J Econ Growth (2014) 19:141–181
DOI 10.1007/s10887-013-9100-6

The democratic transition

Fabrice Murtin · Romain Wacziarg

Published online: 17 December 2013

Abstract Over the last two centuries, many countries experienced regime transitions toward democracy. We document this democratic transition over a long time horizon. We use historical time series of income, education and democracy levels from 1870 to 2000 to explore the economic factors associated with rising levels of democracy. We find that primary schooling, and to a weaker extent per capita income levels, are strong determinants of the quality of political institutions. We find little evidence of causality running the other way, from democracy to income or education.

Keywords Democracy · Modernization · Human capital · GMM

JEL Classification I25 · N30 · N40 · O43

1 Introduction

Over the last two centuries, many countries underwent a democratic transition, moving from autocratic regimes with low popular participation in political decision-making and weak con-

Electronic supplementary material The online version of this article (doi:10.1007/s10887-013-9100-6) contains supplementary material, which is available to authorized users.

F. Murtin
Organisation for Economic Co-operation and Development, Statistics Directorate, and Centre for the Economics of Education (London School of Economics and Political Science),
2 rue André Pascal, 75016 Paris, France
e-mail: fabrice.murtin@oecd.org

R. Wacziarg (✉)
UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095, USA
e-mail: wacziarg@ucla.edu

R. Wacziarg
NBER, Cambridge, MA, USA

R. Wacziarg
CEPR, London, UK

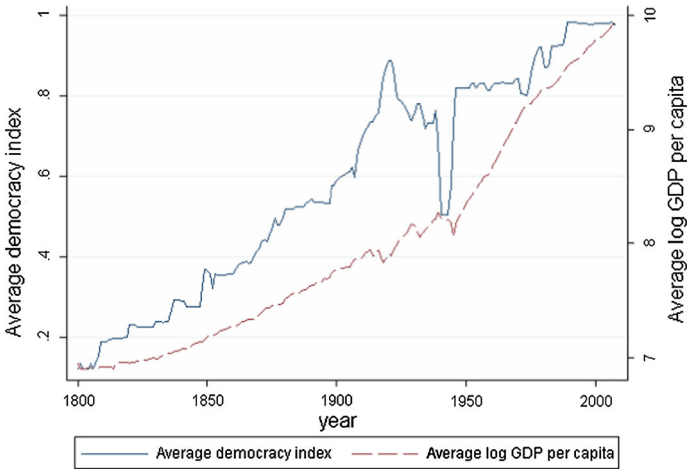

Fig. 1 The democratic and economic transitions. The balanced sample is composed of Austria, Belgium, Chile, Denmark, France, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Turkey, United Kingdom, and the United States over the 1800–2000 period. Log GDP per capita is taken from Maddison (2006) and democracy is the Polity index. Missing observations are interpolated

straints on the exercise of executive power to more democratic regimes with broader political participation and greater limits on the exercise of political power. Figure 1 displays the average democracy score for a balanced panel of countries since 1800, using a commonly used measure of democracy (the Polity index). While democratization happened in fits and starts, and did not happen everywhere, there is a general upward trend: the democratic transition.

The democratic transition partly overlaps with other socioeconomic trends, chiefly the demographic transition, and the process of industrialization and modernization with which it is associated. As shown in Fig. 1, the average level of GDP per has followed the same upward trend as the democracy index in a balanced panel. What are the links between these two transitions? In particular, to what extent did socioeconomic modernization affect the transition toward democratic institutions? This paper takes a long view of history, going back to 1870, to untangle the relationships between democratization and socioeconomic modernization, captured by expanding primary education and rising per capita income levels. We focus mainly on the question of whether education and income are determinants of democracy, and seek to assess their respective impacts on the long-term process of democratization.

We examine the two-way relationship between education (or income) and democracy in a large panel of countries starting in 1870, using a variety of modern dynamic panel data methods, including system GMM and recently developed dynamic panel bias correction methods. In a first step, we address the link from economic development to democracy, concentrating on two dimensions of socioeconomic modernity: GDP per capita and average years of schooling among the adult population. In a second stage, we investigate the potential effect of democracy on childhood education, measured by the average years of schooling completed by young cohorts, as well as on the (log) level or the growth rate of GDP per capita.

This paper is related to a large literature on the dual relationship between economic development and democracy. One strand of this literature investigated the relationship going from economic development to democracy. In "A Bill for the More General Diffusion of Knowledge" (1779), Thomas Jefferson argued that education was a cornerstone of democracy: "the

most effectual means of preventing tyranny would be, to illuminate, as far as practicable, the minds of people at large, and more especially to give them knowledge of those facts, which history exhibiteth, that, possessed thereby of the experience of other ages and countries, they may be enabled to know ambition under all its shapes, and prompt to exert their natural powers to defeat its purposes". This view was shared by other observers throughout history. For instance, Tocqueville (1835) noted that mass education was a significant factor in sustaining the vigor of democracy in the United States, a leading country in terms of educational attainment over the nineteenth century:[1] "the education of the people powerfully contributes to the maintenance of the democratic republic. That will always be so, in my view, wherever education to enlighten the mind is not separated from that responsible for teaching morality".[2] Building on ideas that go back in time as far as Aristotle, Lipset (1959) argued in a seminal article that improvements in economic standards would ultimately lead to democratization: "the more well-to-do a nation, the greater the chances that it will sustain democracy" . Lipset's modernization hypothesis received empirical support from Barro (1999), who found, among other variables, that GDP per capita and primary schooling were positive determinants of democracy in a large sample of countries spanning the period from 1960 to 1995. More recently, Boix (2011) found evidence of a positive effect of income per capita on democratization over a long period, and Treisman (2011) found evidence of a positive effect of income on democracy, more strongly over the medium run than over the short run.[3] However, neither contribution uses dynamic panel data estimation methods nor examines the effect of education variables.

In contrast to studies finding evidence supportive of the modernization hypothesis, in two recent studies Acemoglu et al. (2005, 2008) revisited and questioned the empirical significance of the effect of income on democracy. They found that, after controlling for country fixed-effects, GDP per capita was no longer a significant determinant of democracy.[4] In other recent studies, Bobba and Coviello (2007) as well as Castelló-Climent (2008) re-examined the evidence. The former isolated a significant effect of GDP per capita using a system-GMM estimator, while the latter found a significant effect of education attained by the majority of population, even after controlling for fixed-effects. However, Benhabib et al. (2011) recover a significant effect of GDP per capita even when controlling for fixed effects once taking into account the censoring of the democracy index. In sum, there remains a debate on the validity of Lipset's hypothesis, a debate to which the current paper contributes by taking a long view of history, applying dynamic panel econometric methods, and focusing on the role of education in addition to income.

Another strand of the literature explored the consequences of democratization on economic performance and on the expansion of schooling. The evidence in favor of such effects is mixed. For instance, Barro (1997) found a nonlinear effect of democracy on growth, with an initial increase in growth and a negative relation once a given level of democracy has been

[1] See Lindert (2004), Murtin and Viarengo (2010), and Morrisson and Murtin (2009).

[2] Democracy in America, Volume 1, chapter 9, "the main causes that tend to maintain a democratic republic in the United States".

[3] See also Gundlach and Paldam (2012) for a discussion of long run relationship between economic development and democracy, arguing that causality runs from the former to the latter.

[4] Using a more micro approach in Kenya, Friedman et al. (2011) also find that "increased human capital did not produce more pro-democratic or secular attitudes and strengthened ethnic identification". Whether this holds more broadly at the level of countries and realized democratic institutions rather than attitudes remains debated. Another view is Fayad, Bates and Hoeffler (2011), who uncover a *negative* effect of income shocks on democracy. Finally, Grosjean and Senik (2011) fail to find a significant effect of economic liberalization on democratization, suggesting instead that democratization may facilitate economic liberalization.

reached. Tavares and Wacziarg (2001) uncovered a positive effect of democracy on human capital accumulation and a negative effect on physical capital accumulation, while Rodrik and Wacziarg (2005) found a positive short-term effect of democratization on economic growth, but no effect at longer horizons and no effect when democracy was sustained for long periods. More specifically related to education, Lindert (2004) documented how the extension of the franchise in Europe's rising democracies gradually led to the introduction of public funding for education over the nineteenth and twentieth centuries, but he did not provide empirical evidence on the direction of causality between schooling and democracy.

The two-way relationship between education and democracy is at the core of an important debate on the fundamental sources of economic prosperity. Does education help raise the quality of institutions as well as productivity, or is an efficient institutional framework a prerequisite for expanding education levels and economic growth? The issue of the direction of causality between education and democracy has ignited a debate between the advocates of institutions as the prime engine of growth (Acemoglu et al. 2001, 2002), and the proponents of a different thesis, who view human capital as the root cause of economic development (Glaeser et al. 2004, 2007; Glaeser and Campante 2009). Sorting out the respective effects of democracy and education upon each other would shed some light on this broader debate.

This paper makes three contributions that jointly set it apart from the existing literature. First, using the Morrisson and Murtin (2009) dataset of educational attainment in 74 countries since 1870, we find strong empirical support for the modernization hypothesis. Second, we show that primary schooling, more so than GDP per capita, and more so than secondary and tertiary education, has been a major factor in the democratic transition over 1870–2000. Third, using state-of-the-art dynamic panel data techniques such as system GMM and a recently developed bias-correction estimator by Bun and Carree (2005), we address the issue of reverse causality and find no robust evidence of an effect of democracy on education or GDP per capita. In doing so, we incorporate recent advances in dynamic panel data methods, and address issues of weak instruments, instrument proliferation and violations of stationarity assumptions that have been found to be endemic to cross-country analyses relying on dynamic panel specifications.

Section 2 describes empirical regularities related to democracy and development over the 1870–2000 period. Section 3 describes the econometric specification and estimation methods used in this paper. Section 4 reports empirical tests the modernization hypothesis using a long time series, while Sect. 5 addresses the reverse relationship, from democracy to education and income. The last section concludes.

2 Democracy and development, 1870–2000

In this section, we describe the data and take a bird's-eye view of trends in democracy, income and schooling since 1870. Since differences across countries are central to our analysis, we focus on convergence effects and on three particular periods: 1870–1910, 1910–1960, and 1960–2000. These broadly delineate the two periods of globalization and the protectionist period of the interwar years. The full sample is composed of 70 countries spanning every continent, of which 19 countries form a balanced panel starting in 1870.[5] Table 1 provides descriptive statistics.

[5] When studying the relationship between democracy and income, the maximal sample size is 69 countries, while it is 70 countries when focusing only on education.

Table 1 Descriptive statistics

	Full sample				Balanced sample			
	1870	1910	1960	2000	1870	1910	1960	2000
Democracy								
Average	0.41	0.60	0.58	0.77	0.44	0.61	0.64	0.95
Coefficient of variation	0.74	0.52	0.65	0.36	0.73	0.56	0.62	0.09
p25	0.20	0.38	0.15	0.65	0.25	0.30	0.15	0.90
p75	0.65	0.90	1.00	1.00	0.70	0.95	1.00	1.00
GDP per capita								
Average	1523	2913	3825	8465	1489	2947	6109	14602
Coefficient of variation	0.52	0.48	0.82	0.98	0.53	0.50	0.56	0.48
p25	737	1694	1378	1433	719	1794	3072	7218
p75	2003	4064	6230	16010	1876	4198	8753	20321
Average years of schooling								
Total								
Average	2.80	3.98	4.17	7.28	2.53	4.09	5.95	9.51
Coefficient of variation	0.72	0.63	0.72	0.45	0.74	0.64	0.46	0.25
Primary								
Average	2.16	3.00	3.07	4.45	2.02	3.18	4.24	5.49
Coefficient of variation	0.72	0.60	0.64	0.33	0.73	0.58	0.35	0.12
Secondary								
Average	0.61	0.94	1.02	2.44	0.49	0.87	1.57	3.43
Coefficient of variation	0.98	0.91	1.09	0.68	1.01	0.99	0.83	0.46
Tertiary								
Average	0.03	0.04	0.09	0.39	0.02	0.04	0.14	0.59
Coefficient of variation	1.17	1.00	1.11	0.89	1.06	0.90	0.96	0.65
N[1]	27	32	59	69	19	19	19	19

[1]Number of countries with information available on democracy, average years of schooling as well as GDP per capita

2.1 Democracy

Our main measure of democracy is the combined polity score from the Polity IV dataset (Marshall and Jaggers 2008), rescaled between 0 (full autocracy) and 1 (full democracy). This widely-used index consists of underlying components capturing key characteristics of executive recruitment, constraints on executive authority, and the degree of political competition. We plot the world distribution of democracy scores in Fig. 2 (unbalanced panel) and Fig. 3 (balanced panel). Boxplots represent the 25th and 75th percentiles as well as the median of the distribution. These two figures illustrate an overall increase in democracy between 1870 and 1920, followed by a marked decrease until the Second World War. The immediate postwar period witnessed a new rise in democracy that lasted until the end of our period in the balanced panel. In the unbalanced panel, democracy stagnated or even regressed between 1960 and 1980 but rose again since then. The world distribution of democracy widened from 1930 onwards, while at the end of the period, cross-country differences in democracy fell dramatically.

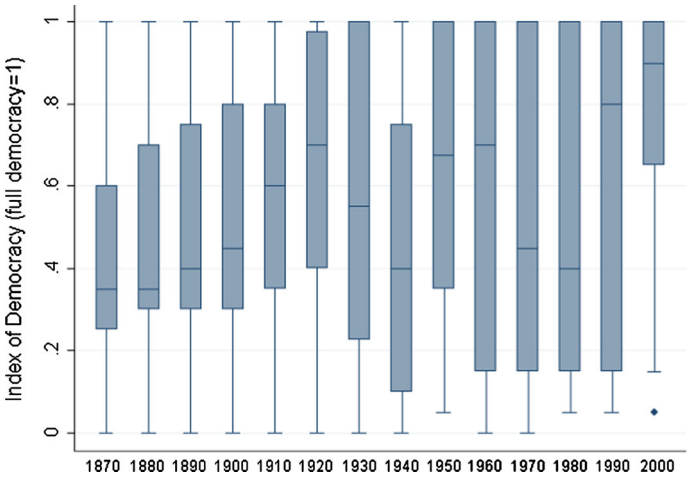

Fig. 2 Distribution of the democracy index over time—unbalanced panel of 70 countries

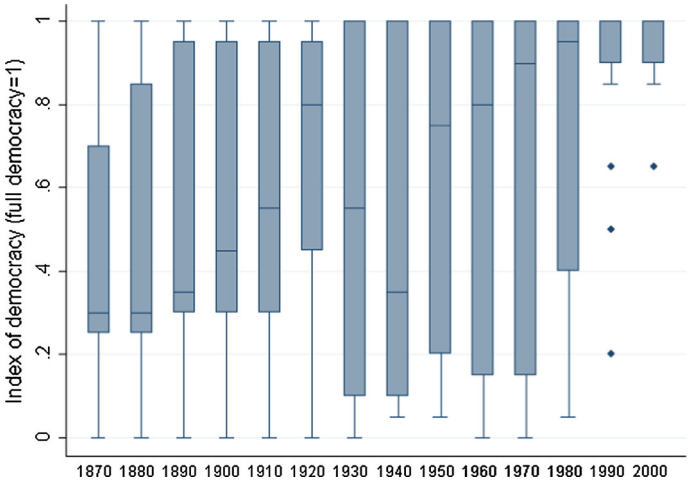

Fig. 3 Distribution of the democracy index over time—balanced panel of 19 countries

The latter observation is related to global inequality in democracy: Consistent with findings in Barro (1999) for the postwar period, results from the full sample show that democracy has indeed converged within any of our three subperiods, at roughly the same pace. Indeed, when calculated on an annual basis, the speed of convergence of democracy is equal to 4.7 % over the 1870–1910 period, 3.4 % in 1910–1960 and 3.7 % in 1960–2000.[6] The three graphs at the top of Fig. 4 illustrate these results.

[6] The annual convergence rate is given by $-log(\rho)/T$ where ρ is the estimated coefficient of initial democracy in an absolute convergence regression, and T is the length of the period over which the difference in democracy is calculated. There are respectively 34, 39 and 59 countries involved in the latter computation. In the balanced panel, the results are almost identical: a convergence process has taken place in the 1910–1960 and 1960–2000 periods at an annual rate of respectively 3.2 % and 3.6 %. In the first period, convergence occurred but only among a club of advanced democracies.

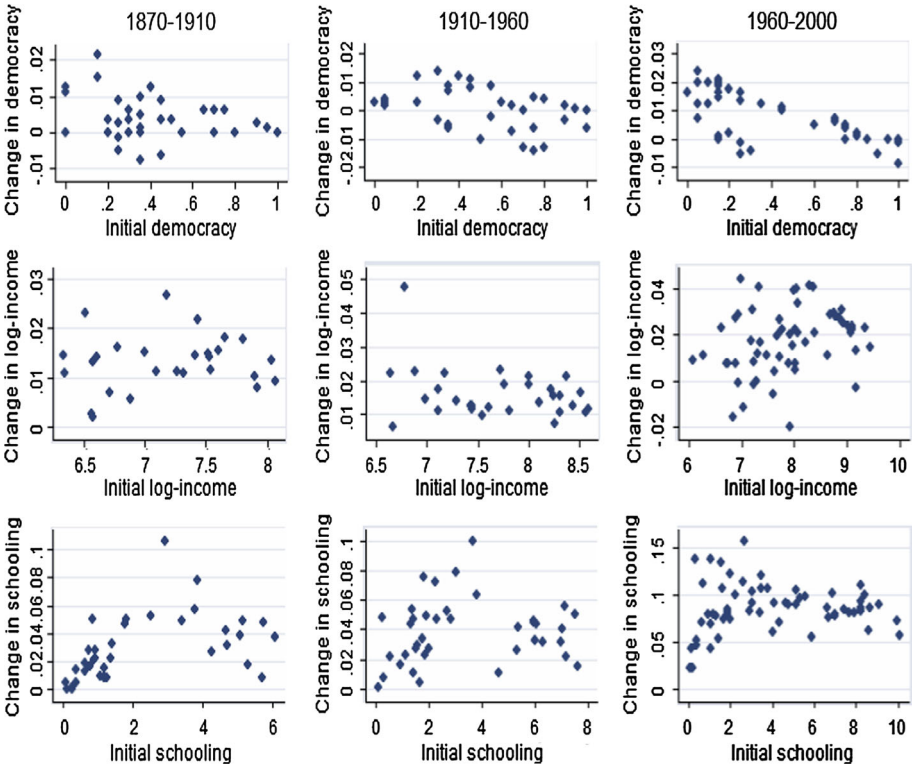

Fig. 4 Convergence in democracy, GDP per capita and average years of schooling by period

How can (beta) convergence in democracy levels be reconciled with the increase in cross-country democracy inequality observed between 1930 and the mid-1980s (sigma divergence)? Beta convergence entails inequality reduction if shocks affecting democracy are stationary over time. But as Figs. 2 and 3 as well as Table 1 make clear, the occurrence of two world wars and the associated political turmoil have significantly widened the distribution of democracy. Hence, the evolution of democracy scores is driven by a convergence effect, but heteroskedastic time-specific shocks have considerably widened the distribution.

2.2 Income and schooling

We capture economic development and modernization either by the level of GDP per capita or by average years of schooling among the adult population, these variables being highly correlated among themselves both across countries and over time. GDP per capita comes from Maddison (2006) while average years of schooling is from Morrisson and Murtin (2009). The latter have constructed a historical database on average years of schooling covering 74 countries since 1870. They combined data on total enrollments in primary, secondary and tertiary schooling with age pyramids, in order to calculate enrollment rates and the average number of years of schooling among the adult population.[7] Morrisson and Murtin (2013) cross-checked these series with historical data on literacy, finding a high level of consistency

[7] Before 1960, the main source for education data is Mitchell (2003)'s statistical yearbooks. After 1960, they relied on the Cohen and Soto (2007) series, adjusted for differential mortality across educational groups.

even at the beginning of the period. In other words, these are the most comprehensive historical data on educational attainment across countries and time available to researchers.

An extensive literature describes the global evolution of key aggregate socio-economic variables over the nineteenth and twentieth centuries, and it is beyond the scope of this paper to comment extensively on all these well-known transformations.[8] Let us simply recall a few facts, mirrored by the summary statistics provided by Table 1: In broad terms, modern economic growth took off in Western Europe in the first half of the nineteenth century, and to varying degrees spread to Asia and Latin America at the beginning of the twentieth century as well as, to perhaps a lesser degree, to Africa after Second World War. In terms of education, schooling was a quasi-exclusive feature of the Western world in 1870, but Eastern Europe as well as some fast-developing Asian countries, most notably Japan, have caught up over the twentieth century. A polarized schooling distribution in 1870 turned to a substantially heterogeneous distribution in 2000: At the top, high-income countries reaching mass enrollment in primary and secondary schooling with growing enrollment in tertiary education. In the middle of the world distribution, Latin America, the Middle-East and North-Africa as well as most of developing Asian countries achieving mass schooling enrollment in primary and partly in secondary school. At the bottom, South Asia and Sub-Saharan Africa have barely achieved extensive levels of enrollment in primary schooling today.

In terms of differences across countries, economic development does not reflect the convergence in democracy described above. As is well-known, income inequality across countries has increased throughout the twentieth century, except among a convergence club of relatively high-income countries before the First World War and after the Second World War. Similarly, there has not been any convergence in average education in any period, except among two clubs of advanced countries during the first and second globalization periods (see Morrisson and Murtin 2009). These two facts are clearly illustrated by the second and third panels of Fig. 4. We now examine the correlation between education, income and democracy in a more quantitative fashion.

2.3 The joint distribution of democracy and development

Figure 5 presents a kernel estimation of the bivariate distribution of democracy and average years of schooling within four different sub-periods: 1870–1970, 1920–1940, 1950–2000 and 1980–2000.[9] The sequence of graphs illustrates joint movements in democracy and schooling over time. A similar graph can be obtained with log GDP per capita instead of education, and results are qualitatively identical. Some interesting facts emerge: On the eve of the twentieth century, the bivariate distribution of democracy and schooling was clearly bimodal, splitting rich and educated democracies on the one hand from poor and low education autocracies on the other. Over the next sub-period, a large share of initial autocracies witnessed economic growth and joined the club of democracies, ensuring a strong positive correlation between democracy and income or education. After 1950, and even more visibly after 1980, there have been two distinct groups of unequal size, poor and

[8] A short list of references might include Bourguignon and Morrisson (2002), describing the world income distribution since 1820; Galor and Weil (2000) and Galor (2005, 2011), analyzing the joint variations of income and population over the long run as well as the structural forces that have triggered the Industrial Revolution; O'Rourke and Williamson (1999) and Hatton and Williamson (2005), focusing on the effect of globalization on economic performance; Morrisson and Murtin (2009, 2013), describing the spread of education at a global level; Murtin (2013) investigating the determinants of the demographic transition over the twentieth century.

[9] We have used an Epanchenikov kernel with bandwidth adjusted to the finite sample size. We used the balanced panel of countries, but results are qualitatively the same with the unbalanced panel.

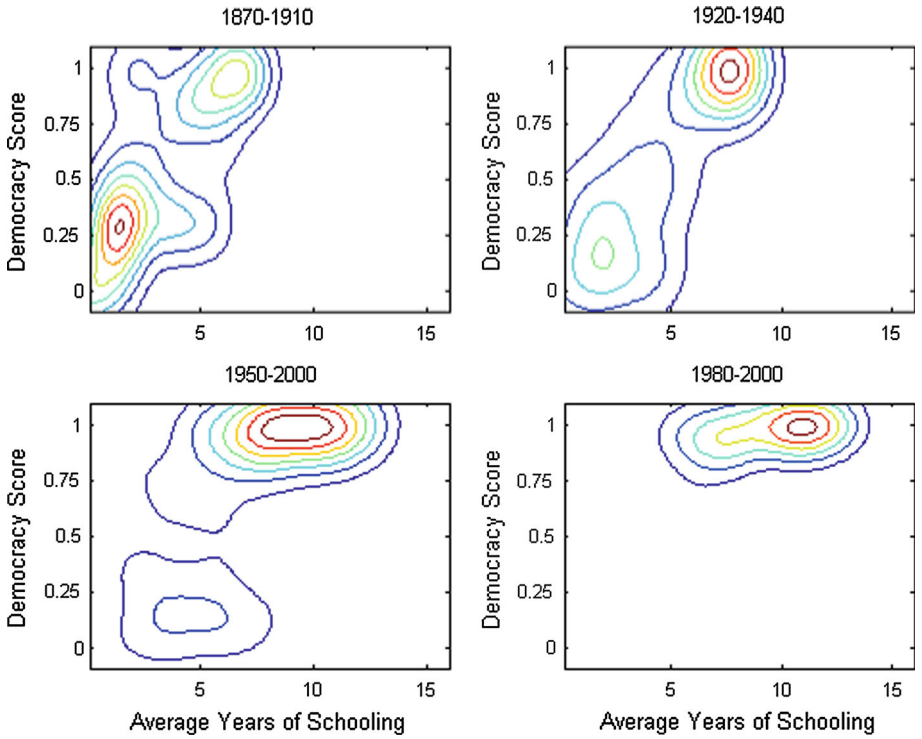

Fig. 5 The joint distribution of democracy and schooling by period—Kernel estimation

low-educated autocracies on the one hand and rich and highly-educated democracies on the other. Among the latter, the correlation between democracy and education or income has flattened out as countries had already reached high levels of democracy in 1950 while experiencing continuous economic development. In sum, this evidence is consistent with the idea that both development and democracy levels experienced a transition from a regime of low education, low income and low democracy to a regime of high education, high income, and high democracy. In what follows we seek to disentangle the directions of causality accounting for this transition.

3 Econometric approach

3.1 Specification

We test the modernization hypothesis by regressing the level of democracy on a variable proxying the level of economic development such as income of education. In addition, we introduce both country fixed effects and the lagged level of democracy. Controlling for country-level fixed effects allows us to partial out the effect of country specific, time invariant factors. Including the lagged value of the dependent variable in the regression reflects the fact that the political structure of a country changes slowly over time, as the introduction or the modification of laws, constitutions and the changes in political regimes meet resistance (Barro 1999). Consequently, the democracy score is highly time persistent, and we account

for autocorrelation in the dependent variable using a dynamic panel data model:

$$D_{it} = a_i + b_t + \rho D_{i,t-1} + cX_{i,t-1} + u_{it} \quad (1)$$

where $D_{i,t}$ is an index of democracy in country i in period t, ranging from 0 to 1, a_i and b_t are respectively country-specific and time-specific effects, u_{it} an idiosyncratic shock and X a variable that proxies for economic development such as log GDP per capita or average education among the adult population, lagged one period. As a benchmark, we consider a 10-year time span between subsequent observations (i.e. in our panel a period is defined as 10 years).

3.2 Estimation by fixed effects and GMM

In our baseline analysis, we use three alternative methods to estimate (1), methods that have become widepread in the empirical literature estimating dynamic panel data models. The first estimator is dynamic fixed effects, that is, fixed effects applied directly to the dynamic model (henceforth denoted DFE). It is well-known that in a classical dynamic panel data model, this within estimator delivers biased estimates of Eq. (1) due to Nickell bias (Nickell 1981). These estimates should therefore be taken with caution. However, DFE can be used to detect estimation problems in more sophisticated dynamic panel estimators, to which we now turn.[10]

The second estimator we employ is Arellano and Bond's (1991) GMM estimator (henceforth denoted AB). This estimator differences away time-invariant, country specific effects and relies on the dynamic structure of the model for identification by using lagged levels of the independent variables as instruments for current differences. The literature on dynamic panel data methods has identified the problem of weak instruments as a serious issue with the AB estimator. A comparison between estimates of ρ under DFE and AB can be used to detect the existence of a weak instruments problem in AB. Since DFE delivers an underestimation of the degree of persistence ρ, the DFE estimate of ρ can be viewed as a lower bound on the true persistence coefficient. Under AB estimation, weak instruments have the well-known property of biasing the estimates towards their DFE counterparts, that is to say to underestimate the degree of autocorrelation in the dependent variable. Therefore, finding similarly low persistence coefficients across DFE and AB regressions can be an indication of a weak instruments problem in the AB regression.[11]

The third estimator is the system GMM estimators of Blundell and Bond (1998) (henceforth BB). Compared to AB, this estimator involves additional moment conditions, which amount to using lagged differences as instruments for current levels. BB has the advantage of addressing the small sample bias issues inherent in AB estimation, and has thus become the current preferred estimator in the literature. However, as we further discuss below, the validity of the additional moment conditions depends on a stationarity assumption, that the country fixed effects are uncorrelated with lagged differences in the dependent variable, i.e. $E(a_i(D_{i,t-s} - D_{i,t-r})) = 0$. This assumption is hard to evaluate with prior theoretical arguments. It is also hard to evaluate empirically, since unbiased estimates of a_i are not available. However, Hauk and Wacziarg (2009) present simulation results applied to cross–country

[10] We also considered a simple FE estimator, without including lagged democracy on the right hand side. These results appeared in the working paper version of this study (Murtin and Wacziarg 2011).

[11] These insights are well-known from the extensive empirical growth literature on estimates of the rate of convergence. See the successive papers of Islam (1995), Caselli, Esquivel and Lefort (1996) and Hauk and Wacziarg (2009).

tests of the Solow growth model, finding that while the stationarity condition in this application is unlikely to hold exactly, the biases on estimated model parameters from BB are systematically smaller in magnitude than those that result from both the weak instruments problem under AB and Nickell bias under DFE. While our application is not exactly the same as cross-country growth regressions, salient features of the empirical exercise are similar, namely our application features highly time persistent variables with a similar dynamic structure in a cross-country context. These simulation results argue is favor of the BB estimator over both AB and DFE. However, throughout this paper we report estimates based on all three estimators.[12]

Another problem with GMM estimators, particularly BB, is that their validity is subject to the use of a relatively small number of instruments. As described by Anderson and Sorenson (1996), Bowsher (2002) and Roodman (2009a), instruments proliferation generate implausibly low values of Hansen tests of instruments exogeneity. This is because the size of the variance matrix of the moments is too large to be estimated accurately within a finite sample. Reducing the number of instruments limits this problem. As a rule of thumb, Roodman (2009a) suggests that the number of instruments should be strictly lower than the number of countries in the sample. The BB regressions presented below abide strictly by the following rule: one should avoid selecting too many or too few instruments. Having too few instruments is likely to generate a weak instruments problem and to deliver inaccurate estimates, while using too many instruments hinders the validity of the Hansen test and runs the risk of endogeneity bias. To strike a balance between estimates consistency and tests validity, we target about 50 instruments, a reasonable objective in a sample that includes about 70 countries. In practice, the number of selected instruments is strictly comprised between 36 and 59, and it is on average equal to 53. In the baseline regressions, the number of instruments is never larger than the number of countries in the sample, conforming to Roodman's rule of thumb. Furthermore, the selection of instruments obeys a transparent rule. We systematically select some lags of the democracy variable D and some lags of the modernization variable X as instruments. When the AB estimator is applied, the second and third lags of D and X are used by default, while the fourth and fifth lags are added in the more recent periods (1930–2000 and 1960–2000) in order to obtain broadly the desired number of instruments. In the context of a BB estimation, we select the third lag of democracy by default, and further lags over more recent periods. In addition, we use the second lag of income and the third or fourth lag of education as default instruments. Education requires more distant lags as this variable displays much time persistence. The maintained assumption is that the education level of the adult population observed 30 or 40 years ago is a valid instrument.

As a diagnostic for the quality of our estimates, for each GMM regression we report below the p value of the Hansen test for joint exogeneity of all the instruments. We also test for the joint exogeneity of the *additional* instruments used in BB regressions relative to AB regressions (i.e. the first-differenced variables used as instruments for the level variables), and report the corresponding p values for the corresponding difference-in-Hansen tests (as described in Roodman 2009b). A similar exogeneity test can be conducted for each individual instrumental variable under BB estimation, such as lags of democracy, income, and education. We also report those, where applicable, in the tables displaying our empirical results.

[12] We also use Windmeijer (2005) finite sample correction of standard errors in order to increase robustness. For further details on the AB and BB estimators, and of their econometric properties in the context of cross-country dynamic panel applications, see Hauk and Wacziarg (2009).

3.3 Drawbacks of system-GMM estimation

Despite our precautions in applying the BB estimator, as described above, econometric concerns may still remain about GMM-based estimation methods.[13] These concerns are of various sorts:

1. The first issue is the validity of the stationarity condition, which can be questioned. Indeed, if the evolution of democracy levels is characterized by a process of differential convergence, as Fig. 5 suggests was the case for much of the period we consider, and if the different patterns of convergence are systematically related to the time invariant country effect, then the stationarity assumption for the BB estimator comes into question since the lagged difference in democracy could be correlated with the country fixed effects.[14] This may be a particular problem for the 1960–2000 regressions, where the time horizon may be too short for the influence of initial conditions to have disappeared and for countries to be sufficiently close to their steady-state democracy levels (this is another argument for focusing on long periods such as 1870–2000). In principle a failure of the stationarity condition could impact not only estimates of the coefficient on lagged democracy, but also those on schooling and per capita income.
2. The second issue concerns the number of instruments. While we choose a number of instruments to conform strictly to Roodman (2009a,b) rule, the total number of instruments remains large. As expected, when we perform an extreme reduction in the number of instruments, the validity of the instruments in levels comes into question, as the Hansen test p values typically fall to levels suggesting the instrument set is invalid (this problem is worse in regressions where $X_{i,t-1}$ is measured by income rather than education). Moreover, when performing such a further reduction in the number of instruments to a very low number of instruments, our estimates of Eq. (1) become unstable. Whether this is due to the fact that a reduction in the number of instruments leads to weak instruments or whether it is due to invalid instruments is open to question, but irrespective it is a further consideration that can lead to questioning the validity of GMM approaches.
3. The third issue concerns the choice of the lags for the included instruments. We try to use distant lags (third lag at least) to limit the potential for instrument endogeneity, but alternative choices of the lag structure in the instrument set, for instance including the second lag, can lead to instability in the results. This instability mirrors that which is obtained when reducing or collapsing the set of instruments, and in general demonstrates the difficulty in appropriately choosing the set of instruments under BB estimation—in this specific context as well as many other contexts.
4. Finally, in the estimates we report below, DFE and BB estimates on either education or income can be quite close to each other (DFE and BB estimates on lagged democracy, on the other hand, are quite different from each other). There are two possible interpretations for this finding. The first is that BB instruments are valid and result in the same estimates as DFE estimation, but the severity of Nickell bias for the latter estimates was not severe, so that the two sets happen to coincide (Nickell bias could be a more serious explanation for differential estimates when it comes to lagged democracy, but not seriously affect estimates of the impacts of education or income). Another explanation is that the set of instruments used under BB are invalid, so BB does little to correct the

[13] We are grateful to an anonymous referee for pointing out these concerns to us in terms close to the discussion that follows.

[14] Acemoglu et al. (2008) raise a similar point.

biases inherent in DFE, and BB estimates collapse to DFE. It is hard to adjudicate *a priori* between these interpretations, but this further ambiguity argues for looking at a different estimation method as an alternative to GMM approaches.

3.4 Bias-corrected estimation: the Bun-Carree estimator

Given the aforementioned limitations of GMM methods, we explore an alternative estimation method that was recently developed by Bun and Carree (2005) for dynamic panels with a small T and a large N. This method is based on calculating the exact DFE bias term, and proposing a procedure to correct DFE estimates by removing this bias. Hence, the estimator is known as "bias-corrected dynamic fixed effects" (henceforth BC). This estimator has several attractive features. First, there is no need to choose instruments or moment conditions, as this is not a GMM/IV based estimation strategy. Second, the estimator has good finite sample properties in Monte Carlo simulations: Bun and Carree (2005) show in such simulations that it removes virtually all of the DFE bias. Third, the estimator does not rely on an initially consistent estimator to derive the final estimates, contrary to bias-correction procedures that had been proposed in the past. Limitations of the initial version of this estimator is that it was developed under the assumption that the additional regressors included in the regression specification are strictly exogenous (which is not the case in our application) and under the assumption that the panel is balanced. We relax the latter assumption and test for the performance of our extended estimator under the assumption that the additional regressors are *not* strictly exogenous by performing Monte Carlo simulations similar to those in Bun and Carree (2005), as well as specific simulations appropriate for our empirical application. A remaining drawback is that implementing the BC estimator, at least in the way we did it, is computationally demanding.[15]

3.4.1 The balanced panel case

We start by describing the BC estimator in the balanced panel case with exogenous regressors. The data is composed of N countries with T observations each. As before we consider the following model:

$$\mathbf{y}_i = \gamma \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{i}_T \eta_i + \boldsymbol{\varepsilon}_i \tag{2}$$

where $\mathbf{y}_i = (y_{i,1}, ..., y_{i,T})'$, $\mathbf{y}_{i,-1} = (y_{i,0}, ..., y_{i,T-1})'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 ... \beta_K)'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, ... \varepsilon_{i,T})'$, $\mathbf{i}_T = (1, ..., 1)'$ a $T \times 1$ vector of 1's and $\mathbf{X}_i$ a matrix with (t, k) element $x_{i,t,k}$. We start by assuming that the regressors are possibly correlated with individual-specific effects η_i but that they are strictly exogenous with respect to the error term $\boldsymbol{\varepsilon}_i$. Stacking over individuals yields:

$$\mathbf{y} = \mathbf{W}\boldsymbol{\delta} + (\mathbf{I}_N \otimes \mathbf{i}_T)\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{3}$$

with $\mathbf{y} = (y_1', ..., y_N')'$, $\mathbf{W} = [\mathbf{y}_{-1} \vdots \mathbf{X}]$, $\mathbf{y}_{-1} = (y_{1,-1}', ..., y_{N,-1}')'$, $\mathbf{X} = (\mathbf{X}_1' ... \mathbf{X}_N')'$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1 ... \boldsymbol{\eta}_N)'$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1' ... \boldsymbol{\varepsilon}_N')'$ and $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$. The DFE (least-squares, dummy variables) estimator is given by:

$$\widehat{\boldsymbol{\delta}}_{DFE} = (\mathbf{W}'\mathbf{A}\mathbf{W})^{-1}\mathbf{W}'\mathbf{A}\mathbf{y} \tag{4}$$

where the $NT \times NT$ idempotent matrix $A = \mathbf{I}_N \otimes (\mathbf{I}_N - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T')$ eliminates individual effects.

[15] We programmed all the estimation and simulation routines ourselves. With a regular personal computer, it takes about 40-50 hours to run a single regression. Thus, in our application we only replicate our baseline results using BC, and continue to report system GMM results for other specifications.

The bias-corrected estimator proposed by Bun and Carree (2005) is based on an iterative approach that calculates the DFE bias and recovers the "true" value of parameters of interest $\boldsymbol{\delta}$. Their approach is iterative as the DFE bias depends on the variance of residuals, which needs to be estimated and hence depends on estimates $\widehat{\boldsymbol{\delta}}$. Bun and Carree (2005) derive the asymptotic limiting distribution of this bias-corrected estimator for a finite T. More specifically, BC estimates are the solution to the following system:

$$\begin{aligned}
\widehat{\boldsymbol{\delta}}_{BC} &= \mathbf{g}^{-1}\left(\widehat{\boldsymbol{\delta}}_{DFE}\right) \\
\mathbf{g}(\boldsymbol{\delta}) &= \begin{pmatrix} \gamma - \widetilde{\sigma}_{\varepsilon}^2 h(\gamma, T)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \\ \boldsymbol{\beta}\widetilde{\sigma}_{\varepsilon}^2 \boldsymbol{\varsigma} h(\gamma, T)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \end{pmatrix} \\
h(\gamma, T) &= \frac{(T-1) - T\gamma + \gamma^T}{T(T-1)(1-\gamma)^2} \\
\widetilde{\sigma}_{\varepsilon}^2 &= \frac{(\mathbf{y} - \gamma\mathbf{y}_{-1} - \mathbf{X}\boldsymbol{\beta})' A (\mathbf{y} - \gamma\mathbf{y}_{-1} - \mathbf{X}\boldsymbol{\beta})}{N(T-1)} \\
\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} &= \left(1 - \rho^2_{\mathbf{X}\mathbf{y}_{-1}}\right)\sigma^2_{\mathbf{y}_{-1}} \\
\rho^2_{Xy_{-1}} &= \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}_{-1}} \boldsymbol{\Sigma}^{-1}_{\mathbf{x}\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}} / \sigma^2_{\mathbf{y}_{-1}} \\
\boldsymbol{\varsigma} &= \boldsymbol{\Sigma}^{-1}_{\mathbf{x}\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}}
\end{aligned}$$

where $\boldsymbol{\Sigma}_{\mathbf{xx}} = \frac{1}{N(T-1)}\mathbf{X}'\mathbf{A}\mathbf{X}$, $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}} = \frac{1}{N(T-1)}\mathbf{X}'\mathbf{A}\mathbf{y}_{-1}$, $\sigma^2_{\mathbf{y}_{-1}} = \frac{1}{N(T-1)}\mathbf{y}'_{-1}\mathbf{A}\mathbf{y}_{-1}$. From the expression of function $g(.)$ it is clear that the DFE estimator is inconsistent and that $\widehat{\boldsymbol{\gamma}}_{DFE}$ is downward biased. The severity of the bias depends on $h(\gamma, T)$ and hence is larger when the value of γ is close to 1 or when the number of time periods T is low. As argued below, it is debatable whether the two latter situations apply to our democracy and development data when using a decennial time span over the 1870–2000 period.

3.4.2 The case of unbalanced panel

We now turn to extending this method to the unbalanced panel case. For unbalanced panels, Bun and Carree (2005) propose in the concluding remarks to their paper to split the unbalanced panel into at most $T-1$ balanced panels of length $(2, ..., T)$, an approach also adopted by Lokshin (2009). [16] Let n_p be the number of countries composing the balanced panel of size $n_p \times p$, with $2 \leq p \leq T$, and $(\mathbf{y}^p, \mathbf{y}^p_{-1}, \mathbf{X}^p)$ be the corresponding data. Following Bun and Carree (2005) and Lokshin (2009) we denote $\varphi(p)$ the fraction of observations in each balanced sub-panel:

$$\varphi(p) = \frac{(p-1)n_p}{\sum_{p=2}^{T}(p-1)n_p} \tag{5}$$

Following Bruno (2005), let us define the indicator $s_{i,t}$ taking value 1 if both dependent and explanatory variables (including the lagged dependent variable) are observed. For each i define the $(T \times 1)$ vector $s_i = [s_{i,1}, ..., s_{i,T}]'$, the $(T \times T)$ diagonal matrix S_i having the vector s_i on its diagonal, and the $(NT \times NT)$ block-diagonal matrix $S = diag(S_i)$. The equivalent of the BC estimator for unbalanced panels is given by the following system:

$$\widehat{\boldsymbol{\delta}}_{BC} = \mathbf{g}^{-1}\left(\widehat{\boldsymbol{\delta}}_{DFE}\right)$$

[16] Lokshin (2009) is the only other empirical application of the Bun-Carree estimator that we are aware of.

$$
\begin{aligned}
\mathbf{g}(\delta) &= \begin{pmatrix} \gamma - \widetilde{\sigma}_{\varepsilon}^2 h(\gamma, T)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \\ \boldsymbol{\beta} + \widetilde{\sigma}_{\varepsilon}^2 \boldsymbol{\varsigma} h(\gamma, T)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \end{pmatrix} \\
h(\gamma, T) &= \sum_{p=2}^{T} \varphi(p) \frac{(p-1) - p\gamma + \gamma^p}{p(p-1)(1-\gamma)^2} \\
\widetilde{\sigma}_{\varepsilon}^2 &= \sum_{p=2}^{T} \varphi(p) \widetilde{\sigma}_{\varepsilon}^2(p) \\
\widetilde{\sigma}_{\varepsilon}^2(p) &= \frac{(\mathbf{y}^p - \gamma \mathbf{y}^p_{-1} - \mathbf{X}^p \boldsymbol{\beta})' A_p (\mathbf{y}^p - \gamma \mathbf{y}^p_{-1} - \mathbf{X}^p \boldsymbol{\beta})}{n_p(p-1)} \\
\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} &= \left(1 - \rho^2_{\mathbf{X}\mathbf{y}_{-1}}\right) \sigma^2_{\mathbf{y}_{-1}} \\
\rho^2_{Xy_{-1}} &= \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}_{-1}} \boldsymbol{\Sigma}^{-1}_{\mathbf{x}\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}} / \sigma^2_{\mathbf{y}_{-1}} \\
\boldsymbol{\varsigma} &= \boldsymbol{\Sigma}^{-1}_{\mathbf{x}\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}}
\end{aligned}
$$

with $\boldsymbol{\Sigma}_{\mathbf{xx}} = \sum_{p=2}^{T} \varphi(p) \frac{1}{n(p)(p-1)} \mathbf{X}^{p\prime} \mathbf{A}_p \mathbf{X}^p$, $\boldsymbol{\Sigma}_{\mathbf{xy}_{-1}} = \sum_{p=2}^{T} \varphi(p) \frac{1}{n(p)(p-1)} \mathbf{X}^{p\prime} \mathbf{A}_p \mathbf{y}^p_{-1}$, $\sigma^2_{\mathbf{y}_{-1}} = \sum_{p=2}^{T} \varphi(p) \frac{1}{n(p)(p-1)} \mathbf{y}^{p\prime}_{-1} \mathbf{A}_p \mathbf{y}^p_{-1}$, $A_p = S(I_{n_p T} - D(D'SD)^{-1} D')S$, $I_{n_p T}$ the identity matrix of size $n(p)T$, $D = \mathbf{I}_N \otimes i_p$, i_p the $(p \times 1)$ vector of 1's and $S = diag(S_i)$. This provides a straightforward extension of the BC estimator to the case that concerns us, namely an unbalanced panel.

3.4.3 Monte Carlo evidence on the properties of the Bun-Carree estimator

We now examine the performance of the bias-corrected LSDV estimator proposed by Bun and Carree (2005) in the case of unbalanced panel data, using Monte Carlo simulations. We proceed in two steps. First, we replicate exactly the procedure that Bun and Carree applied to balanced panel, but this time applying it to unbalanced panels. Second, we run a simulation using a data generating process (DGP) that more closely mirrors the one in our application to democracy and development. The latter simulation features an extension to account for several specific characteristics of the DGP in our application: weakly exogenous regressors, censored dependent variable, time fixed effects and the specific moments of our data.

As a first step, we replicate the Monte-Carlo experiment described in Bun and Carree (2005) for different degrees of unbalancedness and different sample sizes. Specifically, we generate (N, T) panel data for (y, x) from latent variables (y^*, x^*) according to the following data generating process (DGP):

$$
\begin{aligned}
y^*_{i,t} &= \gamma y^*_{i,t-1} + \beta x^*_{i,t} + \eta_i + \varepsilon_{i,t} \\
x^*_{i,t} &= \rho x^*_{i,t-1} + \xi_{i,t} \\
\eta_i &\sim N\left(0, \sigma^2_{\eta}\right) \\
\varepsilon_{i,t} &\sim N\left(0, \sigma^2_{\varepsilon}\right) \\
\xi_{i,t} &\sim N\left(0, \sigma^2_{\xi}\right) \\
(y_{i,t}, x_{i,t}) &= \left(s_{i,t} y^*_{i,t}, s_{i,t} x^*_{i,t}\right) \\
s_{i,t} &= I\left(\sum_{k=1}^{t} I\left(u_{i,k} > \tau\right) \geq 1\right), \; u_{i,k} \sim U[0,1]
\end{aligned}
$$

Table 2 Monte Carlo test of Bun-carree estimator with unbalanced panel data

(N, T^*)	(150,4)	(100,6)	(60,10)	(40,15)
Balanced panel				
T^*/T	1	1	1	1
Bias $\hat{\gamma} - \gamma$				
DFE	−0.194	−0.071	−0.036	−0.022
BC	0.001	0.001	0.001	0
Bias $\hat{\beta} - \beta$				
DFE	−0.015	0.012	0.016	0.014
BC	−0.001	0	0.001	−0.001
Unbalanced panel				
T^*/T	0.55	0.6	0.67	0.75
Bias $\hat{\gamma} - \gamma$		–		
DFE	−0.273	−0.112	−0.053	−0.03
BC	0.002	0.001	0	0
Bias $\hat{\beta} - \beta$				
DFE	−0.053	0.001	0.015	0.016
BC	0.001	0.001	0.001	0

where τ is a parameter controlling for the timing of inclusion in the sample.[17] Unbalancedness is measured by average group size $T^* = \sum_{i=1}^{N} T_i/N$ where $T_i = \sum_{t=1}^{T} s_{i,t}$ is the number of usable observations for country i. As in Bun and Carree (2005), the simulation design is such that $\beta = 1$, $\gamma = \rho = 0.8$, $\sigma_\eta = \sigma_\varepsilon = \sigma_\xi = 1$, $N \times T^* = 600$. We consider the following four (maximal) sample sizes $(N, T^*) \in \{(150, 4), (100, 6), (60, 10), (40, 15)\}$.[18] Homoskedasticity is assumed. For each experiment, we perform 1,000 Monte-Carlo replications. Table 2 reports both the (biased) DFE estimate and the bias-corrected BC estimates in the cases of a balanced panel (upper part) and an unbalanced one (lower case). As expected, the size of the LSDV bias diminishes as the time dimension T^* of the panel increases. Moreover, the BC bias terms are very small irrespective of the degree of unbalancedness, so the BC estimator extended to unbalanced panels seems to work well.

In a second step, we test the bias-corrected estimator on simulated data that display the same statistical properties as our observed data. In particular, we introduce time dummies and potential censoring of the dependent and explanatory variables. Censoring takes two forms: first, latent variables (y^*, x^*) can be included in the sample or not; second, included variables (y, x) can be top-coded or not. The thresholds for top-coded variables (y^c, x^c) are 1 for democracy y and 6 years for average years of primary schooling x. We allow for non-zero correlation between democracy and education-related country-level fixed-effects. Moreover, simulations allow us to examine the influence of two potential sources of bias, namely heteroskedasticity and the weak exogeneity of regressors. As in Bun and Carree (2005) the variance of democracy innovations is assumed to vary over time. We also allow for non-zero correlation between lagged democracy innovations $\varepsilon_{i,t-L}$ and education innovation $\xi_{i,t}$, making education a weakly exogenous variable. The latter source of bias is crucial for us

[17] Inclusion into the sample is assumed to be an absorbing state, namely countries do not exit the sample once they have entered it. Inclusion takes place at the first date t_0 such that $u_{i,t_0} > \tau$.

[18] We do not consider very short panels with $T = 2$ or $T = 3$ as Lokshin (2009) exhibits non-convergence patterns for such very short panels.

to consider, as we are unwilling to assume that the regressors are strictly exogenous. Hence the DGP allows for multiple correlations between regressors and country fixed-effects or residuals, as follows:

$$
\begin{aligned}
y^*_{i,t} &= \gamma y^*_{i,t-1} + \beta x^c_{i,t} + \eta^y_i + \delta^y_t + \varepsilon_{i,t} \\
x^*_{i,t} &= \rho x^*_{i,t-1} + \eta^x_i + \delta^x_t + \xi_{i,t} \\
x^c_{i,t} &= x_{i,t}.I\left(x_{i,t} \leq 6\right) + 6.I\left(x_{i,t} > 6\right) \\
\delta^y_t &= \mu^y + g^y t \text{ and } \delta^x_t = \mu^x + g^x t \\
\left(\eta^y_i, \eta^x_i\right) &\sim N\left(0, \begin{bmatrix} \sigma^2_{\eta^y} & \rho^\eta \sigma_{\eta^y} \sigma_{\eta^x} \\ \rho^\eta \sigma_{\eta^y} \sigma_{\eta^x} & \sigma^2_{\eta^x} \end{bmatrix}\right) \\
\varepsilon_{i,t} &\sim N\left(0, \sigma^2_\varepsilon(t)\right) \\
\sigma^2_\varepsilon(t) &= \sigma^2_\varepsilon(0) + g_\varepsilon.(t - T/2) \\
\xi_{i,t} &= \frac{\lambda \sigma_\xi}{\sigma_\varepsilon(t-L)} \varepsilon_{i,t-L} + u_{i,t}, u_{i,t} \sim N\left(0, \left(1-\lambda^2\right)\sigma^2_\xi\right) \\
\left(y_{i,t}, x_{i,t}\right) &= \left(s_{i,t} y^*_{i,t}, s_{i,t} x^*_{i,t}\right) \\
s_{i,t} &= I\left(\sum_{k=1}^{t} I\left(u_{i,k} > \tau\right) \geq 1\right), \quad u_{i,k} \sim U[0,1] \\
y^c_{i,t} &= y_{i,t} I\left(y_{i,t} \leq 1\right) + I\left(y_{i,t} > 1\right)
\end{aligned}
$$

The structural parameters of the above DGP and the initial distributions of $(y^*_{i,0}, x^*_{i,0})$ are estimated using the Simulated Method of Moments, while fixing $\gamma = 0.5$, $\beta = 0.05$, $\rho = 0.9$, $\rho^\eta = 0.5$, $\tau = 0.85$, $g_\varepsilon = 0.006$, $\lambda = 0.5$ and $L = 2$. Given initial conditions on parameters $(\mu^y, \mu^x, g^y, g^x, \sigma^2_{\eta^y}, \sigma^2_{\eta^x}, \sigma^2_\varepsilon(0), \sigma^2_\xi)$, we minimize the sum of squared differences between some observed statistical moments and the corresponding simulated moments obtained from 1,000 Monte Carlo draws. Each moment is normalized by the inverse variance of the simulated vector. The set of statistical moments includes: i) the means of democracy and education taken both in levels and in first differences; ii) the covariance matrices of observed variables in levels and first-differences (i.e. variances and correlations); iii) the autocorrelation of the observed variables; iv) the shares of censored observations for the democracy variable. Table 3 Panel A reports the results. [19] Simulated and observed moments are almost perfectly in line, so that the estimated DGP constitutes a credible replication of our democracy and education data set as well as a convenient benchmark to assess the influence of some specific features of the data such as top-coding or the exogeneity of innovations.

Panel B of Table 3 reports the results of the BC estimation conducted on various sets of simulated data. Starting from a DGP with no censoring, strict exogeneity of the education covariate and homoskedastic innovations of democracy, we gradually introduce censoring and top-coding, weak exogeneity of education and heteroskedasticity, as in last column where the latter features are jointly considered. In all cases, the BC bias on the autoregressive parameter γ is much smaller than its DFE counterpart. For both LSDV and BC estimates, the bias on the slope parameter on education β can be relatively large in absolute value, but it is always much smaller with the BC estimator. Moreover, with censoring and top-coding, the BC and LSDV estimates underestimate the true value of the slope parameter β as shown in the last column. This result was already noted by Benhabib et al. (2011), and suggests viewing the

[19] The estimation yields $(\mu^y = 0.097, \mu^x = 1.7e-4, g^y = 0.014, g^x = 0.055, \sigma^2_{\eta^y} = 0.021, \sigma^2_{\eta^x} = 0.091, \sigma^2_\varepsilon(0) = 0.263, \sigma^2_\xi = 0.295)$.

Table 3 Monte Carlo test of the Bun-Carree estimator

Panel A: Statistical fit of the data generating process		
	Observed	Simulated
Share of selected observations	0.62	0.63
Mean democracy	0.572	0.594
Mean change in democracy	0.034	0.036
Variance of democracy	0.126	0.119
Variance of change in democracy	0.06	0.058
Share of top-coded democracy observations	0.25	0.24
Mean primary education	2.937	2.337
Mean change in primary education	0.306	0.302
Variance of education	3.648	3.195
Variance of change in education	0.081	0.104
Correlation of education and its lag	0.98	0.98
Correlation between democracy and education	0.635	0.586
Correlation between changes in democracy and education	0.112	0.06

Panel B: Test of the Bun-Carree estimator					
	No censoring and no top-coding			With censoring and top-coding	
	Strict exogeneity		Weak exogeneity	Strict exogeneity	Weak exogeneity
	Homoskedastic	Heteroskedastic	Heteroskedastic	Heteroskedastic	Heteroskedastic
Bias $\hat{\gamma} - \gamma$					
DFE	−0.191	−0.196	−0.181	−0.22	−0.227
BC	0.023	−0.025	0.02	−0.01	−0.047
Bias $\hat{\beta} - \beta$					
DFE	−0.027	−0.025	0.05	−0.07	−0.071
BC	0.015	0.013	0.02	−0.013	−0.048

BC estimates presented below as lower bounds on the true effects of education and income on democracy.

4 Testing the modernization hypothesis over the long run

4.1 Democracy and income

We start with DFE and GMM estimates. Table 4 focuses on log GDP per capita as the main explanatory variable and displays the estimates of Eq. (1) over the whole period as well as over three sub-periods, 1960–2000, 1930–2000 and 1870–2000, enabling us to assess the influence of time-specific sample selection. Regarding the choice of countries, we first use the whole sample, then a sample that excludes countries already at the maximum level of democracy as of the initial year. Thus, Table 4 examines the relationships between democracy and development using a wide range of econometric procedures, time-periods, and country samples. Column 1 of Table 5 displays the corresponding estimate using the BC estimator for the whole sample of countries over the 1870–2000 period.

Table 4 Democracy and GDP per Capita 1870–2000

	1870–2000			1900–2000			1930–2000			1960–2000		
	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB
Dependent variable is standardized Polity IV index of democracy												
Full sample												
D(−1)	0.332***	0.266**	0.509***	0.265***	0.253**	0.493***	0.155***	0.193*	0.419***	0.120*	0.264**	0.612***
	(0.04)	(0.11)	(0.10)	(0.05)	(0.12)	(0.11)	(0.06)	(0.10)	(0.15)	(0.07)	(0.10)	(0.17)
Log y(−1)	0.105**	−0.091	0.118***	0.128***	−0.061	0.120***	0.135**	0.137	0.138***	0.084	0.106	0.129***
	(0.04)	(0.17)	(0.03)	(0.04)	(0.19)	(0.03)	(0.05)	(0.14)	(0.04)	(0.06)	(0.16)	(0.05)
N	560	483	560	505	456	505	417	372	417	307	275	307
N countries	69	69	69	69	69	69	69	69	69	69	68	69
N instruments		58	59		55	56		56	49		45	46
AB1 *p* value		0	0		0	0		0	0		0	0
AB2 *p* value		0.34	0.24		0.36	0.25		0.29	0.19		0.71	0.48
Hansen *p* value		0.08	0.14		0.07	0.17		0.18	0.13		0.04	0.10
Difference-in-Hansen *p* value			0.80			0.91			0.61			0.46
Nature of instruments and		D(−2/ − 3)	D(−3)/0.86		D(−2/ − 3)	D(−3)/0.98		D(−2/ − 4)	D(−3/ − 4) /0.93		D(−2/ − 5)	D(−3/ − 5) /0.57
Difference-in-Hansen *p* value		ly(−2/ − 3)	ly(−2)/0.19		ly(−2/ − 3)	ly(−2)/0.24		ly(−2/ − 4)	ly(−2)/0.60		ly(−2/ − 5)	ly(−2/ − 4) /0.42

Table 4 continued

	1870–2000			1900–2000			1930–2000			1960-2000		
	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB
Excluding initial democracies												
D(−1)	0.322***	0.252**	0.509***	0.265***	0.259**	0.438***	0.153**	0.172	0.301**	0.035	0.150*	0.345***
	(0.05)	(0.11)	(0.11)	(0.05)	(0.11)	(0.10)	(0.06)	(0.11)	(0.13)	(0.08)	(0.09)	(0.13)
Log y(−1)	0.108**	−0.088	0.124***	0.139***	0.005	0.139***	0.185***	0.403**	0.159***	0.257***	0.413***	0.202***
	(0.04)	(0.18)	(0.03)	(0.05)	(0.23)	(0.03)	(0.06)	(0.18)	(0.04)	(0.08)	(0.15)	(0.03)
N	534	459	534	444	397	444	306	267	306	215	190	215
N countries	67	67	67	63	63	63	54	54	54	50	49	50
N instruments		58	59		55	56		56	49		45	46
AB1 *p* value		0	0		0	0		0.01	0		0	0
AB2 *p* value		0.35	0.23		0.44	0.34		0.47	0.41		0.80	0.72
Hansen *p* value		0.12	0.17		0.13	0.16		0.65	0.28		0.27	0.48
Difference-in-Hansen *p* value			0.83			0.82			0.92			0.54
Nature of instruments and		D(−2/ − 3)	D(−3)/0.88		D(−2/ − 3)	D(−3)/0.93		D(−2/ − 4)	D(−3/ − 4)/0.98		D(−2/ − 5)	D(−3/ − 5) /0.93
Difference-in-Hansen *p* value		ly(−2/ − 3)	ly(−2)/0.21		ly(−2/ − 3)	ly(−2)/0.17		ly(−2/ − 4)	ly(−2)/0.66		ly(−2/ − 5)	ly(−2/ − 4) /0.57

Robust standard errors and Windjmeier correction for standard errors in BB. All regressions include country and time specific effects. Hansen test null hypothesis is the joint exogeneity of all instruments. Difference-in-Hansen test null hypothesis is the joint exogeneity of additional instruments used in SYS-GMM. Difference-in-Hansen test is also used separately for each instrumenting variable

Table 5 Bun-Carree estimation of the modernization hypothesis 1870–2000

	I	II	III
Dependent variable is standardized Polity IV index of democracy			
Lagged democracy	0.526***	0.540***	0.522***
	(0.052)	(0.055)	(0.050)
Lagged log income	0.078*	–	0.076*
	(0.043)	–	(0.042)
Lagged primary schooling	–	0.034*	0.036**
	–	(0.019)	(0.019)
Country fixed effects	yes	yes	yes
Time dummies	yes	yes	yes

The estimation uses the full sample. *** (respectively ** and *) denotes significance at the 1 % (5 and 10 %) confidence level

The overall conclusion is as follows: with the BB estimator, lagged income always displays a positive and significant coefficient. In terms of magnitudes, when we use the BB estimate drawn from the full sample over the 1870–2000 period, the coefficient on lagged income is equal to 0.118 while the persistence coefficient is 0.509. This implies that doubling GDP per capita entails closing about 17 % of the gap between the most dictatorial regime and the most democratic one.[20] With a DFE estimator, lagged income is significant in all but one regression. The only one where income fails to be significant corresponds to the full sample over the 1960–2000 period, which is incidentally the baseline regression in Acemoglu et al. (2008). In contrast, AB estimates on lagged income are only significant in 2 of 8 regressions. However, as already noted this estimator is afflicted by problems of weak instruments, the major reason why the BB estimator was developed. When comparing the DFE and AB results, we do find suggestive evidence of weak instruments, as the persistence coefficients ρ are very close across the two estimators. In contrast, the persistence coefficient is significantly higher under BB estimation.

The robust significance of lagged income when using BB confirms with a longer panel the results found in Bobba and Coviello (2007). Moreover, BB estimation passes all of the specification tests when using the full sample, even though a limited number of instruments was chosen in order to avoid instruments proliferation problems. In particular, the Hansen tests p values are comprised between 0.10 and 0.17 with the full sample, and between 0.16 and 0.48 with the sample excluding initial democracies. As already acknowledged, these p values fall when the instrument set is collapsed/reduced to few instruments.

A feature that could explain the non-significance of some estimates in the previous exercise is the top coding of the democracy variable. Countries already at the "democratic frontier" at initial date are unlikely to experience large decreases in democracy (even if this has actually been observed several times in history, notably after the First World War). Conversely, for the same set of countries increases in democracy will not be captured by the index, since it is already at its maximal value.[21] To assess the influence of this group of countries, we

[20] As the magnitude of the latter effect pertaining to *steady-state* democracy levels amounts to $0.118/(1 - 0.509) \times log(2) = 0.166$

[21] In statistical terms, this raises the issue of the measurement of democracy, which is proxied by a bounded variable. Even if some countries have already converged towards the maximum reported level of democracy at the initial date, institutions have kept on evolving, most likely improving, within these countries.

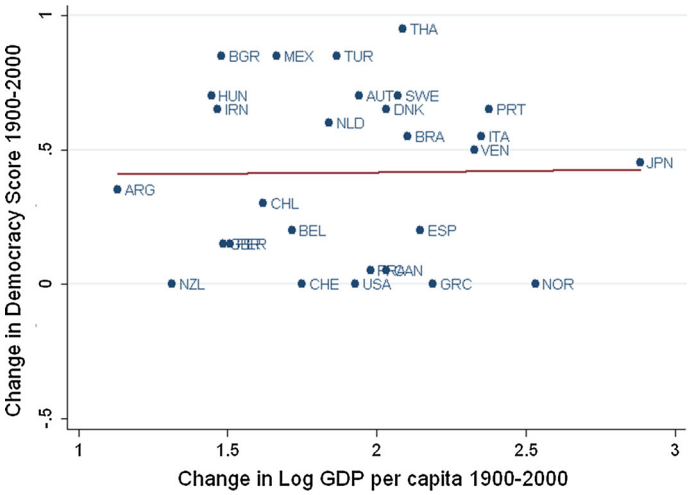

Fig. 6 Change in democracy and economic growth 1900–2000

ran the former regressions while excluding the countries that were already at the maximum level of democracy at the initial date. As shown in the second part of Table 4, excluding these countries from the analysis has an important effect on the results, as the 1960–2000 fixed-effects estimates and the two most recent AB estimates of the coefficient on log per capita income now become significant.

Column 1 of Table 5 confirms our main results using the bias-corrected estimator of Bun and Carree (2005). Here, the coefficient on lagged democracy is in line with that obtained under BB (Table 4, column 3, 1870–2000). The coefficient on log income falls slightly but remains positive and significant. The BC results give some confidence that the BB estimates are valid, since the two sets of estimates are close.

These findings differ from those in Acemoglu et al. (2008) who showed that the relationship between democracy and lagged income was statistically insignificant when calculated over the twentieth century, and turned significant only over the very long term, say between 1500 and 2000. [22] Figure 6 best illustrates their result. It shows the seemingly null correlation between the change in the Polity IV score of democracy and growth in GDP per capita between 1900 and 2000. The non-significant correlation depicted in Fig. 6 can be explained by the omission of the initial level of democracy from the regression. Indeed, the pairwise correlation between growth of GDP per capita and the change in democracy can be spuriously contaminated by unobserved variables. Among those, the initial level of democracy is a potential candidate, since, as described earlier, convergence in democracy has occurred in every period. The negative correlation between the change and the initial level of democracy might therefore contaminate the relationship between the change in democracy and per capita GDP growth.

To further examine this issue, Table 6 presents a simple set of purely cross-sectional regressions where the dependent variable is the change in democracy between 1900 and 2000, while explanatory variables include per capita GDP growth (column I), then adding

Footnote 21 contineued
Benhabib et al. (2011) account for the censoring of the democracy variable and find, similarly to us, a significant coefficient on log GDP per capita.

[22] In their paper, Acemoglu et al. (2008) include both a set of country fixed effects and of time period firxed effects. We include time period fixed effects in all of our panel regressions.

Table 6 OLS results using long-differences between 1900 and 2000—log GDP per capita and primary schooling

	I	II	III	IV	V	VI	VII
Dependent variable: change in score of democracy between 1900 and 2000							
Change in log GDP per capita	0.009	0.066*	0.118***				0.088**
	(0.129)	(0.035)	(0.032)				(0.039)
Initial level of log GDP per capita			0.108***				0.065
			(0.022)				(0.045)
Initial level of the score of democracy		−0.887***	−0.997***		−0.910***	−0.984***	−1.021***
		(0.041)	(0.036)		(0.052)	(0.041)	(0.026)
Change in average years of primary Schooling				0.140***	−0.004	0.086***	0.100***
				(0.038)	(0.018)	(0.016)	(0.025)
Initial level of average years of primary Schooling						0.087***	0.088***
						(0.011)	(0.027)
R^2	0.00	0.95	0.97	0.27	0.9	0.97	0.99
N	28	28	28	35	35	35	28

Robust standard errors. * (respectively ** and ***) stand for significance at the 10 % (5 and 1 %) confidence level

initial democracy (column II), and finally adding initial log GDP per capita (column III). As is clear from these simple cross-country regressions, per capita GDP growth is highly significant once the initial level of democracy is included in the specification, as shown in columns II and III. This suggests that the non-significance of per capita GDP growth in column I results from the omitted variable bias arising from the exclusion of the initial level of democracy. Interestingly, both the initial level and the growth in GDP per capita bear a significant positive association with the change in democracy over the period, suggesting the presence of both accumulation and level effects.[23] We now examine the findings obtained with a different proxy for economic development: education.

4.2 Democracy and education

In this subsection, we analyze the determinants of democracy focusing on education. We consider years of education at various levels (primary, secondary and tertiary), focusing particularly on average years of primary schooling in the adult population (aged 15 years and older).

Columns IV–VII of Table 6 display cross-sectional regressions of the change in democracy between 1900 and 2000 on the corresponding change in mean years of primary schooling among the adult population, initial democracy and initial average years of primary schooling. We find qualitatively the same results as for income. The change in primary schooling is highly significant in column IV and remains significant at the 5 percent confidence level in columns VI and VII once other explanatory variables are added. Controlling for the whole set of income and education-related variables as in column VII, we find positive and significant

[23] The divergence in results with Acemoglu et al. (2008) can also in part be explained by the difference in the time span used across the two analyses. Acemoglu et al. (2008) consider 25 year time spans, while the present study focuses on decennial time spans. When using a longer time span of 30 years, we also found that lagged income was insignificant with a DFE estimator. However, it was still highly significant when using a BB estimator.

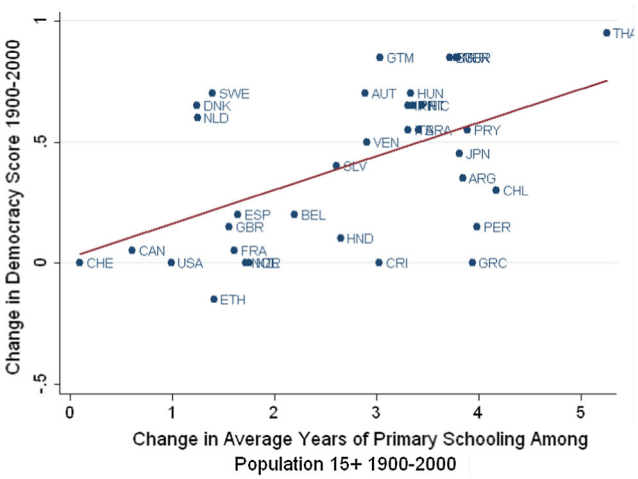

Fig. 7 Change in democracy and average years of primary schooling 1900–2000

coefficients for both the changes in log GDP per capita and average primary schooling, as well as for the initial level of primary schooling. The relationship between democracy and primary schooling is illustrated by Fig. 7, which depicts a significant and positive correlation between changes in democracy and changes in average years of primary schooling between 1900 and 2000.

Returning to higher data frequency (using a panel with decade averages), Table 7 reports results that can be directly compared to those of Table 4. In this setting, we estimate Eq. (1) using various samples and econometric methods, replacing log GDP per capita by average years of primary schooling. Again, we find that in all BB regressions, the coefficient on lagged primary schooling is positive and statistically significant at the 1 % level. As before, for the chosen set of instruments, the BB estimator passes all specification tests. The DFE estimates also generally display a positive effect of primary schooling on democracy in 6 out of 8 regressions. Both DFE and BB estimates contrast with AB estimates, which are never positive and significant at the same time. A comparison between DFE and AB estimates of the persistence coefficient ρ suggests that a weak instruments problem is once again plaguing the AB estimator.

Column 2 of Table 5 confirms the BB results for the 1870–2000 period using the BC estimator instead. Here, the coefficient on primary schooling is 0.034 under BC versus 0.036 under BB estimation, while the coefficient on lagged democracy falls from 0.749 unde BB to 0.540 under BC. The relative similarity between these numbers further reinforces our confidence that there was something to the BB estimates.

In terms of magnitudes, the effect associated with primary schooling is large. When looking at the full sample over the whole period, the BC coefficient on primary schooling (equal to 0.034) and the persistence coefficient (0.54) jointly entail a steady-state effect equal to $0.034/(1-0.54) = 0.074$, and an average increase in $0.074 \times 2.29 = 0.17$ points in the democracy score over the 1870–2000 period where mean years of primary schooling have increased by an average 2.29 years, and average democracy by 0.36 points. Therefore, the diffusion of literacy accounts for about half of the democratic transition over that period.

To see if primary schooling is the main variable driving democracy or if education at other levels of attainment matters separately, we reran the regressions (over the full sample)

Table 7 Democracy and primary schooling 1870–2000

	1870–2000			1900–2000			1930–2000			1960–2000		
	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB
Dependent variable is standardized Polity IV index of democracy												
Full sample												
D(−1)	0.376***	0.423***	0.749***	0.295***	0.392***	0.749***	0.149***	0.224**	0.511***	0.097	0.189*	0.614***
	(0.04)	(0.11)	(0.07)	(0.04)	(0.11)	(0.07)	(0.05)	(0.11)	(0.11)	(0.07)	(0.10)	(0.11)
P(−1)	0.044**	0.036	0.036***	0.062***	0.050	0.036***	0.071***	0.036	0.064***	0.085**	0.013	0.072***
	(0.02)	(0.04)	(0.01)	(0.02)	(0.05)	(0.01)	(0.03)	(0.04)	(0.02)	(0.04)	(0.04)	(0.02)
N	615	534	613	542	498	540	429	389	427	312	278	310
N countries	70	70	70	70	70	70	70	70	70	70	69	70
N instruments		58	57		55	56		56	49		45	46
AB1 *p* value		0	0		0	0		0	0		0	0
AB2 *p* value		0.21	0.15		0.22	0.15		0.22	0.15		0.92	0.56
Hansen *p* value		0.12	0.41		0.12	0.41		0.18	0.31		0.15	0.21
Difference-in-Hansen *p* value			0.77			0.77			0.90			0.53
Nature of instruments and		D(−2/ − 3)	D(−3)/0.85		D(−2/ − 3)	D(−3)/0.85		D(−2/ − 4)	D(−3/ − 4)/0.88		D(−2/ − 5)	D(−3/ − 5) /0.66
Difference-in-Hansen *p* value		P(−2/ − 3)	P(−3)/0.53		P(−2/ − 3)	P(−3)/0.54		P(−2/ − 4)	P(−3)/0.84		P(−2/ − 5)	P(−3/ − 5) /0.73

Table 7 continued

	1870–2000			1900–2000			1930–2000			1960–2000		
	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB	DFE	AB	BB
Excluding initial democracies												
D(−1)	0.373***	0.411***	0.731***	0.307***	0.392***	0.694***	0.183***	0.255**	0.530***	0.051	0.150	0.473***
	(0.04)	(0.10)	(0.07)	(0.05)	(0.12)	(0.11)	(0.06)	(0.12)	(0.14)	(0.08)	(0.11)	(0.08)
P(−1)	0.042**	0.061	0.036***	0.057**	0.075	0.043***	−0.003	−0.005	0.053***	−0.039	−0.154*	0.059***
	(0.02)	(0.05)	(0.01)	(0.02)	(0.05)	(0.01)	(0.04)	(0.06)	(0.02)	(0.06)	(0.08)	(0.02)
N	589	510	587	478	435	476	318	283	316	220	193	218
N countries	68	68	68	64	64	64	55	55	55	51	50	51
N instruments		58	57		55	56		56	49		45	46
AB1 *p* value		0	0		0	0		0	0		0	0
AB2 *p* value		0.2	0.15		0.27	0.21		0.29	0.23		0.99	0.64
Hansen *p* value		0.15	0.47		0.18	0.30		0.67	0.38		0.46	0.44
Difference-in-Hansen *p* value			0.83			0.53			0.96			0.80
Nature of instruments and		D(−2/ − 3)	D(−3)/0.94		D(−2/ − 3)	D(−3)/0.66		D(−2/ − 4)	D(−3/ − 4)/0.99		D(−2/ − 5)	D(−3/ − 5)/1
Difference-in-Hansen *p* value		P(−2/ − 3)	P(−3)/0.54		P(−2/ − 3)	P(−3)/0.25		P(−2/ − 4)	P(−3)/0.68		P(−2/ − 5)	P(−3/ − 5) /0.53

Robust standard errors and Windjmeier correction for standard errors in BB. All regressions include country and time specific effects. Hansen test null hypothesis is the joint exogeneity of all instruments. Difference-in-Hansen test null hypothesis is the joint exogeneity of additional instruments used in SYS-GMM. Difference-in-Hansen test is also used separately for each instrumenting variable

using secondary and tertiary schooling. We also ran a horse race between primary, secondary and tertiary schooling. Results are presented in Table 8. As in Tables 4 and 7, most Hansen specification tests for AB regressions fail, so we rule out the AB estimator in the remainder of this paper, and focus on the DFE, BB and BC estimators. [24] The first panel of Table 8 confirms that average years of primary schooling is a strongly significant determinant of democracy in both DFE and BB regressions, and across all periods. All specification tests are valid and point-estimates of primary schooling are consistent across all regressions. By contrast, we find little evidence that secondary and tertiary schooling matter much for democracy—the signs on these variables are sometimes negative, often insignificant, and generally unstable across time periods and estimation methods.[25] We do find that, using the BB estimator, secondary schooling turns out positive and significant in all but one case, with satisfactory specification tests, but this is not robust to using the simpler DFE estimator, where we even get sign reversals.

An important new finding of this paper stems from the horse race between education measures at various levels, in the bottom panel of Table 8: When disaggregating average years of schooling into its three main components, primary schooling wins out—it bears a consistently positive and statistically significant estimated coefficient while the other measures of schooling are generally insignificant statistically.[26]

Our empirical results on the positive impact of primary education on democracy over the long period 1870–2000 is the central contribution of this paper and the result that sets it apart from existing research testing the modernization hypothesis.[27] It suggests that progress in democracy is achieved mostly when countries are still in their infancy in terms of educational development. In other words, increased political participation might involve the transition between illiteracy and literacy rather than further developments of secondary schooling and higher education, which often take place in already mature societies, and do not necessarily yield more stable political regimes over the short run. What matters for democracy is the average number of years of primary schooling in the adult population rather than the average total number of years of schooling.[28] Since higher education reached mass enrollment rates in high-income countries over the last quarter of the twentieth century, when democracy was already well-established, it is not surprising that higher education cannot explain the emergence of democracy.

4.3 Channels of democratization: income or primary education?

The results discussed above provide empirical support for a central issue in political economy: The quantitative evidence demonstrates that economic development and modernization,

[24] However, AB estimates for the corresponding specifications are available upon request.

[25] This finding may be interpreted as the sign of multicollinearity problems when variables are introduced altogether.

[26] We also ran regressions using the total number of years of education in the adult population (the sum of primary, secondary and tertiary). We found mixed evidence that the overall stock of education was significantly associated with democracy. Results were strongest using the state-of-the-art BB estimator irrespective of the period under consideration. Results for overall educational attainment are available upon request.

[27] Recent papers reaching the conclusion that education is an important determinant of democratization, albeit in contexts that are very different from ours, include Milligan, Moretti and Oreopoulos (2004) for the US and the UK, and Campante and Chor (2012) for the Arab Spring.

[28] Over the period 1960–2000, Castelló-Climent (2008) find that the education level of the first three quintiles of the education distribution is a more robust determinant of democracy than average years of schooling in the population. As the population at the bottom of the education distribution receive only primary education in early stages of economic development, her measure and ours are quite similar to each other.

Table 8 Democracy and average years of primary, secondary and tertiary schooling

	1870–2000		1900–2000		1930–2000		1960–2000	
	DFE	BB	DFE	BB	DFE	BB	DFE	BB
Dependent variable is standardized Polity IV index of democracy								
Primary								
D(−1)	0.376***	0.547***	0.295***	0.544***	0.149***	0.323***	0.097	0.382***
	(0.04)	(0.08)	(0.04)	(0.08)	(0.05)	(0.12)	(0.07)	(0.15)
P(−1)	0.044**	0.047***	0.062***	0.048***	0.071***	0.068***	0.085**	0.080***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)	(0.03)
N	615	613	542	540	429	427	312	310
N countries	70	70	70	70	70	70	70	70
N instruments		55		54		49		36
AB1 *p* value		0		0		0		0
AB2 *p* value		0.17		0.17		0.20		0.73
Hansen *p* value		0.09		0.09		0.21		0.11
Difference-in-Hansen *p* value		0.28		0.29		0.72		0.81
Nature of instruments and		D(−3)/0.87		D(−3)/0.88		D(−3/ − 4)/0.75		D(−3/ − 5) /0.70
Differences-in-Hansen *p* value		P(−4)/0.12		P(−4)/0.13		P(−4)/0.69		P(−4)/0.63
Secondary								
D(−1)	0.390***	0.660***	0.313***	0.659***	0.169***	0.417***	0.103	0.611***
	(0.04)	(0.08)	(0.04)	(0.08)	(0.05)	(0.14)	(0.06)	(0.16)
S(−1)	−0.026	0.044***	−0.033	0.044***	−0.070**	0.079***	−0.136***	0.046
	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)	(0.04)	(0.03)
N	615	613	542	540	429	427	312	310
N countries	70	70	70	70	70	70	70	70
N instruments		55		54		49		36

Table 8 continued

	1870–2000		1900–2000		1930–2000		1960–2000	
	DFE	BB	DFE	BB	DFE	BB	DFE	BB
AB1 *p* value		0		0		0		0
AB2 *p* value		0.14		0.14		0.14		0.50
Hansen *p* value		0.18		0.18		0.15		0.05
Difference-in-Hansen *p* value		0.74		0.74		0.63		0.36
Nature of instruments and		D(−3)/0.91		D(−3)/0.91		D(−3/ − 4)/0.74		D(−3/ − 5)/0.41
Differences-in-Hansen *p* value		S(−4)/0.17		S(−4)/0.17		S(−4)/0.58		S(−4)/0.14
Tertiary								
D(−1)	0.391***	0.873***	0.313***	0.874***	0.168***	0.755***	0.118*	0.906***
	(0.04)	(0.06)	(0.04)	(0.06)	(0.05)	(0.13)	(0.07)	(0.09)
H(−1)	−0.255***	0.010	−0.287***	0.010	−0.388***	0.143	−0.457***	−0.030
	(0.09)	(0.09)	(0.10)	(0.09)	(0.12)	(0.15)	(0.14)	(0.11)
N	615	613	542	540	429	427	312	310
N countries	70	70	70	70	70	70	70	70
N instruments		55		54		49		36
AB1 *p* value		0		0		0		0
AB2 *p* value		0.14		0.14		0.10		0.46
Hansen *p* value		0.17		0.17		0.10		0.15
Difference-in-Hansen *p* value		0.49		0.50		0.37		0.22
Nature of instruments and		D(−3)/0.89		D(−3)/0.90		D(−3/ − 4)/0.64		D(−3/ − 5)/0.97
Differences-in-Hansen *p* value		H(−4)/0.20		H(−4)/0.21		H(−4)/0.21		H(−4)/0.06

Table 8 continued

	1870–2000		1900–2000		1930–2000		1960–2000	
	DFE	BB	DFE	BB	DFE	BB	DFE	BB
All								
D(−1)	0.383***	0.603***	0.301***	0.600***	0.157***	0.316**	0.103	0.493***
	(0.04)	(0.09)	(0.04)	(0.09)	(0.05)	(0.15)	(0.07)	(0.19)
P(−1)	0.034*	0.074***	0.049**	0.074***	0.047*	0.086**	0.051	0.083***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)
S(−1)	0.002	−0.090	0.000	−0.091	−0.022	−0.058	−0.092*	−0.065
	(0.03)	(0.07)	(0.03)	(0.07)	(0.04)	(0.07)	(0.05)	(0.06)
H(−1)	−0.217*	0.209	−0.214	0.216	−0.257*	0.268	−0.185	0.052
	(0.12)	(0.32)	(0.13)	(0.33)	(0.15)	(0.35)	(0.19)	(0.33)
N	615	613	542	540	429	427	312	310
N countries	70	70	70	70	70	70	70	70
N instruments		55		54		49		36
AB1 *p* value		0		0		0		0
AB2 *p* value		0.16		0.16		0.21		0.62
Hansen *p* value		0.10		0.11		0.20		0.13
Difference-in-Hansen *p* value		0.29		0.29		0.83		0.78
Nature of instruments and		D(−3)/0.94		D(−3)/0.94		D(−3/ − 4)/0.70		D(−3/ − 5)/0.46
Difference-in-Hansen *p* value		P(−4)/0.09		P(−4)/0.09		P(−4)/0.44		P(−4)/0.16

Robust standard errors and Windjmeier correction for standard errors in BB. Comparable results are obtained with BB estimator when excluding initial democracies. All regressions include country and time specific effects. Hansen test null hypothesis is the joint exogeneity of all instruments. Difference-in-Hansen test null hypothesis is the joint exogeneity of additional instruments used in SYS-GMM. Difference-in-Hansen test is also used separately for each instrumenting variable

captured either by log GDP per capita or average years of primary schooling, are strongly significant determinants of democracy. This result is robust to controlling for country fixed-effects, accounting for persistence in the dependent variable and instrumenting using lagged explanatory variables in both levels and first differences.

Table 9 investigates which of the two variables—log GDP per capita or average years of primary schooling—has the greatest explanatory power using the full sample of countries. Primary education appears to be a more robust determinant. Indeed, it is significant in all 8 specifications with either DFE or BB estimators, while lagged log GDP per capita is significant in 3 DFE regressions but is never significant in BB regressions. Focusing on the BB estimator, we find satisfactory specification tests in all cases as the Hansen test p value is comprised between 0.1 and 0.25, limiting the number of instruments as in former tables. The coefficient on primary schooling is rather similar across periods and is comprised between 0.06 and 0.09 in BB regressions. Finally, turning to the BC estimator, in column 3 of Table 5 we find that both log income and primary schooling retain their significance, but that primary schooling is more significant than log income—in line with the results in Table 9.

4.4 Robustness analysis

We ran several robustness checks and they largely confirmed the validity of our results. First, we defined subperiods differently, and found that our main results still hold over the two globalization periods (1870–1910 and 1960–2000) and the period of retreat from globalization that corresponds roughly to the Interwar period (1910–1960).[29] Using a BB estimator, we found that primary education has been the key determinant of democracy in all these sub-periods. In terms of the magnitude of the effect, it was comparable across all three sub-periods and similar to the one described above. Interestingly, income has not been a significant determinant of democracy until the postwar period.

Second, we found that our results are unchanged when using a different democracy index, namely the Freedom House indices of political rights and civil liberties over the period 1960–2000. This result was largely expected as it was already highlighted in Bobba and Coviello (2007), although the latter authors used different education data.

Finally, sample composition could further affect our results as the sample does not remain fixed over time. In practice, countries join the sample at their date of independence, but many young countries have experienced erratic political processes, including declines in democracy in some cases. Their inclusion in the sample could create composition effects and spuriously affect the long-term interpretation of our results. We therefore restricted the sample to a balanced panel of 19 countries observed over the two periods 1870–2000 and 1960–2000, using alternatively DFE or BB. [30] In the latter case, to limit the problem of instrument proliferation we only use one lag per independent variable to generate the moment conditions, namely the fourth lags of democracy and primary education. Even with this important constraint, the number of instruments is larger than the number of countries, which hinders the validity of the Hansen test due to the instruments proliferation problem.

[29] The corresponding empirical results are available upon request.

[30] Results are available upon request. The 19 countries in the balanced panel were Argentina, Austria, Brazil, Canada, Switzerland, Spain, France, the United Kindgom, Greece, Hungary, Iran, Italy, Mexico, New Zealand, Portugal, Sweden, Thailand, the United States, and Venezuela. While it may be tempting to assume that countries for which data is constinuously available since 1870 may not have experienced large shifts in democracy ovr the 1870–2000 period, this is not the case, as demonstrated by the right-hand panel of Table 1, which summarizes the level of democracy per period for this subsample.

Table 9 Democracy, GDP per capita and average years of primary schooling

	1870–2000		1900–2000		1930–2000		1960–2000	
	DFE	BB	DFE	BB	DFE	BB	DFE	BB
Dependent variable is standardized Polity IV index of democracy								
D(−1)	0.331***	0.722***	0.257***	0.723***	0.139**	0.638***	0.111	0.662***
	(0.04)	(0.11)	(0.05)	(0.11)	(0.06)	(0.21)	(0.07)	(0.14)
Log y(−1)	0.102**	−0.013	0.123***	−0.013	0.131**	−0.033	0.079	−0.018
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.07)	(0.06)	(0.06)
P(−1)	0.040**	0.061***	0.064***	0.061***	0.075***	0.088***	0.087**	0.076**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)
N	560	560	505	505	417	381	307	307
N countries	69	69	69	69	69	69	69	69
N instruments		57		56		50		46
AB1 *p* value		0		0		0		0
AB2 *p* value		0.19		0.19		0.20		0.56
Hansen *p* value		0.24		0.24		0.10		0.16
Difference-in-Hansen *p* value		0.95		0.95		0.31		0.08
Nature of instruments and		ly(−3)/0.63		ly(−3)/0.63		ly(−3/ − 4)/0.40		ly(−3/ − 5)/0.40
Difference-in-Hansen *p* value		P(−3)/0.69		P(−3)/0.69		P(−3/ − 4)/0.31		P(−3/ − 5)/0.38

All regressions include country and time specific effects. Hansen test null hypothesis is the joint exogeneity of all instruments. Difference-in-Hansen test null hypothesis is the joint exogeneity of additional instruments used in SYS-GMM. Difference-in-Hansen test is also used separately for each instrumenting variable

This is a problem that arises when the number of countries is small, but it is unavoidable here.

Despite this problem, we found evidence that primary education, and in a less robust way income, have been significant determinants of democracy over both periods. With DFE, both variables were always highly significant. With BB, income was significant over 1870–2000 but not over 1960–2000, while primary education was always significant, including in the horse race with income. Caution should be exercised in drawing strong inferences given the instrument proliferation problem in this context, but overall the balanced sample results confirm that primary education can be viewed as the most robust determinant of democracy.

5 From democracy to modernity?

5.1 From democracy to education

This section focuses on reverse causality from democracy to economic development. We start by examining the link from democracy to schooling. Two classes of theories aiming to explain the historical increase in schooling have emerged from past studies. The first one is again linked to Lipset's modernization hypothesis, as it emphasizes the role played by economic development and technological progress as determinants of education. It includes Unified Growth Theory (Galor and Weil 1996, 2000, Galor 2011), explaining the take-off in public schooling among advanced countries by the rising demand for skills at the onset of the second Industrial Revolution. To say this differently, the spread of education originates from "the race between education and technological progress" (Tinbergen 1975) that started in the second half of the nineteenth century.

The second class of theories highlights the role played by cultural or religious factors (Weber 1904/05; Becker and Woessmann 2009) and institutions (Acemoglu and Robinson 2000). In particular, Lindert (2004, chapters 5 and 15) documents how the gradual extension of the suffrage as well as decentralization fostered the development of a publicly funded mass education system among advanced countries over the nineteenth century. He concludes that "there is a strong link from the spread of political voice to the rise of tax-based primary education" (Lindert 2004, chapter 15, p. 50). According to "Lindert's hypothesis//, the extension of the suffrage should entail an increase in schooling. Although our Polity IV index captures several other dimensions besides voting rights (such as the rule of law), the latter hypothesis is consistent with, but not necessarily equivalent to, the hypothesis that democratization triggers the spread of education.

In this subsection we aim to formally test whether educational attainment of young cohorts is determined by economic progress or by democratization, while accounting for cross-country differences in cultural factors. Measurement issues are addressed as follows:

First, the educational attainment of the young is defined as the average years of schooling of pupils aged between 5 and 14. In practice, the latter variable is proxied by the observed average value of schooling among the cohort aged between 25 and 34 years twenty years later (i.e. observed at time $t + 2$).[31] It is worth noting than at date $t + 1$, the latter cohort corresponds to the population aged between 15 and 24 years, in other words pupils in mid-secondary or at the end of tertiary schooling. Consequently, changes in democracy taking

[31] This variable is taken from the dataset on education by age borrowed from the same sources as those underlying Morrisson and Murtin (2009).

place between two subsequent observations of our panel should affect the whole spectrum of pupils in school.

Second, standards of economic development are proxied by two variables, namely log GDP per capita and average years of schooling of the adult population. We can control for the lagged dependent variable to account for the autocorrelation of the flow of education (i.e. education of young generations). However, as we already control for the stock of education (i.e. education of the adult population), the lags of the dependent variable are implicitely accounted for. Having tried both the specification with and without the lagged dependent variable, we found qualitatively identical results. Here we choose to report the regressions without the lagged dependent variable that are less subject to multi-collinearity problems. Finally, country fixed-effects account for time-invariant factors such as religion or culture. We assess the following dynamic panel data model using the same instrumentation techniques as before:

$$E_{it} = a_i + b_t + \alpha D_{it} + \beta S_{it} + \gamma y_{it} + \upsilon_{it} \tag{6}$$

where E_{it} is average years of schooling completed by the cohort aged between 25 and 34 twenty years later, D_{it} the democracy score, S_{it} average years of schooling of population older than 15, y_{it} log GDP per capita, a_i and b_t respectively country and time fixed-effects. This specification allows us to test whether $\alpha > 0$, namely whether the democracy-child education linkage turns out significant holding fixed past increases in schooling and improvements in living standards.[32]

Table 10 presents the results of OLS estimates (imposing $a_i = 0$), panel fixed effects ($a_i \neq 0$) and BB estimation, controlling or not for the level of development. We ran two sets of regressions, one spanning the whole period, the other focusing on the period 1960–2000. In BB regressions, the Hansen test of joint exogeneity of the instruments were satisfied in all cases (the same rules as before concerning the selection of the number of lags, i.e. the number of moment conditions, apply—namely we minimize the problem of instrument proliferaton by limiting the number of moment conditions to be smaller than the number of countries).

Overall, weak results arise regarding the potential role of democracy as a determinant of human capital. Over the 1870–2000 period, democracy is a significant determinant of average years of schooling attained by young cohorts in OLS and FE regressions in columns I, III and V, but it is no longer significant once the level of development is controlled for, as columns II, IV and VI demonstrate. Over the 1960–2000 period, democracy is significant in the OLS (Column VII) and BB (Column XI) regressions, but switches sign when other variables are introduced (Column VIII and XII). In FE estimation, it is never significant (columns IX and X). In sum, democracy is a positive and significant determinant of schooling attained by young cohorts when it is the only one explanatory variable. When the level of development is controlled for, it is never significant among the 6 corresponding specifications. Hence, we find little evidence that democracy leads to higher educational attainment. [33]

[32] In results available upon request, we introduced a cubic in the level of democracy to detect potentially non-linear effects as suggested by Lindert (2004). We could not find any evidence of non-linearity.

[33] In contrast to Lindert's view that the extension of the suffrage should lead to differences in educational policy and spending, Mulligan, Gil and Sala-i-Martin (2004, 2010) show that, on a wide array of policy metrics, including Social Security, democracies do not significantly differ from non-democracies. This is consistent with our finding that democracy does not causally affect educational attainment very strongly.

Table 10 Effect of democracy on average years of schooling of young cohorts

	1870–2000						1960–2000					
	OLS		FE		BB[a]		OLS		FE		BB[a]	
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Dependent variable is proxied average years of schooling of population aged 5–14[b]												
Full sample												
Democracy index	5.610***	0.050	0.450**	−0.159	10.486***	−0.546	5.871***	−0.434	−0.235	−0.230	11.169***	−0.724
	(0.322)	(0.156)	(0.194)	(0.173)	(1.037)	(0.466)	(0.547)	(0.269)	(0.377)	(0.372)	(2.367)	(0.938)
Average years of schooling of population 25+		0.974***		0.511***		1.062***		0.939***		−0.318		0.754***
		(0.031)		(0.058)		(0.158)		(0.054)		(0.205)		(0.164)
Log GDP per capita		0.681***		0.884***		0.318		0.733***		1.309***		1.377***
		(0.110)		(0.179)		(0.640)		(0.171)		(0.431)		(0.531)
N	553	500	553	500	553	500	195	193	195	193	195	193
N countries	70	69	70	69	70	69	69	68	69	68	69	68
N instruments					46	46					48	48
Hansen *p* value					0.16	0.14					0.19	0.25
Difference-in-Hansen *p* value					0.06	0.6					0.01	0.14
Nature of instruments and					D(−3)/0.96	D(−3)/0.66					D(−3/ − 8)/0.54	D(−3/ − 8)/0.72
Difference-in-Hansen *p* value					S(−4)/0.16	S(−4)/0.43					S(−4/ − 10)/0.01	S(−4/ − 10)/0.20

Robust standard errors and Windjmeier correction for standard errors in BB. All regressions include country and time specific effects. Hansen test null hypothesis is the joint exogeneity of all instruments. Difference-in-Hansen test null hypothesis is the joint exogeneity of additional instruments used in SYS-GMM. Difference-in-Hansen test is also used separately for each instrumenting variable

[a] Proxy is average years of schooling of population aged 25–34 observed 20 years later

5.2 From democracy to income

Turning to the relationship between GDP per capita and democracy, we simply run our dynamic panel data specifications, running income as a function of lagged democracy and lagged income. In other words, we "reverse" the baseline specification of Sect. 3, focusing similarly on DFE, AB and BB estimators. However, when applying the BB estimator, we find evidence of unit root as the coefficient on lagged income is very close to or even greater than unity.[34] To deal with this problem, we take the first-difference of all variables when applying the BB estimator. A similar operation applied to other regressions (DFE and AB) yields qualitatively similar results. Hence, we report the results when the following dynamic panel data model is estimated either in levels (DFE and FE) or in first-differences (BB):

$$\log y_{it} = a_i + b_t + \rho \log y_{i,t-1} + \alpha D_{i,t-1} + \upsilon_{it} \tag{7}$$

We find no robust effect of the lagged level of democracy on log per capita income. As reported in Table 11, we do not find any significant correlation between log GDP per capita and lagged democracy when using DFE, although the p value of the coefficient on lagged democracy for the period 1870–2000 is equal to 0.104. In unreported results disclosed in the working paper version of this study when using FE (i.e. excluding lagged income from the regression), we do find significant coefficients on democracy for the periods 1870–2000, 1900–2000 and 1930–2000. However, as shown in Table 11, this correlation vanishes once we control for the lagged level of log GDP per capita and use the AB estimator, whatever the period and country sample. Similarly, we find that growth of GDP per capita and the lagged change in democracy are never significantly associated, whatever the period and country sample, when using the BB estimator. The same result prevails independently from whether we control for lagged per capita GDP growth with AB, BB and DFE estimators.

In sum, the empirical evidence suggest that the level of income is more likely to be a determinant of the level of democracy than the other way around once we control for income persistence, and that the level of democracy has hardly been a significant determinant of economic growth over the twentieth century. This finding is consistent with a vast literature that seeks to explain economic growth as a function of the degree of democracy in the context of cross-country specifications similar to Eq. (7), finding generally inconclusive results. Among salient contributions, see Barro (1996), Helliwell (1994), Minier (1998), Tavares and Wacziarg (2001), and the survey in Borner, Brunetti and Weder (1995), among many others. More recent contributions include Persson and Tabellini (2006), who argue that the sequencing of economic and political reforms matters: democratic reforms are more conducive to growth if they are preceded by economic liberalizations. Also providing a nuanced view of the effects of democracy on economic performance, which includes an explicit consideration of the bi-directional causality linking the two concepts, Persson and Tabellini (2009) argue that a long experience with democratic institutions ("democratic capital") is conducive to development, and that in turn development consolidates democratic capital, leading to a virtuous circle.

Finally, some caveats are in order. First, our analysis is conditional on the definition of democracy as expressed by the Polity IV composite variable, which is to some degree subject to judgment calls and measurement error. Moreover, as suggested by Lindert's analysis, some specific dimensions of democracy such as voting rights may be more relevant as determinants

[34] This finding may be explained by the existence of country-specific deterministic trends that are not well captured by the set of time dummies, leaving room for the existence of an integrated process of order 1 that contaminates the residuals and biases upward the autocorrelation coefficient.

Table 11 Effect of democracy on log GDP per Capita

	1870–2000			1900–2000			1930–2000			1960–2000		
	DFE	AB	BB[a]	DFE	AB	BB[a]	DFE	AB	BB[a]	DFE	AB	BB[a]
Dependent variable is log GDP per capita												
Full sample												
Lagged log GDP per capita	0.812***	0.487***	0.207**	0.796***	0.506***	0.207**	0.759***	0.410***	0.175**	0.744***	0.493***	0.375***
	(0.032)	(0.119)	(0.101)	(0.036)	(0.128)	(0.100)	(0.043)	(0.121)	(0.082)	(0.051)	(0.171)	(0.116)
Lagged democracy	0.056	−0.095	0.119	0.049	−0.095	0.119	0.025	−0.048	0.070	−0.019	−0.160	0.166
	(0.035)	(0.070)	(0.084)	(0.038)	(0.071)	(0.084)	(0.046)	(0.084)	(0.108)	(0.055)	(0.116)	(0.109)
N	567	489	489	512	462	462	423	377	377	308	275	275
N countries	69	69	69	69	69	69	69	69	69	69	68	68
N instruments		58	52	69	55	52		56	57		35	36
AB1 *p* value		0.02	0		0.02	0		0.09	0		0.43	0.02
AB2 *p* value		0.53	0.41		0.52	0.41		0.37	0.31		0.5	0.59
Hansen *p* value		0.43	0.55		0.39	0.54		0.44	0.62		0.14	0.51
Difference-in-Hansen *p* value			0.27			0.27			0.22			0.16
Nature of instruments and		D(−2/ − 3)	D(−3)/0.75		D(−2/ − 3)	D(−3)/0.75		D(−2/ − 4)	D(−3/ − 4)/0.55		D(−2/ − 4)	D(−3/ − 4)/0.32
Difference-in-Hansen *p* value		ly(−2/ − 3)	ly(−3)/0.89		ly(−2/ − 3)	ly(−3)/0.89		ly(−2/ − 4)	ly(−3/ − 4)/0.93		ly(−2/ − 4)	ly(−3/ − 4)/0.98

Robust standard errors and Windjmeier correction for standard errors in BB. Dependent and explanatory variables as well as instruments are taken in first-difference in BB. All regressions include country and time specific effects. Hansen test null hypothesis is the joint exogeneity of all instruments. Difference-in-Hansen test null hypothesis is the joint exogeneity of additional instruments used in SYS-GMM. Difference-in-Hansen test is also used separately for each instrumenting variable

[a] Because of unit roots problems, all variables are taken in first-difference in BB regressions

of the spread of education than others. We leave the question of whether some subcomponents of the composite democracy index may causally affect education open for future research. Second, the effect of democracy on economic growth may depend on the state of development or some other interaction variable. For instance, Aghion, Alesina and Trebbi (2007) suggest that institutional features become favorable to economic growth when the economy is closer to the technological frontier, or for more technologically advanced sectors. Also pointing to a conditional effect, Aghion, Persson and Rouzet (2012) find that democracy displays a positive impact on child enrollment only in countries engaged into fierce economic or military rivalry with competing nations. Finally, Cervelatti and Sunde (2011) argue that violent conflict is a mediating variable linking democracy and economic performance: peaceful transitions to democracy are conducive to growth, while violent conflict reduces the effect of democracy on growth. Overall, these caveats do not undermine our general conclusion that over the long-run, primary education and to a lesser extent income levels have had a substantial impact on (broadly defined) democracy standards, while the converse hypothesis does not hold in its simple, unconditional form.

6 Conclusion

This paper provides empirical support for the modernization hypothesis in the context of an ongoing debate about the direction of causality linking democracy and economic development. Three features jointly set apart our approach from existing studies on this topic. First, in contrast with most existing research, we take a longer historical view, going back to 1870. This allows us to examine how the relationship might have changed over time. Second, we use modern dynamic panel data methods, including state-of-the-art bias correction methods, in a context where the long duration of the data affords substantial within-country variation, in contrast with studies that feature a compressed time frame. Finally, we are able to run horseraces between various measures of economic modernization (income per capita and human capital) and various new historical measures of education (primary, secondary and tertiary). To our knowledge, no existing paper combines these three novel features together. Doing so, we document a strong empirical link from the level of development, particularly as captured by the level of primary schooling, to democracy. Investigating reverse causality from democracy to the educational attainment of young cohorts over the long run, we find little evidence of a positive and significant link.

Future work should seek to better understand the specific microeconomic mechanisms that link primary education to democracy. One possibility is that illiteracy hinders the demand for democracy because it limits the ability of the population to learn about and monitor the actions of autocratic regimes. Another, not mutually exclusive possibility is that a minimal level of education is necessary for the supply of collective action needed to overthrow dictators. Another potential mechanism is that literacy, and more generally economic modernization, are associated with economic activities less geared toward subsistence and basic survival, freeing time and resources to actively demand more political rights. Literacy, finally, may create the conditions for a well-functioning expression of suffrage, and thus more sustainable democracy.

Acknowledgments We thank Maurice Bun, Anke Hoeffler, Peter Lindert, Casey Mulligan, three anonymous referees and the editor for helpful comments. This paper does not express the official views of the OECD.

References

Acemoglu, D., & Robinson, J. A. (2000). Why did the west extend the Franchise? Democracy, inequality, and growth in historical perspective. *Quarterly Journal of Economics*, *115*(4), 1167–1199.

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

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics*, *118*, 1231–1294.

Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2005). From education to democracy? *American Economic Review Papers and Proceedings*, *95*(2), 44–49.

Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2008). Income and democracy. *American Economic Review*, *98*(3), 808–842.

Aghion, P., Alesina, A., Trebbi, F. (2007). Democracy, Technology, and Growth. *NBER Working Paper* #13180 June.

Aghion, P., Persson, T., Rouzet, D. (2012). Education and Military Rivalry. *NBER Working Paper* #18049, May.

Anderson, T. G., & Sorenson, B. E. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics*, *14*, 328–352.

Arellano, M., & Bond, S. (1991). Some specification tests for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, *58*, 277–298.

Barro, R. (1996). Democracy and growth. *Journal of Economic Growth*, *1*(1), 1–27.

Barro, R. J. (1997). *The determinants of economic growth: A cross-country empirical study*. Cambridge: MIT Press.

Barro, R. J. (1999). The determinants of democracy. *Journal of Political Economy*, *107*, S158–S183.

Becker, S. O., & Woessmann, L. (2009). Was weber wrong? A human capital theory of protestant economic history. *Quarterly Journal of Economics*, *124*(2), 531–596.

Benhabib, J., Corvalan, A., Spiegel, M. (2011). Reestablishing the income-democracy nexus. *NBER Working Paper* #16832.

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

Bobba, M., & Coviello, D. (2007). Weak instruments and weak identification in estimating the effects of education on democracy. *Economics Letters*, *96*(3), 301–306.

Boix, C. (2011). Democracy, development, and the international system. *American Political Science Review*, *105*(4), 809–828.

Borner, S., Brunetti, A., & Weder, B. (1995). *Political credibility and economic development*. New York: Macmillan.

Bourguignon, F., & Morrisson, C. (2002). Inequality among world citizens: 1820–1992. *American Economic Review*, *92*(4), 727–744.

Bowsher, C. G. (2002). On testing overidentifying restrictions in dynamic panel data models. *Economics Letters*, *77*, 211–220.

Bruno, G. (2005). Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models. *Economics Letters*, *87*(3), 361–366.

Bun, Maurice J. G., & Carree, Martin A. (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics*, *23*(2), 200–210.

Bun, Maurice J. G., & Windmeijer, Frank. (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal*, *13*, 95–126.

Campante, F. R., & Chor, D. (2012). Why was the Arab world poised for revolution? Schooling, economic opportunities, and the Arab spring. *Journal of Economic Perspectives*, *26*(2), 167–188.

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

Castelló-Climent, A. (2008). On the distribution of education and democracy. *Journal of Development Economics*, *87*(2), 179–190.

Cervellati, M., Sunde, U. (2011). Democratization, Violent Social Conflicts, and Growth. *IZA Discussion Papers* #5643.

Fayad, G., Bates, R.H., Hoeffler, A. (2011). Income and democracy: Lipset's law inverted. *Oxcarre Working Paper* #61, University of Oxford, April.

Friedman, W., Kremer, M., Miguel, E., & Thornton, R. (2011). *Education as Liberation. Working paper*, UC Berkeley.

Galor, O. (2005). *The transition from stagnation to growth: Unified growth theory*. North Holland: Handbook of Economic Growth.

Galor, O. (2011). *Unified growth theory*. Princeton: Princeton University Press.
Galor, O., & Weil, D. N. (1996). The gender gap, fertility, and growth. *American Economic Review*, *86*, 374–387.
Galor, O., & Weil, D. N. (2000). Population, technology, and growth: From the malthusian regime to the demographic transition and beyond. *American Economic Review*, *90*, 806–828.
Glaeser, E., Campante, F. (2009). Yet Another Tale of Two Cities: Buenos Aires and Chicago. *NBER Working Paper* #15104, June.
Glaeser, E., Ponzetto, G., & Shleifer, A. (2007). Why does democracy need education? *Journal of Economic Growth*, *12*(2), 77–99.
Glaeser, E., La Porta, R., Lopez-de-Silanes, R., & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth*, *9*(3), 271–303.
Grosjean, P., & Senik, C. (2011). Democracy, market liberalization, and political preferences. *Review of Economics and Statistics*, *93*(1), 365–381.
Gundlach, E., & Paldam, M. (2012). The democratic transition: Short-run and long-run causality between income and the gastil index. *European Journal of Development Research*, *24*, 144–168.
Hatton, T. J., & Williamson, J. G. (2005). *Global migration and the world economy: Two centuries of policy and performance*. Cambridge: MIT Press.
Hauk, W. R., & Wacziarg, R. (2009). A Monte Carlo study of growth regressions. *Journal of Economic Growth*, *14*(2), 103–147.
Helliwell, J. (1994). Empirical linkages between democracy and economic growth. *British Journal of Political Science*, *24*, 225–248.
Islam, N. (1995). Growth empirics: A panel data approach. *Quarterly Journal of Economics*, *110*(4), 1127–1170.
Jefferson, T. (1779). *A Bill for the More General Diffusion of Knowledge, Bill 79 in Report of the Committee of Revisors Appointed by the General Assembly of Virginia in 1776* (p. 1784). Richmond, VA: Dixon and Holt.
Kalemli-Ozcan, S., Ryder, H. E., & Weil, D. N. (2000). Mortality decline, human capital investment, and economic growth. *Journal of Development Economics*, *62*(1), 1–23.
Lindert, P. (2004). *Growing public*. Cambridge: Cambridge University Press.
Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review*, *53*, 69–105.
Lokshin, B. (2009). Further results on bias in dynamic unbalanced panel data models with an application to firm R&D investment. *Applied Economics Letters*, *16*(12), 1227–1233.
Maddison, A. (2006). *The world economy*. Paris: OECD.
Marshall, M. G., Jaggers, K. (2008). Polity IV Project. http://www.systemicpeace.org/polity/polity4.htm.
Milligan, K., Moretti, E., & Oreopoulos, P. (2004). Does education improve citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics*, *88*(9–10), 1667–1695.
Minier, J. A. (1998). Democracy and growth: Alternative approaches. *Journal of Economic Growth*, *3*(3), 241–266.
Morrisson, C., & Murtin, F. (2009). The century of education. *Journal of Human Capital*, *3*(1), 1–42.
Morrisson, C., & Murtin, F. (2013). The kuznets curve of education: A global perspective on education inequalities. *Journal of Economic Inequality*, *11*(3), 283–301.
Mulligan, C. B., Gil, R., & Sala-i-Martin, X. (2004). Do democracies have different public policies than nondemocracies? *Journal of Economic Perspectives*, *18*(1), 51–74.
Mulligan, C. B., Gil, R., & Sala-i-Martin, X. (2010). Social security and democracy. *The B.E. Journal of Economic Analysis and Policy*, *10*(1), 1–27.
Murtin, F., & Viarengo, M. (2010). American education in the age of mass migrations 1870–1930. *Cliometrica: Journal of Historical Economics and Econometric History*, *4*(2), 113–139.
Murtin, F. (2013). Long-term determinants of the demographic transition: 1870–2000. *Review of Economics and Statistics*, *95*(2), 617–631.
Murtin, F., Wacziarg, R. (2011). The Democratic Transition. *NBER Working Paper* #17432, August.
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*, 1417–1426.
O'Rourke, K., & Williamson, J. G. (1999). *Globalization and history: The evolution of a 19th century atlantic economy*. Cambridge: MIT Press.
Persson, T., & Tabellini, G. (2009). Democratic capital: The nexus of political and economic change. *American Economic Journal*, *1*(2), 88–126.
Persson, T., & Tabellini, G. (2006). Democracy and development: The devil in the details. *American Economic Review*, *96*(2), 319–324.
Rodrik, D., Wacziarg, R. (2005). Do Democratic Transitions Produce Bad Economic Outcomes?. In *American Economic Review Papers and Proceedings*, May 2005, (pp. 50–55).

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

Roodman, D. M. (2009b). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, *9*(1), 86–136.

Schumpeter, J. (1942). *Capitalism, socialism and democracy*. New York: Penguin Classics.

Tavares, J., & Wacziarg, R. (2001). How democracy affects growth? *European Economic Review*, *45*(8), 1341–1379.

Tinbergen, J. (1975). *Income distribution: Analysis and policies*. Amsterdam: North Holland.

Tocqueville, A. (1835). *Democracy in America*. New York: Penguin Classics.

Treisman, D. (2011). Income, Democracy and the Cunning of Reason. *NBER Working Paper* #17132.

Weber, Max (1904/05). Die protestantische Ethik und der, Geist" des Kapitalismus. Archiv für Sozialwissenschaft und Sozialpolitik 20: 1–54 and 21: 1–110. Reprinted in: Gesammelte Aufsätze zur Religionssoziologie, 1920: 17–206.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, *126*, 25–51.