Gong, Da; Hammar, Olle

Working Paper

Replication Report: A Comment on Gethin, Martínez-Toledano & Piketty (2022)

I4R Discussion Paper Series, No. 19

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Gong, Da; Hammar, Olle (2023) : Replication Report: A Comment on Gethin, Martínez- Toledano & Piketty (2022), I4R Discussion Paper Series, No. 19, Institute for Replication (I4R), s.l.

This Version is available at:
http://hdl.handle.net/10419/268753

INSTITUTE for REPLICATION

No. 19
I4R DISCUSSION PAPER SERIES

Repliation Report: A Comment on Gethin, Martínez-Toledano & Piketty (2022)

Da Gong

Olle Hammar

Februar 2023

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 19

Replication Report: A Comment on Gethin, Martínez-Toledano & Piketty (2022)

Da Gong[1], Olle Hammar[2]

[1]*University of California, Riverside/USA*
[2]*Research Institute of Industrial Engineering and Institute of Future Studies, Stockholm/Sweden*

FEBRUAR 2023

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Replication Report

A comment on Gethin, Martínez-Toledano & Piketty (2022)

Da Gong[1] and Olle Hammar[2]

Abstract

Gethin, Martínez-Toledano and Piketty (2022) analyze the long-run evolution of political cleavages using a new database on socioeconomic determinants of voting from approximately 300 elections in 21 Western democracies between 1948 and 2020. They find that, in the 1950s and 1960s, voting for the "left" was associated with lower-educated and low-income voters. After that, voting for the "left" has gradually become associated with higher-educated voters, while high-income voters have continued to vote for the "right". In the 2010s, there is a disconnection between the effects of income and education on voting. In this replication, we first conduct a computational reproduction, using the replication package provided by the authors. Second, we do a robustness replication testing to what extent the original results are robust to i) restricting the sample to "core" left and right parties, ii) analyzing the top 80% versus bottom 20%, iii) weighting by population, iv) dropping control variables, and v) using country fixed effects. The main results of the paper are found to be largely replicable and robust.

[1] Affiliation: University of California (UC), Riverside. Contact information: dgong005@ucr.edu. No conflict of interest.

[2] Corresponding author. Affiliations: Research Institute of Industrial Economics (IFN) and Institute for Futures Studies. Contact information: olle.hammar@ifn.se. No conflict of interest.

1. Introduction

Amory Gethin, Clara Martínez-Toledano and Thomas Piketty (2022), henceforth GMP, study the evolution of voting preferences in 21 Western democracies between 1948 and 2020. They study this trough a new database on individual-level socioeconomic determinants of voting from approximately 300 elections. The analysis is descriptive rather than causal. Their main finding "is the existence of a gradual process of disconnection between the effects of income and education on the vote" (p. 3). While voting for the "left" used to be associated with lower-educated and low-income voters, it has gradually become associated with higher-educated voters, resulting in a political cleavage between the high-income "Merchant right" versus the higher-educated "Brahmin left".

In this replication, we test the computational reproducibility and robustness replicability of the main GMP results. First, we test the reproducibility of the results using the replication package provided by the authors. Second, we test the replicability of the results through the following robustness tests: i) restricting the sample to "core" left and right parties, ii) analyzing top 20% versus bottom 80%, iii) weighting by population, iv) dropping control variables, and v) using country fixed effects. Finally, we also do a replication exercise where we test to what extent the GMP results are affected by some ex-post recoding that we noticed when going through the authors' do-files.

With the exception of a very minor issue in the codes, we are able to run the full replication package provided by the authors, and we find the GMP results to be fully reproducible. Naturally, our robustness replications change the results somewhat, but our overall conclusion is that the main GMP results are replicable and robust to our checks.

2. Reproducibility

We describe in this section one minor error that we uncovered while reproducing the study. When running the codes in the replication package provided by the authors and

through the *Quarterly Journal of Economics* (QJE),[3] we found that one of the do-files (**results-all-1-region.do**) did not run correctly but stopped for Belgium (*BE*) with the error code "*convergence not achieved*". When excluding Belgium from that do-file,[4] we were able to run all codes from beginning to end. This did not change the results. Hence, the results were reproducible.

3. Replication

We now turn to our replication. We test the replicability of the GMP main results through five additional robustness tests. The decision to conduct these robustness analyses was taken after reading the paper. The decision to also conduct an ex-post recoding replication was taken after observing the codes. We did not pre-register our replication analysis.

3.1 Robustness replication

3.1.1 Restricting the sample to "core" left and right parties

As illustrated in GMP Appendix Figures A37 and A38, there has been a change in the composition of left- and right-wing parties over the studied period. As a robustness check, to see whether the main results in the paper are driven by this compositional change, we replicate the main analysis but only include the "core" left-wing (i.e., social democrats/socialists) and "core" right-wing (i.e., conservatives/Christian democrats) parties in the analysis. As such, this robustness check is an extension to GMP Appendix Figures A21 and A22, in which the authors exclude the green parties. The difference in our replication is that we exclude even more parties, namely the greens, communists, liberals/social-liberals and anti-immigration parties.[5] The results from this robustness replication are presented in **Figure 1** (balanced panel) and **Figure 2** (unbalanced panel).

[3] Available at Harvard Dataverse: https://doi.org/10.7910/DVN/XUSWG6 (2022-11-26).
[4] We do this by changing line 12 in **results-all-1-region.do** from:
`levelsof iso, local(countries)`
to:
`levelsof iso if iso != "BE", local(countries).`
[5] We do this by changing the "*redefine old left as left – green, and old right as right – anti-immigration*" to "*redefine Old left as Left – Green – Communist – Liberal – Regional*" and "*redefine old right as Right –*

Excluding these parties from the sample reduces the income/education gap compared to baseline. Compared with the authors' analysis where they exclude the green parties, however, our robustness replication where we exclude even more parties gives very similar results. The main results thus seem to be robust to only including the "core" left- and right-wing parties.

3.1.2 Top 20% vs. bottom 80%

Most of the analysis in GMP focuses on top 10% versus bottom 90%. As a robustness check, they also do the analysis for top 50% versus bottom 50% (GMP Appendix Figures A3 and A4). In this robustness replication, we additionally analyze the main results for top 20% versus bottom 80%,[6] to check whether the main results are driven by the authors' choice of quantile groups. The results from this robustness replication are presented in **Figure 3** (balanced panel) and **Figure 4** (unbalanced panel).

Similar patterns are found when analyzing top 20% versus bottom 80% instead of top 10% versus bottom 90% (or top 50% versus bottom 50%), which suggests that the main results are robust to this replication.

3.1.3 Weighted by population

The main results displayed in GMP Figure 1 is calculated as the following procedure: First, they get the coefficient β for each year and each country. Second, for each country, they calculate the average β within the 5-year time interval. Third, for each 5-year interval,

Anti-immigration – Liberal – Regional" or, more explicitly, by adding the following after lines 254 and 255 in the **results-all-1-educ.do** and after lines 253 and 254 in **results-all-1-inc.do**, respectively:

```
replace voteolf = 0 if votegroup == "Communist"
replace voteolf = 0 if voteolf == 1 & votegroup == "Liberal"
replace voteolf = 0 if voteolf == 1 & votegroup == "Regional"

replace voteolr = 0 if voteolr == 1 & votegroup == "Liberal"
replace voteolr = 0 if voteolr == 1 & votegroup == "Regional".
```

[6] We do this by changing the "*Top 50% – bottom 50%*" analysis to "*Top 20% – bottom 80%*" or, more explicitly, by adding the following before line 47 in **results-all-1-educ.do**:

```
replace geduc_1 = qeduc_1 + qeduc_2 + qeduc_3 + qeduc_4,
```

and, similarly, the following before line 47 in **results-all-1-inc.do**:

```
replace ginc_1 = qinc_1 + qinc_2 + qinc_3 + qinc_4.
```

they calculate the β across countries without weighting and display the coefficients. In this subsection, we modify the third step by weighting the average coefficients with the population of each country at that time.

To be specific, the population data is obtained from the World Bank[7] and the average population for each country within each 5-year time interval is calculated. Next, the population data is merged with the coefficients for GMP Figure 1 in the GMP data replication file (gmp-macro.dta). Finally, the population-weighted coefficients across countries for each 5-year time interval are calculated and displayed in **Figure 5**.

Compared to GMP's main results, the weighted results show a similar trend but with more fluctuation. As depicted by the upper dashed line, the highest-educated voters were less likely to vote for social democratic parties than were lowest-educated voters by 12 percentage points in 1960s, which is 3 percentage points lower than GMP's findings. Furthermore, the reversal in the educational divide occurred in the 1980s, which is 10 years earlier than GMP's estimate. The evolution of income also stays robust, though with additional fluctuations. On average, the population-weighted results show smaller gaps between the top-income voters and low-income voters than the original results. This difference becomes noticeable after the 2010s, where the gap narrows to –5 percentage points compared to –10 percentage points in GMP's results.

3.1.4 Drop control variables

The estimates displayed in GMP Figure 1 control for income/education, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status and marital status. Although their results remain robust even when controls are dropped, as demonstrated in their paper, they do not examine the extent to which the main results would be impacted if any single control were omitted from the baseline model.

In this section, we evaluate the primary findings by omitting each control variable one by one and present the results for education and income separately in **Figure 6** and **Figure**

[7] Available at the World DataBank: https://data.worldbank.org/indicator/SP.POP.TOTL (2022-11-26).

7. For each estimate, we omit one of the following controls: age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status and marital status.[8] The GMP results for education and income are represented by the red and blue curves respectively. The green curves show the results when the corresponding control variable is omitted, as indicated in the figure legend. A similar pattern is observed for all the models, implying that the primary results are robust to the exclusion of any individual control.

3.1.5 Country fixed effects

As discussed in section 3.1.3, the coefficients presented in GMP's main results are two-way averages at the country-year level. In this subsection, we modify their specification by using a country fixed effect model. Our equation is:

$$y_{ict} = \alpha + \beta_t x_{ict} + \boldsymbol{C}_{ict}\gamma + \mu_c + \varepsilon_{ict}$$

where y_{ict} is a binary variable of interest for individual *i* in country *c* in election *t*, x_{ict} is a binary explanatory variable of interest and $\boldsymbol{C}_{ict}$ is a vector of controls (for more details, see original paper). This setting differs from the original study in that we control for country fixed effects, μ_c, by using pooled samples from each year. To be specific, for each year, we run the regression using a sample of all 12 countries while including all the controls and controlling for country fixed effects. Then, we calculate the average of β_t within the 5-year time interval and present the results in **Figure 8**.

In general, similar patterns emerge when the country fixed effect approach is employed. As depicted in **Figure 8**, the dashed lines represent results that account for country fixed effects, while the solid lines represent the main results from GMP. The only discrepancy is observed from the 1960s to 1970s, where the fixed effect model portrays a declining trend in education, while the original results depict an upward trend. Despite both models displaying a downward trend in income, the fixed effect approach indicates a steeper

[8] To stay consistent with the main idea of this paper, when estimating trends for education, income is always included in the regression model and vice versa.

decline in comparison to GMP's results. Overall, however, the results are very similar and we consider the original results to be robust to this specification.

3.2 Ex-post recoding replication

While doing the replication, we noticed that some codes (mainly for GMP Figures 2 and 3 in the main paper) included a few potentially ad hoc data adjustments. While we do not take a standpoint on to what extent these manual adjustments are reasonable (which they may very well be), here we rather check whether they affect the main results or not. We do this by simply rerunning the codes without these adjustments.[9] The results from this ex-post recoding replication are shown in **Figure 9** (corresponding to Figure 2 in GMP) and **Figure 10** (corresponding to Figure 3 in GMP).

We can see that these adjustments change a few data points, e.g., Canada 1960–69 and Iceland 1970–79 in GMP Figure 2A, Canada 1960–79 and Norway 1960–69 in GMP Figure 3A, and Luxembourg 2000–09 in GMP Figure 3B. While the motivation for some of these adjustments are partly provided in the code (under "*Small fixes by variables*" in **results-paper-main.do**), they are mostly motivated as "*outliers*", and sometimes only as "*too inconsistent*" or "*bad data*". As such, this information is not sufficient to evaluate the validity of these adjustments. Moreover, it would have been more transparent to describe them directly in the main paper or appendix. Nevertheless, the main results and overall conclusions of the paper are not affected by including/excluding these minor adjustments.

Finally, we also did another ex-post recoding replication by excluding similar adjustments in some of the other do-files.[10] These adjustments did not affect any of the main results or conclusions of the paper.

[9] We do this by excluding lines 167–204 in **results-paper-main.do** (see Appendix I for the exact code).
[10] The additional adjustments we tried excluding are line 65 in **results-all-1-religion.do** and lines 10–13, 35–38 and 45–46 in **results-all-1-race.do** (see Appendix I for the exact code).

4. Conclusion

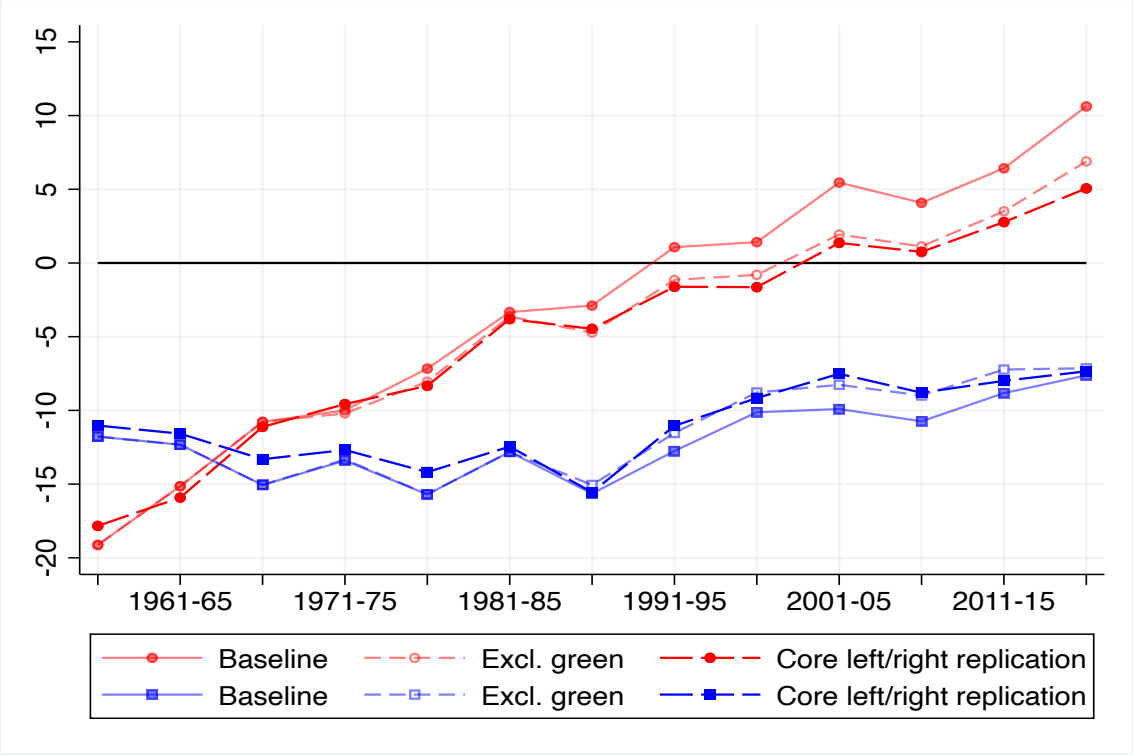
In this report we have presented the results from a replication of Gethin, Martínez-Toledano and Piketty (GMP, 2022). We have tested the computational reproducibility and robustness replicability of the GMP main results. Our replication has been limited in the sense that we have only focused on the paper's main result and only in the form of additional robustness tests. Based on our analyses, we conclude that the GMP main results are reproducible and robust to our replications.

References

Gethin, A., Martínez-Toledano, C., and Piketty, T. (2022). Brahmin Left versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948–2020. *The Quarterly Journal of Economics*, 137(1), 1–48.

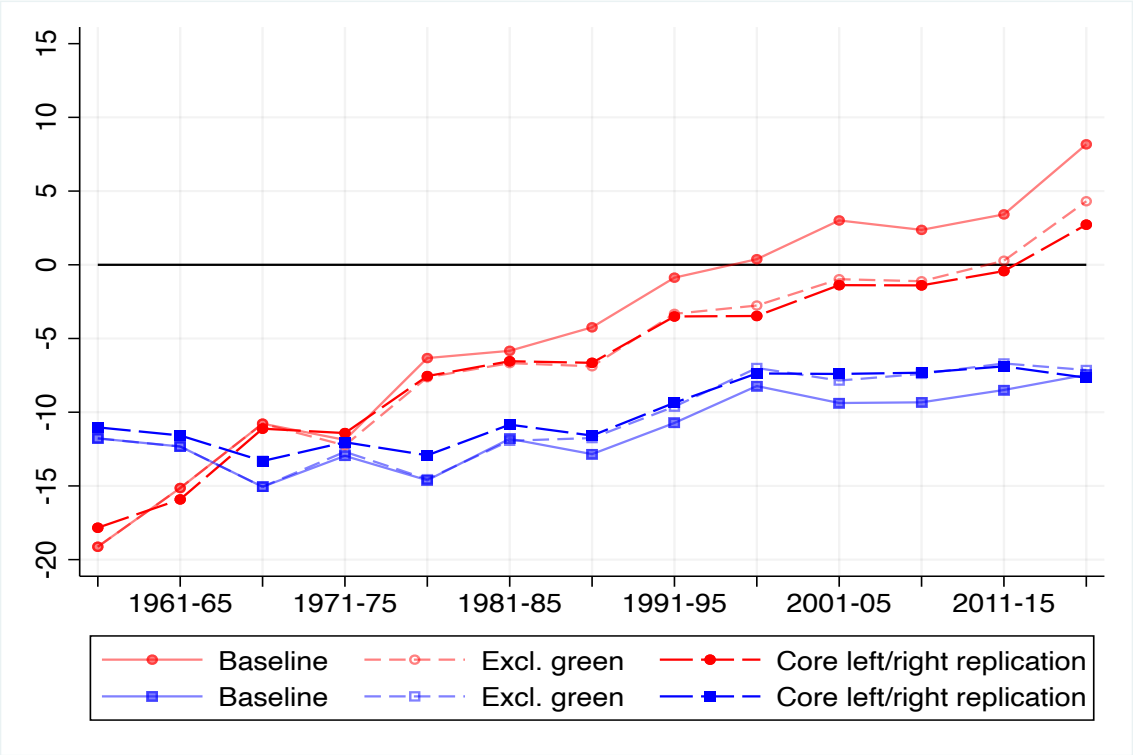
Figures

Figure 1 – The disconnection of income and education cleavages in Western democracies: Including/excluding parties (balanced panel)



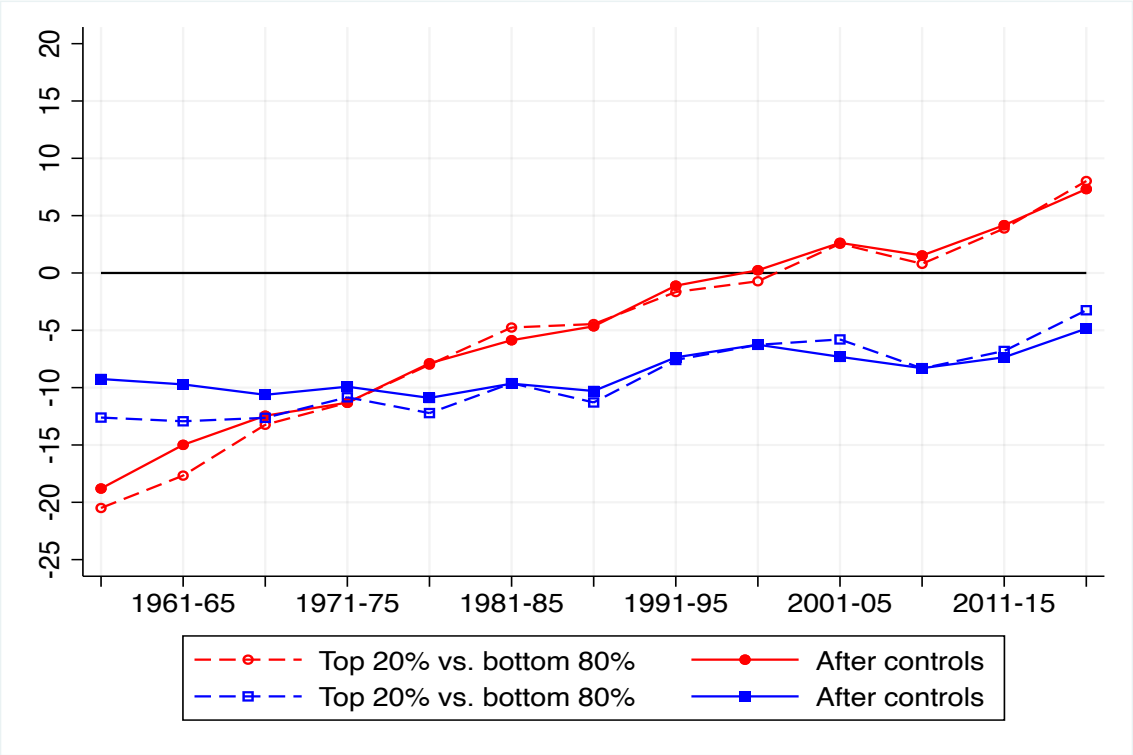
Note: This figure replicates and extends Appendix Figure A21 in GMP. "Baseline" is from the original study and shows the difference between (% of top 10% educated/earners voting left) and (% of bottom 90% educated/earners voting left). Red lines show results for education and blue lines show results for income. "Excl. green" is also from the original study and shows the results excluding green parties. "Core left/right replication" is our robustness replication excluding greens, communists, liberals/social-liberals and anti-immigration parties (i.e., only including social democrats/socialists as the "core" left-wing parties and conservatives/Christian democrats as the "core" right-wing parties). The results in this figure are for the balanced panel of countries.

Figure 2 – The disconnection of income and education cleavages in Western democracies: Including/excluding parties (unbalanced panel)



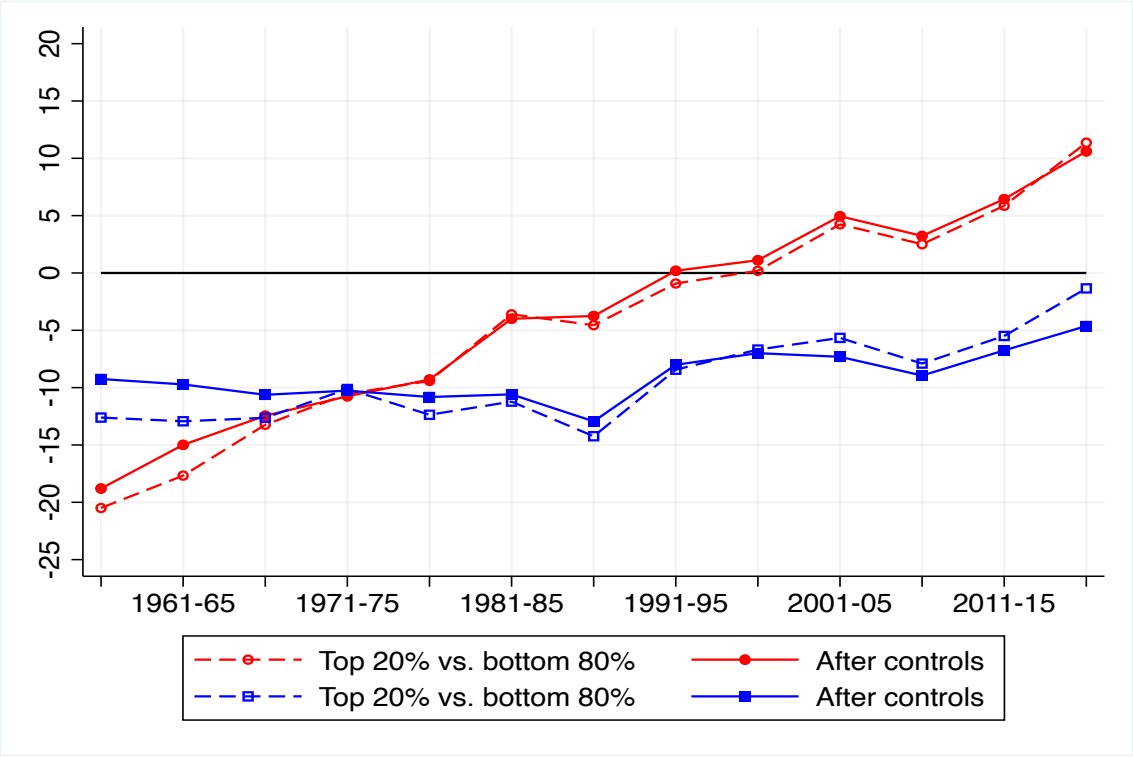
Note: This figure replicates and extends Appendix Figure A22 in GMP. "Baseline" is from the original study and shows the difference between (% of top 10% educated/earners voting left) and (% of bottom 90% educated/earners voting left). Red lines show results for education and blue lines show results for income. "Excl. green" is also from the original study and shows the results excluding green parties. "Core left/right replication" is our robustness replication excluding greens, communists, liberals/social-liberals and anti-immigration parties (i.e., only including social democrats/socialists as the "core" left-wing parties and conservatives/Christian democrats as the "core" right-wing parties). The results in this figure are for the unbalanced panel of countries.

Figure 3 – The disconnection of income and education cleavages in Western democracies: Top 20% vs. bottom 80% (balanced panel)



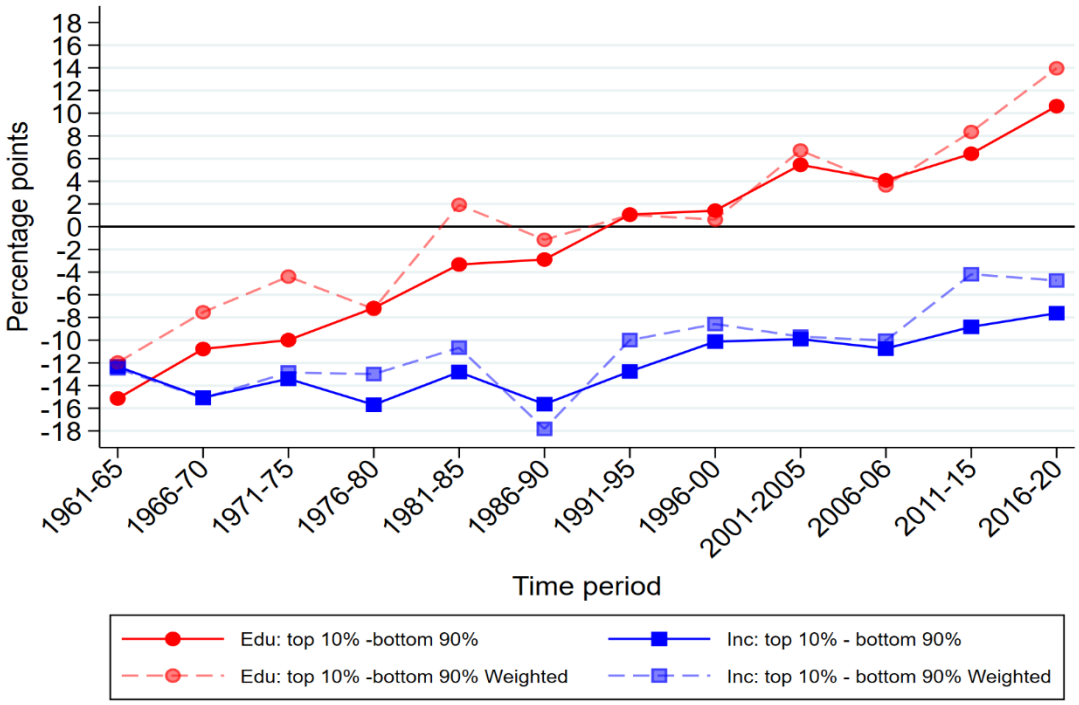
Note: This figure is similar to Appendix Figures A1 and A3 in GMP, but shows the results for top 20% vs. bottom 80% (instead of top 10% vs. bottom 90% or top 50% vs. bottom 50%). Dashed lines show the difference between (% of top 20% educated/earners voting left) and (% of bottom 80% educated/earners voting left) before controls. Solid lines show the same results after controls. Red lines show results for education and blue lines show results for income. The results in this figure are for the balanced panel of countries.

Figure 4 – The disconnection of income and education cleavages in Western democracies: Top 20% vs. bottom 80% (unbalanced panel)



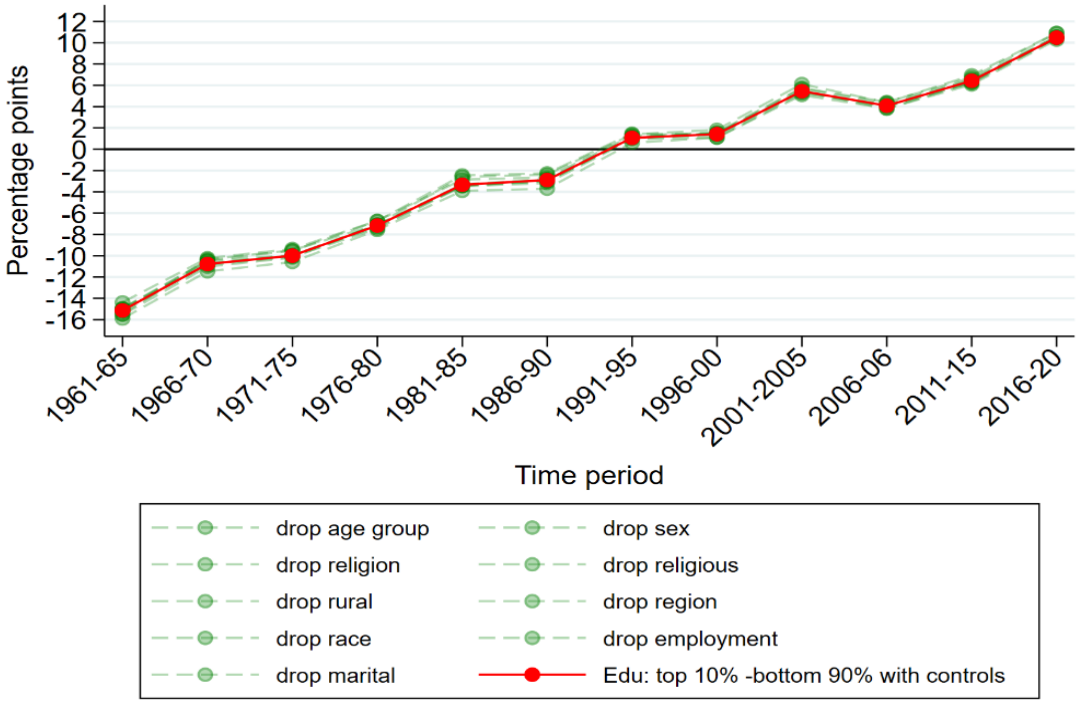
Note: This figure is similar to Appendix Figures A2 and A4 in GMP, but shows the results for top 20% vs. bottom 80% (instead of top 10% vs. bottom 90% or top 50% vs. bottom 50%). Dashed lines show the difference between (% of top 20% educated/earners voting left) and (% of bottom 80% educated/earners voting left) before controls. Solid lines show the same results after controls. Red lines show results for education and blue lines show results for income. The results in this figure are for the unbalanced panel of countries.

Figure 5 – The disconnection of income and education cleavages in Western democracies: Weighted by population (balanced panel)



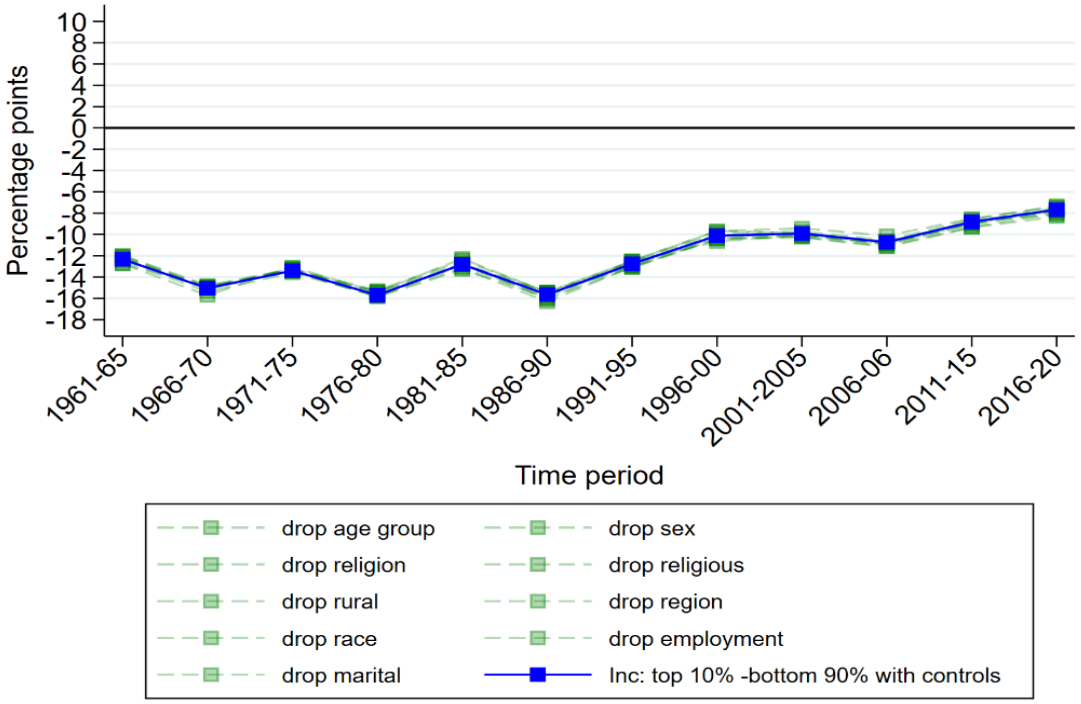
Note: This figure is to examine the robustness of Figure 1 in GMP, by showing the results weighted by the population of each country at that time. Population data is downloaded from the World Bank. Dashed lines show the weighted difference between (% of top 10% educated/earners voting left) and (% of bottom 90% educated/earners voting left) after controls. Solid lines show the original results from Figure 1 of GMP. Red lines show results for education and blue lines show results for income. The results in this figure are for the short-balanced panel of countries (the same 12 countries as in Figure 1 of GMP).

Figure 6 – Omitting control variables one by one: Education



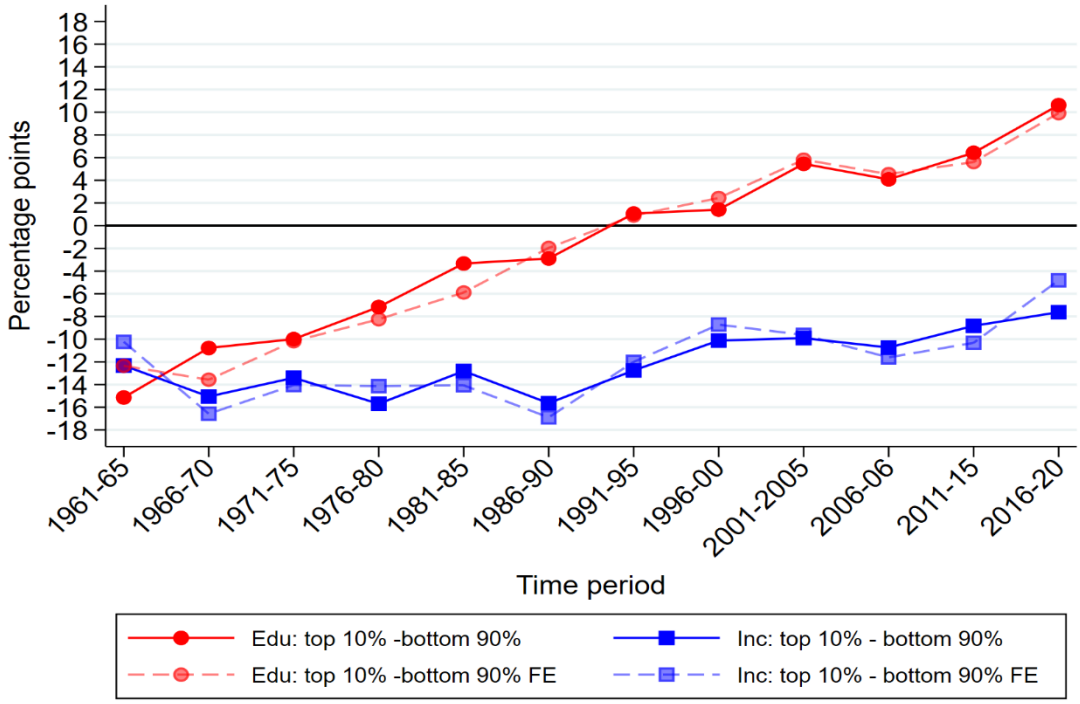
Note: This figure examines the robustness of Figure 1 in GMP, by showing the results of omitting control variables one by one. We exclude one of the controls listed for each estimation, including age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status. Dashed green lines show the difference between (% of top 10% educated voting left) and (% of bottom 90% educated voting left) after excluding one of the controls. Solid red and blue lines are original results in GMP Figure 1. The results in this figure are for the short-balanced panel of countries (the same 12 countries as in Figure 1 of GMP).

Figure 7 – Omitting control variables one by one: Income



Note: This figure examines the robustness of Figure 1 in GMP, by showing the results of omitting control variables one by one. We exclude one of the controls listed for each estimation, including age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status. Dashed green lines show the difference between (% of top 10% earners voting left) and (% of bottom 90% earners voting left) after excluding one of the controls. Solid red and blue lines are original results in GMP Figure 1. The results in this figure are for the short-balanced panel of countries (the same 12 countries as in Figure 1 of GMP).

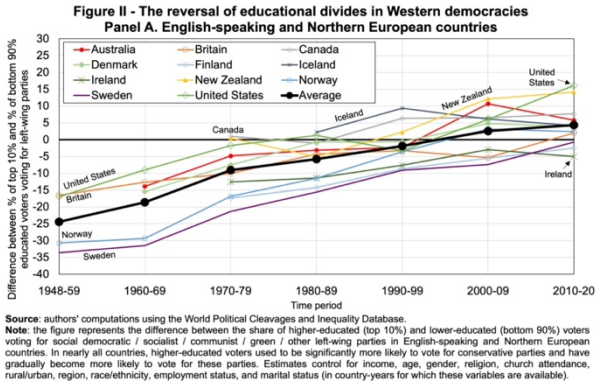
Figure 8 – Country fixed effects



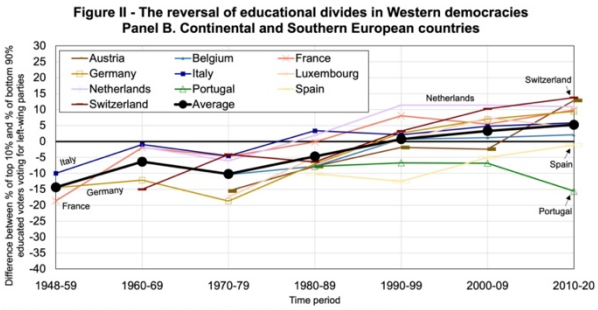
Note: This figure is to examine the robustness of Figure 1 in GMP, by modifying their specification by using a country fixed effect model. We control for country fixed effects by using pooled samples from each year. Dashed lines show the weighted difference between (% of top 10% educated/earners voting left) and (% of bottom 90% educated/earners voting left) after controls. Solid lines show the original results from Figure 1 of GMP. Red lines show results for education and blue lines show results for income. The results in this figure are for the short-balanced panel of countries (the same 12 countries as in Figure 1 of GMP).

Figure 9 – The reversal of educational divides in Western democracies: Original vs. ex-post recoding replication

a) Original

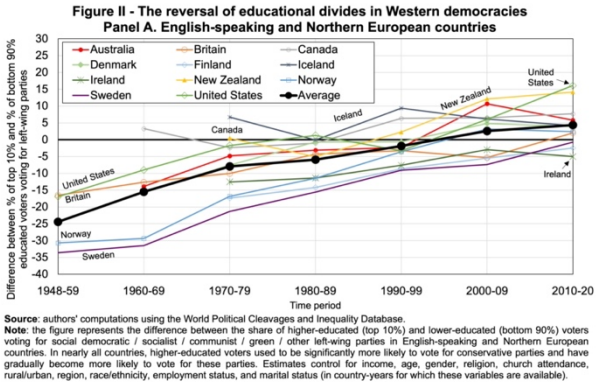


Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of higher-educated (top 10%) and lower-educated (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in English-speaking and Northern European countries. In nearly all countries, higher-educated voters used to be significantly more likely to vote for conservative parties and have gradually become more likely to vote for these parties. Estimates control for income, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).

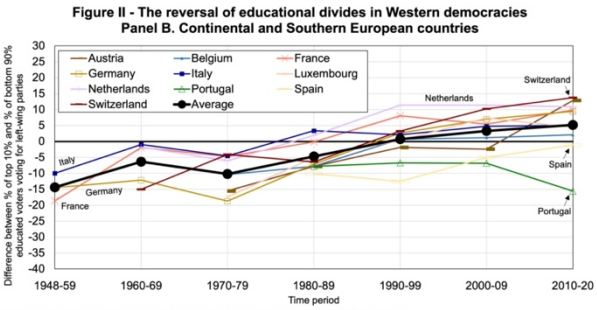


Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of higher-educated (top 10%) and lower-educated (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in Continental and Southern European countries. In nearly all countries, higher-educated voters used to be significantly more likely to vote for conservative parties and have gradually become more likely to vote for these parties. Estimates control for income, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).

b) Replication



Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of higher-educated (top 10%) and lower-educated (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in English-speaking and Northern European countries. In nearly all countries, higher-educated voters used to be significantly more likely to vote for conservative parties and have gradually become more likely to vote for these parties. Estimates control for income, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).

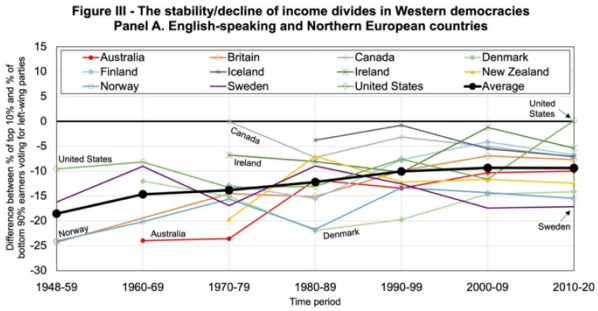


Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of higher-educated (top 10%) and lower-educated (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in Continental and Southern European countries. In nearly all countries, higher-educated voters used to be significantly more likely to vote for conservative parties and have gradually become more likely to vote for these parties. Estimates control for income, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).

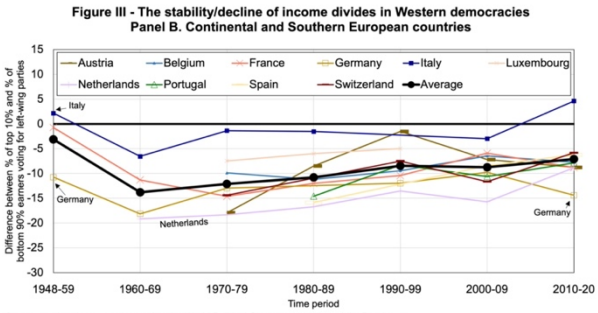
Note: This figure is Figure 2 in GMP. The left-hand panel shows the original figure from the main paper. The right-hand panel shows our ex-post recoding replication, which excludes the manual adjustments ("*Small fixes by variables*") in **results-paper-main.do**.

Figure 10 – The stability/decline of income divides in Western democracies: Original vs. ex-post recoding replication

a) Original

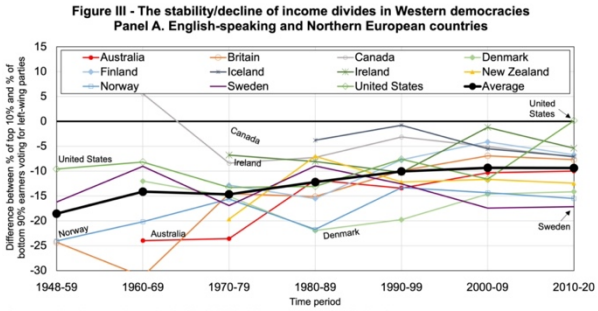


Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of high-income (top 10%) and low-income (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in English-speaking and Northern European countries. In all countries, top-income voters have remained significantly less likely to vote for these parties than low-income voters. Estimates control for education, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).

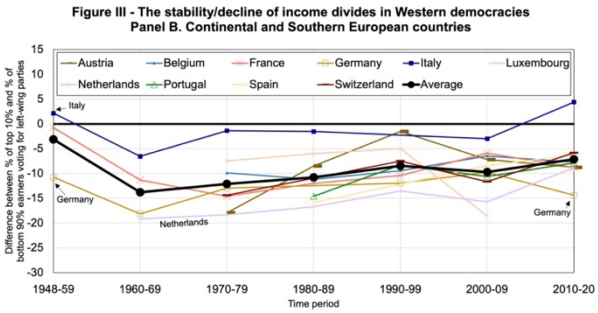


Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of high-income (top 10%) and low-income (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in Continental and Southern European countries. In all countries, top-income voters have remained significantly less likely to vote for these parties than low-income voters. Estimates control for education, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).

b) Replication



Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of high-income (top 10%) and low-income (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in English-speaking and Northern European countries. In all countries, top-income voters have remained significantly less likely to vote for these parties than low-income voters. Estimates control for education, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).



Source: authors' computations using the World Political Cleavages and Inequality Database.
Note: the figure represents the difference between the share of high-income (top 10%) and low-income (bottom 90%) voters voting for social democratic / socialist / communist / green / other left-wing parties in Continental and Southern European countries. In all countries, top-income voters have remained significantly less likely to vote for these parties than low-income voters. Estimates control for education, age, gender, religion, church attendance, rural/urban, region, race/ethnicity, employment status, and marital status (in country-years for which these variables are available).

Note: This figure is Figure 3 in GMP. The left-hand panel shows the original figure from the main paper. The right-hand panel shows our ex-post recoding replication, which excludes the manual adjustments ("*Small fixes by variables*") in **results-paper-main.do**.

Appendix

Appendix I – Code removed from do-files for the ex-post recoding replication.

For the main ex-post recoding replication, we excluded the following section (lines 167–204) in **results-paper-main.do**:

```
* aggregate some low-quality decades
replace year2 = 1970 if iso == "CA" & year2 == 1960
replace year2 = 1980 if iso == "IS" & year2 == 1970

* income outliers
drop if year2 == 2000 & iso == "LU" & strpos(var, "vote_inc") > 0
drop if year2 == 1960 & iso == "GB" & strpos(var, "vote_inc") > 0

* class outliers
drop if iso == "IT" & strpos(var, "vote_class") > 0 // this corresponds to
actual occupation
drop if iso == "FR" & year < 1980 & strpos(var, "vote_class") > 0
drop if !(inlist(iso, "US", "GB", "AU", "NZ", "NL", "LU", "SE") | inlist(iso,
"NO", "DK", "FR", "ES", "PT", "IT")) & strpos(var, "vote_class") > 0

* religion outliers
drop if iso == "IS" & strpos(var, "vote_religion") > 0 // only one data
points
drop if iso == "DK" & strpos(var, "vote_religion") > 0 // only data in the
1970s
drop if iso == "NO" & strpos(var, "vote_religion") > 0 // data on religiosity
but not on religion
drop if iso == "AT" & strpos(var, "vote_religion") > 0 // too inconsistent
drop if iso == "IE" & strpos(var, "vote_religion") > 0 // too inconsistent
drop if iso == "DE" & year2 == 1960 & strpos(var, "vote_religion") > 0 //
only partial data, no "no religion"
drop if iso == "PT" & year == 2002 & strpos(var, "vote_religion") > 0 // bad
data

* rural-urban outliers
drop if iso == "PT" & year2 == 1990 & strpos(var, "vote_rural") > 0
drop if iso == "FR" & year2 == 1950 & strpos(var, "vote_rural") > 0
drop if iso == "AT" & year2 == 1980 & strpos(var, "vote_rural") > 0
drop if iso == "FR" & year2 == 1950 & strpos(var, "vote_rural") > 0

* age outliers
drop if iso == "GB" & year2 == 1950 & strpos(var, "vote_age") > 0
drop if iso == "IT" & year2 == 1950 & strpos(var, "vote_age") > 0

* gender outliers
drop if iso == "LU" & year2 == 2010 & strpos(var, "vote_sex") > 0 // negative
but not significant
drop if iso == "CH" & year2 == 1960 & strpos(var, "vote_sex") > 0 // no data
on women

* union outliers
drop if iso == "IT" & year2 == 1950 & strpos(var, "vote_union") > 0.
```

For the additional ex-post recoding replication, we excluded line 65 in **results-all-1-religion.do**:

```
drop if iso == "CA" & region == "Quebec",
```

and lines 10–13, 35–38 and 45–46 in **results-all-1-race.do**:

```
replace race = region if iso == "IQ" // we take region as religion and
ethnicity are bad variables
replace race = religion if iso == "PH" // we take religion to better capture
the disprivileged Mindanao people
replace race = religion if iso == "FR"
replace race = ctrbirth if iso == "HK"

replace race = "Upper" if iso == "IN" & inlist(race, "Brahmins", "Other FC")
drop if iso == "IN" & !inlist(race, "Upper", "SC/ST").
```