PRT**	6	32	38	1	0	**PRT**	6	33	39	1	0	**PRT**	6	28	34	1	0
70	Paraguay	**PRY**	1	18	19	1	0	**PRY**	11	5	16	1	0	PRY						**PRY**	1	11	12	1	0
71	Russia	**RUS**	10	11	21	1	1	**RUS**	11	7	18	1	1	RUS						**RUS**	2	11	13	1	0
72	Sudan	**SDN**	36	3	39	1	1	**SDN**	33	3	36	1	1	**SDN**	34	3	37	1	1	SDN					
73	Senegal	**SEN**	39	11	50	1	0	**SEN**	39	8	47	1	0	**SEN**	39	9	48	1	0	**SEN**	39	4	43	1	0
74	Solomon Isl.	**SLB**	4	7	11	1	1	**SLB**	6	5	11	1	1	SLB						SLB					
75	Sierra Leone	**SLE**	20	11	31	2	1	**SLE**	22	6	28	1	0	**SLE**	17	12	29	2	1	SLE					
76	El Salvador	SLV						SLV						SLV						**SLV**	7	10	17	1	0
77	Suriname	**SUR**	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	**SVK**	5	15	20	1	0	**SVK**	5	16	21	1	0	SVK					
79	Slovenia	**SVN**	1	19	20	1	0	SVN						SVN						**SVN**	1	12	13	1	0
80	Thailand	**THA**	18	32	50	4	3	**THA**	24	23	47	3	3	**THA**	20	28	48	4	3	**THA**	31	12	43	1	0
81	Turkey	**TUR**	5	45	50	3	2	**TUR**	3	44	47	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	**UGA**‡	23	3	26	0	1	**UGA**‡	24	3	27	0	1	UGA					
84	Ukraine	**UKR**	4	17	21	1	0	**UKR**	1	17	18	1	0	**UKR**	1	18	19	1	0	**UKR**	3	10	13	1	0
85	Uruguay	**URY**	13	35	48	1	1	**URY**	12	35	47	1	1	**URY**	12	36	48	1	1	**URY**	24	19	43	1	0
86	Venezuela	**VEN**‡	2	48	50	0	1	**VEN**‡	3	44	47	0	1	VEN						VEN					
87	South Africa	**ZAF**	33	17	50	1	0	**ZAF**	33	14	47	1	0	ZAF						**ZAF**	33	10	43	1	0
88	Zambia	**ZMB**	27	20	47	1	0	ZMB						ZMB						**ZMB**	27	13	40	1	0
89	Zimbabwe	**ZWE**	26	9	35	1	1	ZWE						ZWE						ZWE					
Totals		83	1,568	1,484	3,052	105	58	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.[45] 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.[46] 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).[47] 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 Murtin 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.

[45]Many 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).

[46]In 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.

[47]The 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	ΔGDPpc 1960-85	GDP pc (log), investment, schooling (all in 1960, restrictions imposed following MRW)	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	ΔGDPpc 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 ΔGDPpc	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	ΔGDPpc in non-overlapping 5-year periods	GDP pc (log), schooling attainment, (all lagged by 5 years)	n=485, 1960-89	**concave (5% level)**; *insig. +ve changes*, **sig. –ve changes**	Table 2[1]
Tavares & Wacziarg (2001)	3SLS	**Continuous**, Bollen and Freedom House indices	Annual ΔGDPpc	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 ΔGDPpc	GDP pc (log), life expectancy, investment, labour force, HC (all lagged), various lags of ΔGDPpc	N=128, n=548, 1967-97	+ *(insign.)*	Table 1[2]
Gerring et al (2005)	OLS w/ country FE	**Continuous**, Democracy **stock** (1900-2000) based on continuous polity2 $(-10, +10)$	Annual ΔGDPpc	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 ΔGDPpc	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 ΔGDPpc	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 ΔGDPpc	Dummy for socialist regimes (interacted with democratisation), continent dummies	N=138, n=4,387, 1960-2000	+ *(insign.)*	Table 1[8]
Rodrik & Wacziarg (2005)	OLS w/ country FE	**Dummy**: New Democracy, Established Democracy, etc (Polity IV-based)	Annual ΔGDPpc	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
Persson & Tabellini (2006)	OLS w/ period FE	**Dummy** for polity2> 0	Annual ΔGDPpc	Continent dummies, legal origin, lagged GDP pc (log)	N=138, n=4,338, 1960-2000	**+ (5% level)**	Table 1[1]
	dto.	dto.	Annual ΔGDPpc	Continent dummies, legal origin, lagged GDP pc (log)	N=148, n=8,135, 1850-2000	**+ (10% level)**	Table 3[3]
Persson & Tabellini (2009)	2FE	**Continuous**, domestic and foreign democratic **stock** (PIM, 1800-2000) based on dummy (= 1 if polity2> 0)	Annual ΔGDPpc	lagged GDP pc (log)	n=8,379, 1820-2000	**+ (1% level) domestic**; *insign. foreign*	Table 5[1]
Papaioannou & Siourounis (2008)	Diff-in-Diff	**Dummy** building on FHI and polity2	Annual ΔGDPpc	none	N=166, n=5,410, 1960-2005	**+ (1% level)**	Table 2[4]
	Diff-in-Diff	dto.	Annual ΔGDPpc	lagged GDP pc (log) and lagged growth rate, investment	N=166, n=5,410, 1960-2005	**+ (1% level)**	Table 3[1]
Knutsen (2013)	OLS w/ period FE	**Continuous**, Freedom House Index*	Annual ΔGDPpc	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 ΔGDPpc	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	+ *(insign.)*	Table 11[1]
	2FE	dto.	dto.	dto.	N=69, n=308, 1960-2000	– *(insign.)*	Table 11[10]
	GMM BB	dto.	dto.	dto.	N=69, n=489, 1870-2000	+ *(insign.)*	Table 11[3]
	GMM BB	dto.	dto.	dto.	N=68, n=275, 1960-2000	+ *(insign.)*	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 $\rightarrow$ +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.	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 Main Results — Static Specification

Table C-1: Main Results – Static Specifications

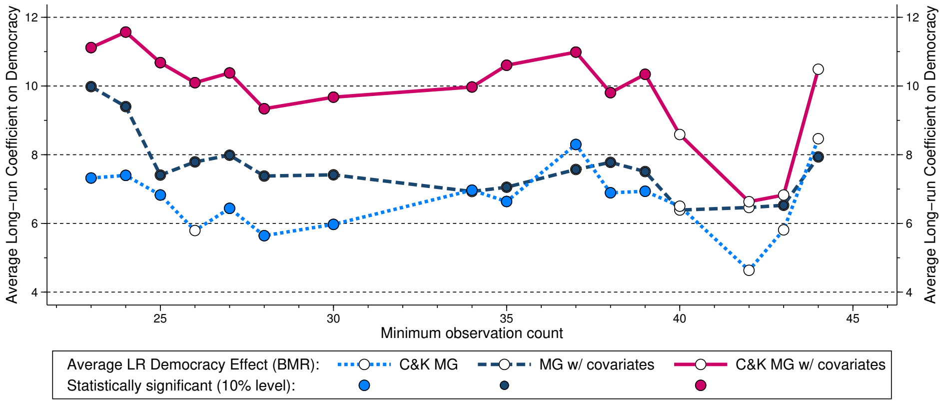
	Plain Vanilla		With Covariates	
	(1)	(2)	(3)	(4)
Implementation	MG	C&K MG	MG	C&K MG
Parameters estimated ‡	$2 \times N$	$5 \times N$	$4 \times N$	$7 \times N$
Democracy (ANRR)	10.249	4.402	3.846	4.016
	(3.277)***	(2.088)**	(2.598)	(1.983)**
Observations	3052	3052	3052	3052
Countries (N)	83	83	83	83
Democratisations	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
Democracy (BMR)	10.629	4.168	5.505	4.260
	(3.394)***	(2.242)*	(2.847)*	(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
Democracy (CGV)	12.849	2.862	6.853	4.991
	(3.739)***	(2.730)	(2.837)**	(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
Democracy (PS)	21.990	4.669	11.296	4.874
	(4.636)***	(3.146)	(3.538)***	(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.

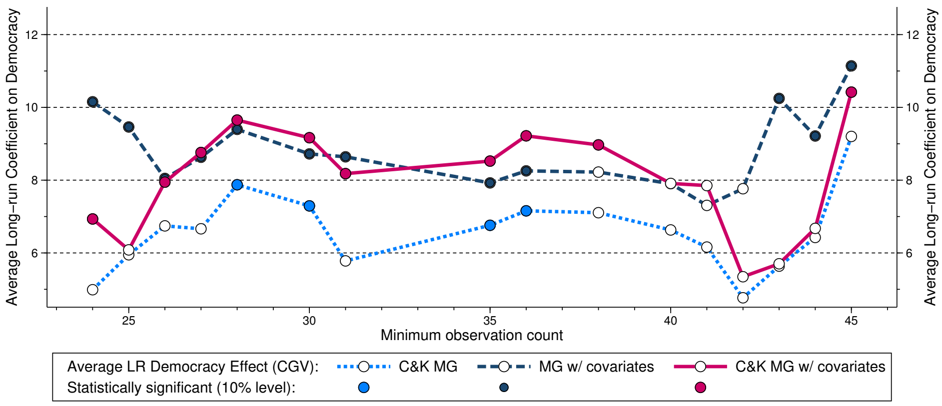
D Sample Reduction Exercises – More Results

D.1 Alternative Definitions of Democracy

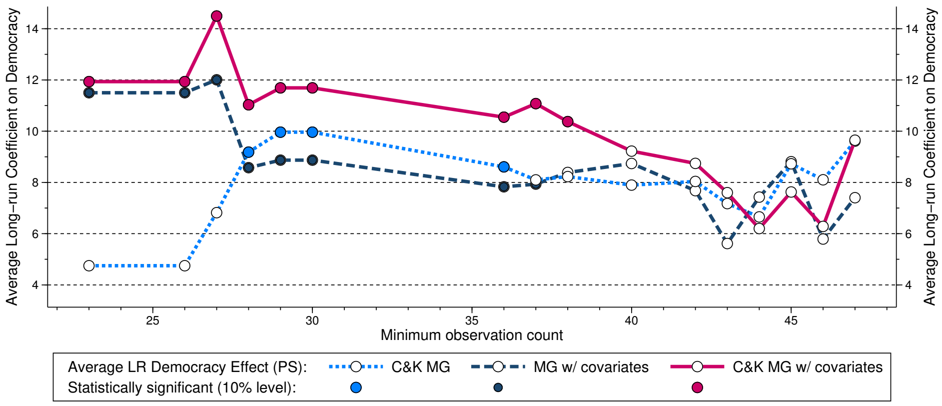
Figure D-1: Sample Reductions — minimum T_i



(a) Boix et al (2013)



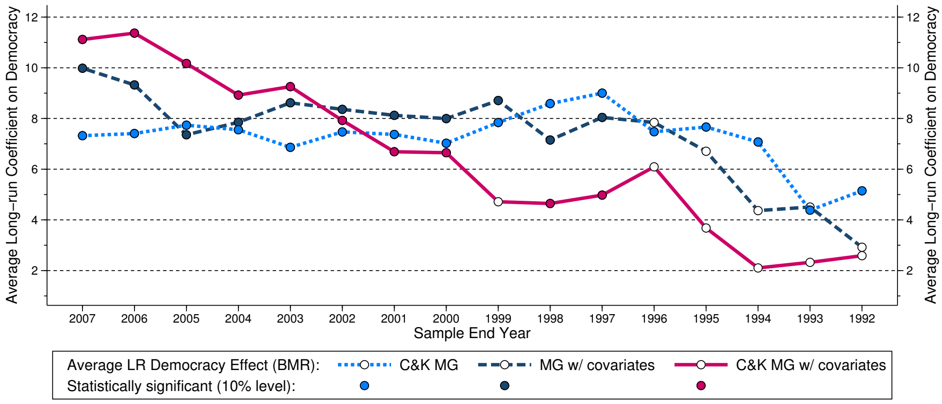
(b) Cheibub et al (2010)



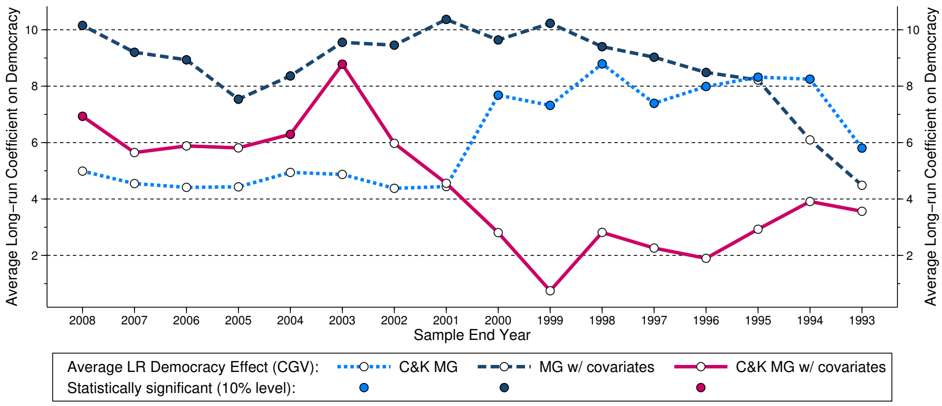
(c) Papaioannou and Siourounis (2008)

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 4 for a comparison with the results for the ANRR definition of democracy.

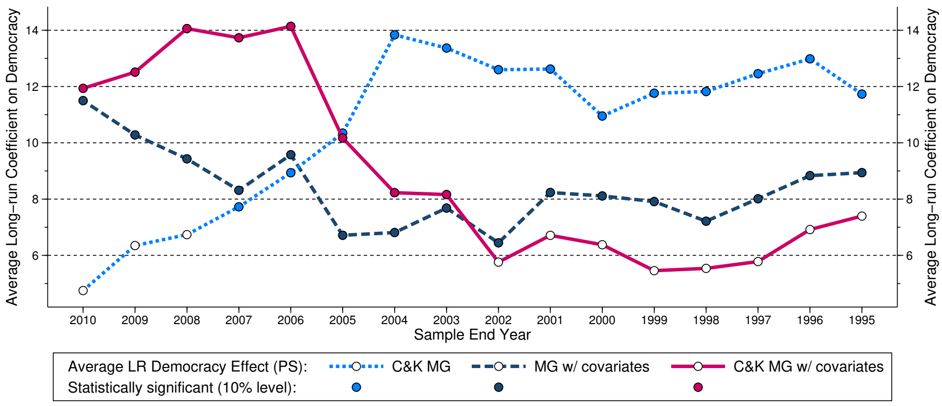
Figure D-2: Sample Reductions — end year



(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 4 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 3 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 E-3 I speculate about one potential souce 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:[48]

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

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.[49] 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 disporportionate 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%

[48] My 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.

[49] ANRR 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.[50] 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%.[51]

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 3 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.[52]

[50] Results 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.

[51] Note 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!

[52] ANRR 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 **??** 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 (a) charts the distribution of sample observations and democratisation/reversal events by year — here and in panel (b) 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.[53]

Panel (b) 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 3 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 signficance 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 to (statistically-speaking) disappear by the omission of three former Soviet Republics

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

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.

E.2 ANRR: Some forensic analysis

What are the reasons for this sensitivity of results to relatively small numbers of observations? In the following I indicate that the source of this puzzle is possibly related to parameter heterogeneity — my focus here is not on the estimates for democracy, but on the estimates for the GDP dynamics as the sample changes: *a priori* we do not know what the magnitude or even the sign of the democracy coefficient $\hat{\beta}$ in the dynamic 2SLS regression 'should' be (the literature has argued for positive or negative effects), but we know that the estimate for the GDP dynamics should be positive and somewhere below but fairly close to 1. I limit my attention to the 4-lag 2SLS specification, where I plot the estimate and 90% confidence interval for the GDP dynamics (solid line), i.e. $\sum_{\ell=1}^{4} \hat{\rho}_\ell$: in panel (a) of Figure E-3 I drop countries by number of observations, and in panel (b) I drop observations by end year.

Since all of the parametric models studied above are pooled models, the democracy coefficient as well as the GDP dynamics are assumed to be *common* across countries. A high (low) coefficient on the GDP dynamics *ceteris paribus* implies a higher (lower) long-run coefficient on democracy in absolute terms: $\hat{\beta}^{LR} = \hat{\beta}/(1 - \sum_{\ell=1}^{4} \hat{\rho}_\ell)$. Figure E-2 plots (among others) the 2SLS $\hat{\beta}^{LR}$ for democracy, while Figure E-3 plots the estimated GDP dynamics used in computing these long-run democracy estimates and associated standard errors. For either sample reduction strategy the estimate on the GDP dynamics (solid line) is remarkably stable across samples, especially given the sensitivity of the long-run democracy coefficients in Figure E-2.

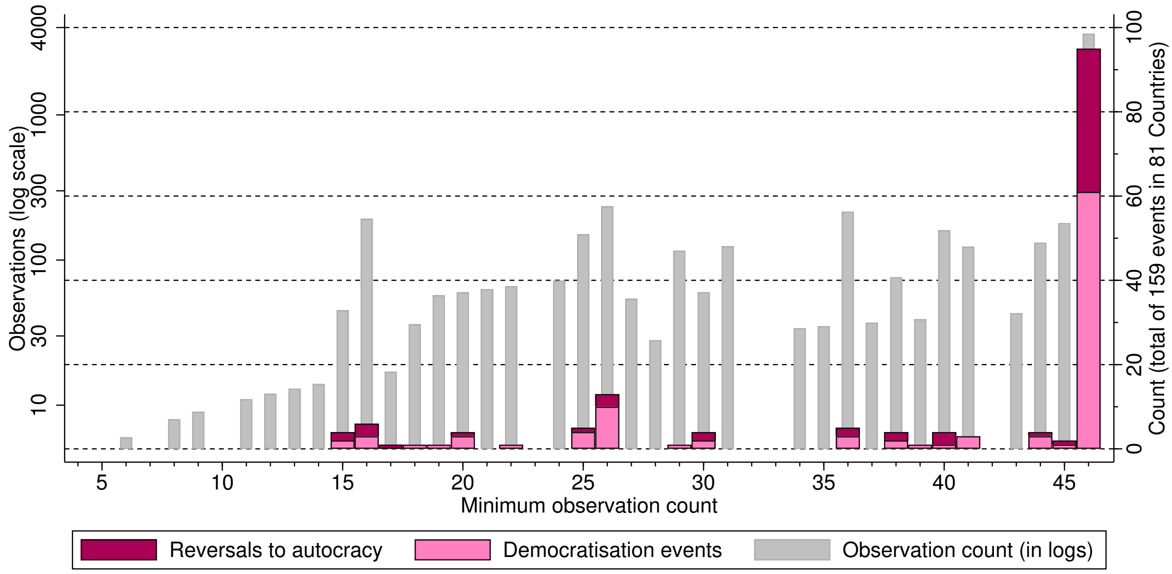
So what is the estimate on the GDP dynamics in the countries or years I omit? The dashed line in panel (a) of Figure E-3 represents the estimated GDP dynamics for all countries with a minimum observation count *lower* than that indicated on the x-axis:[54] as we move to the right these countries are dropped from the sample estimating the solid line and included in the sample estimating the dashed line. It is noticeable that, with the exception of two, all of these estimates for GDP dynamics in the sample of 'dropped' countries are *below* those for the 'included' countries. For some samples toward the right end of the graph the confidence intervals of the two sets of estimates do not overlap.[55] Similarly, in panel (b) the estimates on the GDP dynamics for the omitted end years are substantially below those of the included years, the patterns for 2008 and 2009 even speak to those of the results in panel (b) of Figure E-2.

Thus, if GDP dynamics differ between countries in general, and between my samples of countries/years included and omitted in particular, then the inclusion of these 'omitted' countries or years may inflate the long-run democracy coefficients in the full sample results.

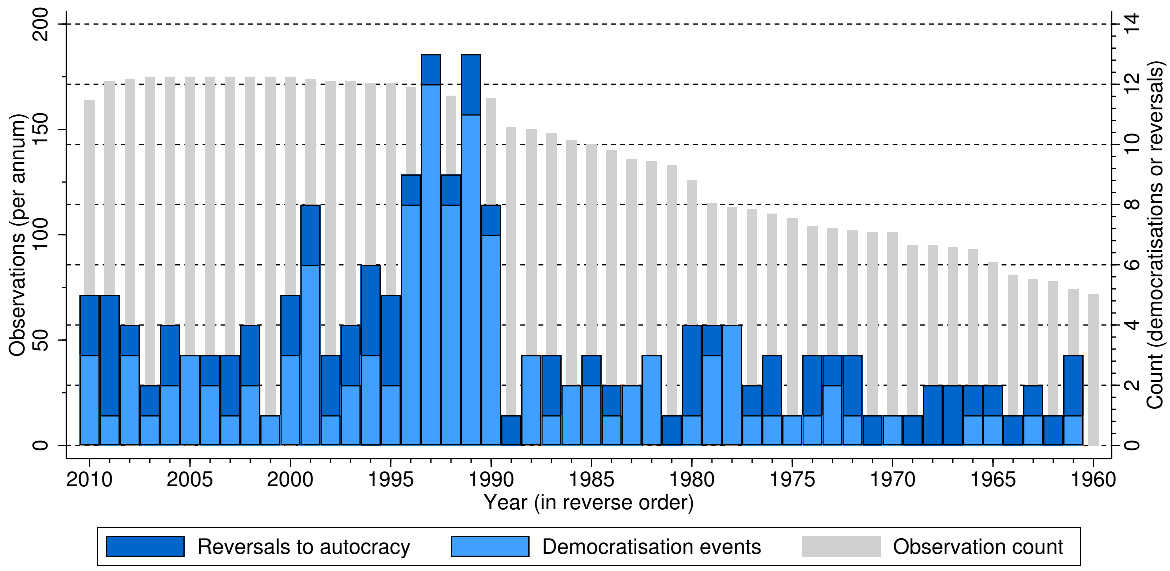
[54]For $T_{min} = 6$ this estimate would be constructed from 6 observations in one country. I therefore only begin charting this estimate for countries with 17 or fewer observations (338 observations in 23 countries).

[55]I do not show the confidence intervals for the 'drop-out' estimates for ease of illustration.

Figure E-1: Sample and Event Distribution – ANRR



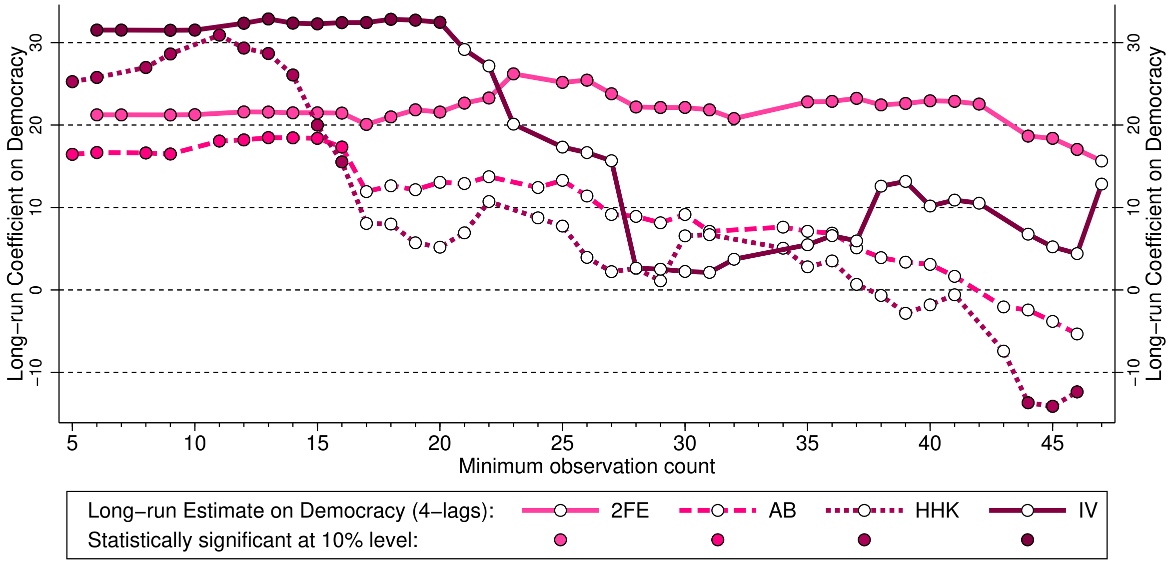
(a) Sample Reduction by T_i count



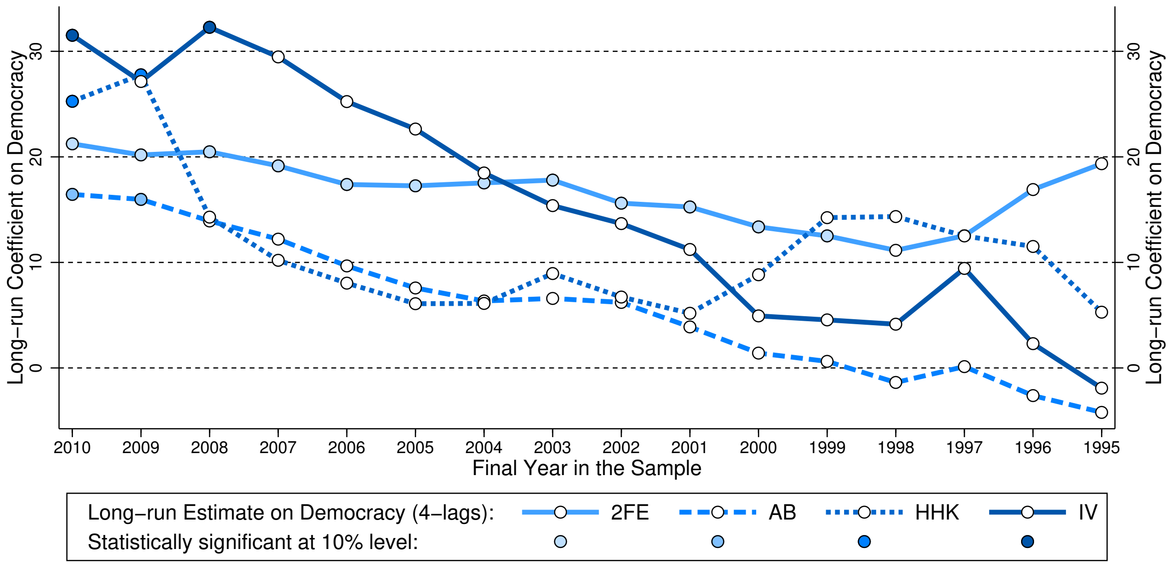
(b) Sample Reduction by end year

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



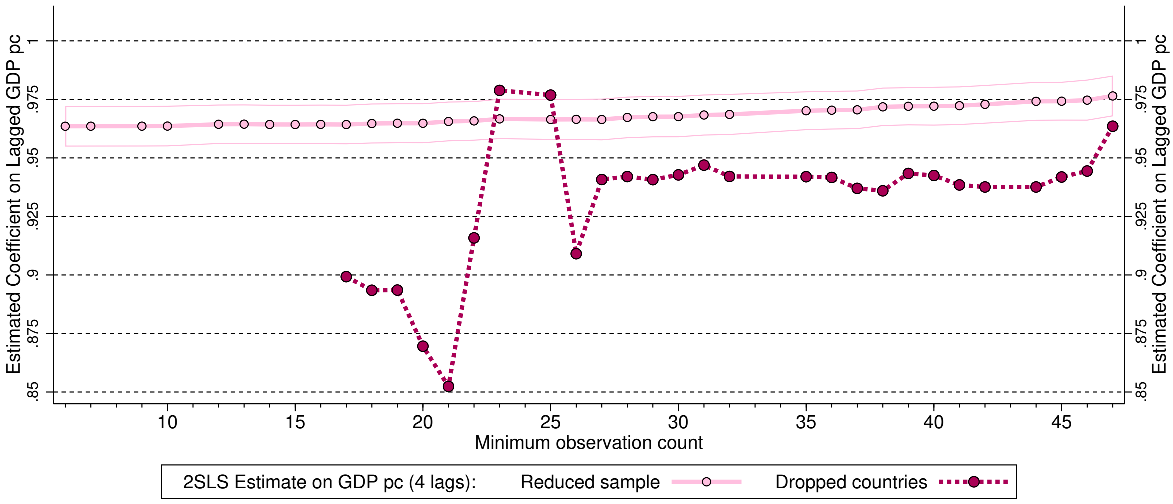
(a) Sample reduction by T_i count



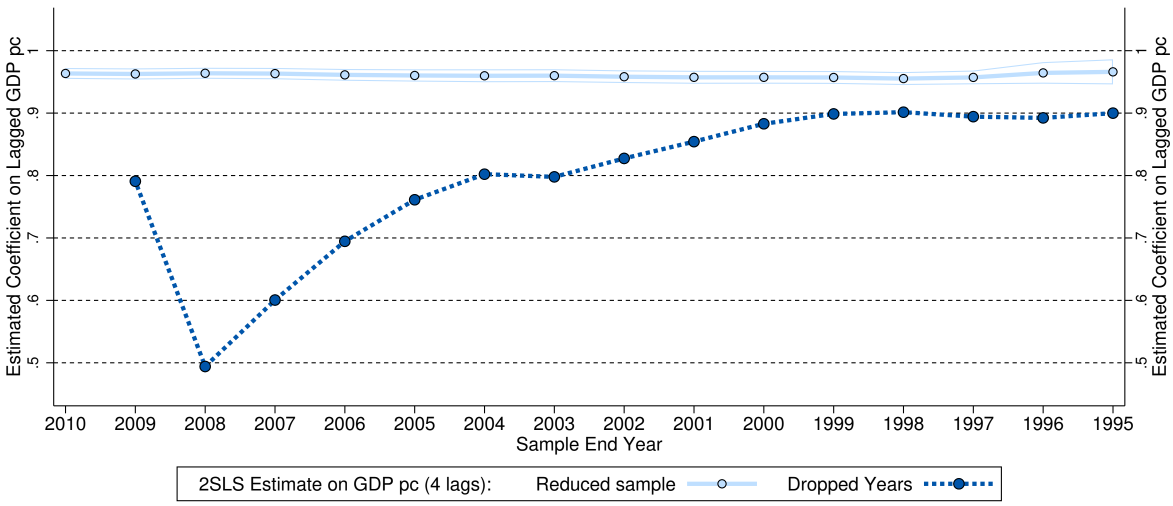
(b) Sample reduction by end year

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%.

Figure E-3: The GDP Dynamics of 'Omitted' Countries — ANRR



(a) Sample reduction by T_i count



(b) Sample reduction by end year

Notes: The plots present estimates on the sum of lagged GDP terms in the 4-lag 2SLS regressions for the sample reduction by T_i count in panel (a) and for the reduction by end year in panel (b): $\sum_{\ell=1}^{4} \hat{\rho}_{i,t-\ell}$, where the $\hat{\rho}$ are the coefficients on the lags of per capita GDP (standard errors are constructed via the Delta method). Each panel plots two series, the estimates (i) for the reduced sample (solid line with 90% CI), and (ii) for those countries or end years which are dropped (dashed line), e.g. the 2009 estimate in panel (b) is for the years 2010 and 2009.

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 YEM	CZE LTU MKD 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 CHL CIV COG DOM ECU ESP GHA GRC GTM HND IDN KEN KOR LBR MDG MEX MRT MWI NER NGA NIC NPL PAK PAN PER PHL PRT PRY SDN SEN SLE SLV THA TUR URY VEN ZAF ZMB	CHN CMR DZA EGY GAB MAR MYS RWA SYR TCD TGO ZAR	AUS AUT BEL CAN COL CRI DNK FIN FRA GBR IND ISL ISR ITA JPN LKA LUX NLD 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

	Sample reduction by T_i count				Sample reduction by end year			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Estimator	2FE	AB	HHK	2SLS	2FE	AB	HHK	2SLS
Panel A: Full ANRR sample estimates								
Long-Run Democracy Effect	21.240	16.448	25.268	31.521	21.240	16.448	25.268	31.521
	[7.215]***	[8.436]*	[10.869]**	[17.425]*	[7.215]***	[8.436]*	[10.869]**	[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	11.932	8.066	29.168	12.516	3.891	14.293	27.145
	[9.867]	[8.071]	[7.047]	[17.733]	[7.386]	[8.131]	[11.504]	[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	3.949	2.651	1.160	3.891	-22.917	4.936
		[7.622]	[5.670]	[16.519]	[6.157]	[8.131]	[28.970]	[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	5.718	2.651	1.160	1.411	6.091	2.305
		[8.722]	[6.287]	[16.519]	[6.157]	[8.409]	[8.090]	[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	-5.337	-12.358	12.843	n/a	n/a	n/a	n/a
	[9.867]	[8.484]	[6.899]*	[23.009]				
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], respectivly (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 prefered 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

	Sample reduction by T_i count							
	[1]		[2]		[3]		[4]	
	2FE		AB		HHK		2SLS	
ANRR Reference	Table 2(3)		Table 2(7)		Table 2(11)		Table 6(2)A	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Long-run effect	21.240	15.637	16.448	12.846	25.032	9.221	31.521	28.574
of democracy	[7.215]***	[9.867]	[8.436]*	[8.023]	[10.581]***	[5.830]	[17.425]*	[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.3 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.):

$$y_{it} = \alpha_i + \gamma_t + \beta\,\text{Democracy}_{i,t-1} + \delta HC_{i,t-1} + \rho y_{i,t-1} + \varepsilon_{it}, \tag{7}$$

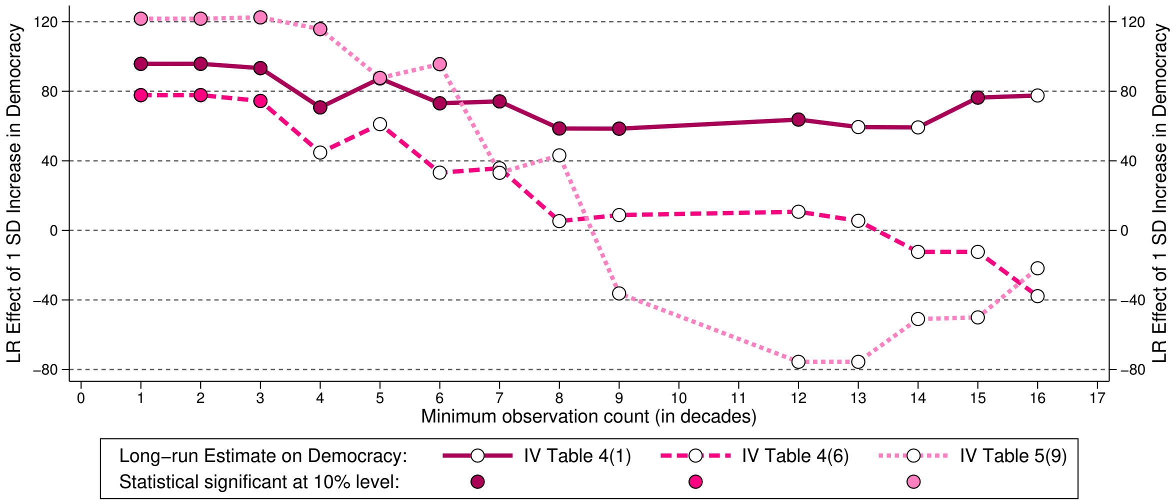
where y is the log of real per capita GDP (in PPP values), Democracy is a the (continuous) polity2 variable, α_i and γ_t are country and time dummies, respectively, and HC is human capital proxied by literacy. Democracy is instrumented using the linguistic distance-weighted average of 'foreign' democracy, 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.

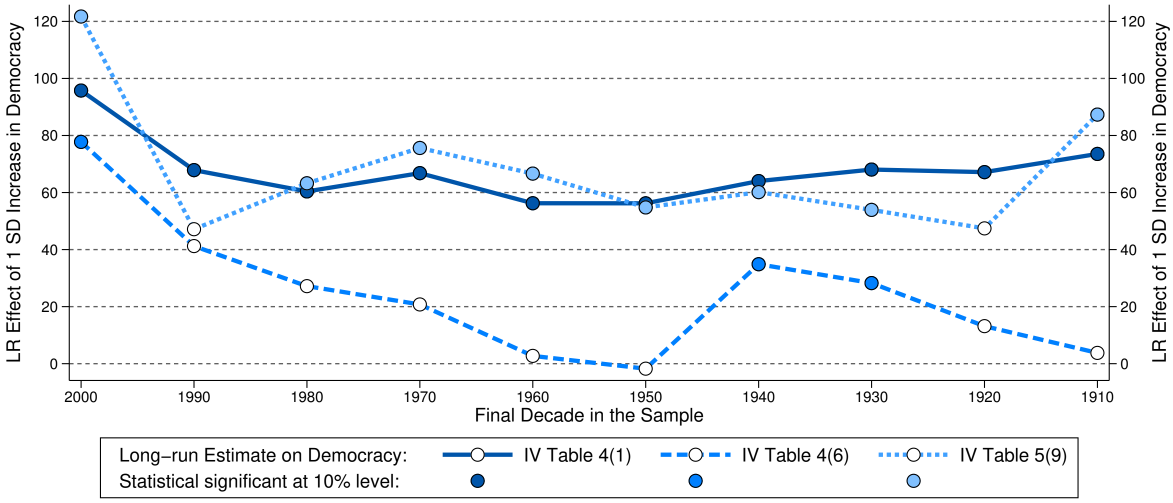
Sample reduction by minimum observation count Panel (a) of Figure E-4 provides the decade-by-decade results for the sample reduction by minimum observation count, columns [1] to [3] of Table E-4 the results for the full sample, for the sample when the democracy estimate turns statistically insignificant 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 throught, 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-4: 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	77.763	121.708	95.758	77.763	121.708
	[25.745]***	[30.512]**	[35.549]***	[25.745]***	[30.512]**	[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	44.752	33.093	n/a	41.243	47.151
	[40.168]	[28.707]	[61.963]		[26.932]	[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	33.093	n/a	27.189	47.151
		[8.722]	[61.963]		[24.587]	[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), respectivly (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.