Vote Your Region or Your Income?

Decomposing Variance in Redistributive Voting

Online Appendix

Contents

1 Data Descriptions and Summary Statistics **1**

2 Additional Detail, Variance Regression **6**

3 Robust to Alternative Estimators **7**

3.1 Restricted Estimation of Maximum Likelihood for a Heteroskedastic Regression . . 7

3.2 Variance Function Regression in a Bayesian Framework 9

4 Robust to the Alternative Geographic Units **12**

5 Robust to Restricted Models **14**

6 Robust to Party System Differences **16**

7 Robust to Additional Socio-Economic Controls **18**

8 Robust to Weighted Income Quintiles **20**

9 Controls for Cultural Factors **22**

10 Regional Productivity Profiles **24**

11 Sensitivity Analysis, Outliers Included **27**

1 Data Descriptions and Summary Statistics

We bring together data from various sources to examine our hypotheses. Our primary data source is the collection of post-electoral surveys compiled by the Comparative Study of Electoral Systems (Integrated Module Dataset, IMD I-IV). The CSES (IMD) includes a record of the survey respondent's party-list vote choice in the previous lower-house election (IMD3004_LH_PL). We then import each respondent's primary electoral district (A2027 from Module I, B2031 from Module II, C2031 from Module III, and D2032 from Module IV).

We merge the CSES data with two additional data sets: 1) the Comparative Manifesto Project (CMP), 2) the OECD regional economy statistics. Using country-election-year specific party statements on economic topics from the CMP, we calculate each party's relative position on redistribution (Benoit and Laver, 2007; Lowe et al., 2011). See the main text for details. We assign this party position value to the party of the CSES respondents' vote choice by cross-referencing the unique party ID codes (IMD3002_LH_PL). This conversion of vote choice allows us to measure a relative position scale of the voter's party choice on redistributive policies across countries and time. Next, we add the OECD regional economy statistics for gross domestic product per capita to match the electoral district. We assure the data for the OECD territorial level 3 (except Italy for level 2) matched the country's full list of electoral districts; thus, our sample is limited to nine Western European countries from 2000 to 2015 in which the electoral districts and TL3 (or TL2 for Italy) measure match.

The decision to match the electoral district to its regional information is theoretical and empirically motivated, but it substantially reduced our sample. In alternative analyses, featured below in Section 3, Table D, we include a sample with regional data at TL3 and TL2 OECD regions to sync with the 'region of residence' variable (A2019 from Module I, B2027 from Module II, C2027 from Module III, and D2028 from Module IV) for a larger sample of countries. While empirically less precise, we show these results because 1) individuals may perceive their "region of residence" as their region more readily than they perceive their electoral district to be a meaningful geographic unit; and 2) this result shows robustness at multiple levels of geographic aggregation, reducing

concerns with the modifiable areal unit problem.

Both individual income and regional productivity variables are coded as quintiles for cross-national comparison. The income information of the respondents from the CSES IMD dataset is available only in quintile measures for national household income (IMD2006). One downside of these data is that income ranges represent sample quintiles, not population quintiles. However, random sampling and large sample size should reduce bias concerns. We also convert the per capita measure of regional GDP into quintiles to directly and intuitively compare household income to regional productivity. Using a quintile also eases cross-national comparison–it eliminates currency unit differences and price changes over time and across countries.

The remaining control variables are available from the CSES IMD or the OECD regional statistics. We use conventional survey controls such as age, gender, education levels, marital status, and religiosity. We also add a control for the respondent's self-rated partisan orientation, measured on a zero to ten scale (with 10 being the most left-leaning). We also use the country's voter turnout rate as a group-level predictor to explain the variation in vote choices. We consider formulations of our household income variable in Section 8 and additional socio-economic controls in Section 7, along with several proxies for regional culturalism as explored in Section 9. The detailed regional productivity profiles are also provided in Section 10.

Table A. Variables, Summary Statistics, and Data Descriptions

Variables	Descriptions	Min	Max	Mean	Std. Dev	Data Sources
Dependent Variables						
Vote Choice (Continuous Measure)	Respondents' electoral choice (political party) in the lower house parliamentary election. All vote choices are transformed to a comparative position of the voter's preferred party on economic redistribution. A higher value is a vote for a party in favor of more redistribution. Calculation of parties' economic policy position is detailed in Equation (1) in the main text.	-2.38	4.93	1.32	1.10	Comparative Study of Electoral Studies (Integrated survey modules 1 through 4); Modified by the authors using Manifesto Project Dataset; Lowe et al. (2011).
Variance of Vote Choice (Continuous measure) Stage 2 Regression Residuals	(Logged) variance of the vote choice regression residuals. Calculation of variance function MLE code (in Stata) provided by Western and Bloom (2009) with authors' modification. The value range is based on Table 1, Model [1], reported in the main text.	-15.80	2.64	-1.41	2.19	
Vote Choice (Binary Measure)	Dummy equal to 1 if the partisan vote choice score based on Equation (1) is above country-election specific policy position mean, implying above mean support of economic redistribution. Based on Rueda and Stegmueller (2019).	0	1	0.47	0.50	
Variance of Vote Choice (Binary Measure) Stage 2 Regression Residuals	(Logged) variance of the binary vote choice regression (Pearson) residuals. Calculation of variance function MLE code (in Stata) provided by Western and Bloom (2009) with authors' modification. The value range is based on Table 1, Model [4], in the main text.	0.11	9.16	1.04	1.08	
Key Independent Variables						
Household Income Quintile	Household income status of the respondent relative to others within a country. Measured in household income quintile, with a value of 5 being the most affluent income group.	1	5	3.04	1.37	Organization for Economic Cooperation and Development (OECD) regional account. Geographic unit: TL3 (matched to electoral districts), TL3, TL2. Quintiles generated by the authors.
Regional Productivity Quintile (Electoral Districts at TL3)	Regional gross productivity quintile (1-5) of the residential area of the respondent. Measured by regional gross domestic product per capita. 1 is the lowest income quintile, 5 the highest.	1	5	3.34	1.46	
Robustness Checks						
Regional Income Quintile (TL3)		1	5	3.37	1.45	
Regional Income Quintile (TL2)		1	5	2.82	1.30	

Continued

Table A. Continued

Variables	Descriptions	Min	Max	Mean	Std. Dev	Data Sources
Controls						
Individual Level						
Self-placement	The respondent's self-placement on a 0 (right) -10 (left) scale.	0	10	5.01	2.48	Comparative Studies of Electoral Studies (Integrated survey modules 1 through 4); modified by the authors.
Age	Age of the respondent. Squared value is generated (Age Squared) to capture a quadratic function.	16	96	50.36	16.85	
Male	Gender of the respondent. Dummy value 1 = male, 0=female.	0	1	0.50	0.50	
Higher Education	Education of the respondent (2 levels) 0=none / primary / secondary education 1=post-secondary (non-university) /university education	0	1	0.34	0.47	
Married	Respondent's marital or civil union status. Dummy value: 1=married or living together as married, 0=others (widowed, divorced/separated, or single)	0	1	0.71	0.45	
Religiosity	Religious denomination dummy 1=religiosity (Catholic, protestant, orthodox/eastern catholic churches, other Christian, Jewish, Islam – Sunni, Islam – Other, Buddhism, Hinduism, Indigenous, or Ethnoreligions), 0=others (non-believers; agnostics; not specified; refused).	0	1	0.77	0.42	
Country-election-year Level						
Vote Turnout	Turnout as a percentage of the registered voters.	40.57	83.62	65.17	11.68	
Robustness Checks						
Individual Level						
Urbanization	Each survey respondent's residential categories: 1=rural area and village, 2=small or middle-sized town, 3=suburbs of large town or city, 4=large town or city. Used for the living expense weights.	1	4	2.54	1.15	
Eurosceptic	Net value of negative and positive references to the European Community/Union: per110-per108. This value is the party stance each individual survey respondent supports.	-8.18	35	-0.34	5.74	Manifesto Project Dataset
Nationalism	Net value of national way of life for positive and negative dimensions: per601-per602. This value is the party stance each individual survey respondent supports.	-1.49	20.43	1.44	2.49	

Continued

Table A. Continued

Variables	Descriptions	Min	Max	Mean	Std. Dev	Data Sources
Anti-multiculturalism	Net value of negative multiculturalism and positive multiculturalism: per608-per607. This value is the party stance each individual survey respondent supports.	-16.98	6.96	-0.21	2.10	Manifesto Project Dataset
Regional Level						
Population Growth	Growth/shrinking index of the total population (2001=100)	81.30	128.40	101.96	5.54	OECD Regional Statistics (Electoral districts at TL3, except Italy at TL2)
Population Density Growth	Growth/shrinking index of the population density per km2 (2001=100).	81.20	141.80	101.53	5.98	
Real GDP Per Capita Growth	Regional GDP per capita growth annual change (percent)	-17.80	21.62	2.76	4.29	
Labour Participation	The proportion of labour force over the population 15-64 (percent)	30.30	97.80	73.34	6.08	
Labour Utilization	The proportion of labour force over the total population (percent)	23.80	103.50	49.14	12.00	
GVA in Agriculture	Gross Value Added in Agriculture, Forestry, and Fishing (ISIC rev4). Millions USD, constant price, constant PPP based year 2005.	0.01	22.92	2.78	3.35	
Employment in Manufacturing	Share of employment in manufacturing (percent)	1.58	44.16	17.53	8.77	
Country-election-year Level						
ENP	Effective number of parties (ENP) on the votes level.	2.82	7.62	4.93	1.23	Comparative Political Dataset
Index of Absolute Proportionality	ENP on votes – ENP on seats.	0.42	2.25	0.85	0.46	
Index of Relative Proportionality	(ENP on votes – ENP on seats) / (ENP on votes).	0.08	0.33	0.17	0.07	

2 Additional Detail, Variance Regression

A potential issue with this two-step method is that even if the point estimates of β and λ were consistent, their standard errors would still be incorrect. This is because the standard errors for the estimates of λ fail to account for the uncertainty in β, and the least squares estimates of β are inefficient due to neglecting heteroskedasticity in the dependent variable (Western and Bloome, 2009).

Given that heteroskedastic errors are a main concern in our estimation, we following the recommendation of Aitkin (1987), iterating the two-step method by allowing for the maximum likelihood estimates under the assumption that $vote\ choice_i$ has a conditionally (and independently) normal distribution with the average vote choice, $\widehat{vote\ choice}_i$ and variance, σ_i^2. In this case, the log-likelihood function for the contribution of an individual survey response is defined as: $\mathscr{L}(\beta, \lambda : vote\ choice_i) = -1/2[\log(\sigma_i^2) + (vote\ choice_i - \widehat{vote\ choice}_i)^2/\sigma_i^2]$, where $(vote\ choice_i - \widehat{vote\ choice}_i)^2$ is the squared residual. Given this revision, our estimation technique evolves into a four-step approach based on Western and Bloome (2009) that extends the two-step method: 1) as an initial setup, we fit a least-squares regression of vote choice on the independent variables, yielding the predicted estimates $\hat{\beta}$ and storing the residuals $\hat{e}_i$. 2) Next, we fit a gamma regression with a log link function of $\hat{e}_i^2$ on the quintiles to predict the coefficient estimates $\hat{\lambda}$ and save the fitted values $\hat{\sigma}_i^2$. 3) Once these initial values are set, we perform a weighted linear regression of vote choice on the covariates, using the weights $\frac{1}{\hat{\sigma}_i^2}$ to inform how much each observation should influence the estimated linear regression coefficient, $\hat{\beta}$, which updates the residuals, $\hat{e}_i$, and re-evaluate the log-likelihood. 4) We then iterate Steps 2 and 3 until the estimates converge to values that maximize the likelihood function.

3 Robust to Alternative Estimators

3.1 Restricted Estimation of Maximum Likelihood for a Heteroskedastic Regression

The (unrestricted) maximum likelihood estimation of variance components can be biased in small samples. Take, for example, cross-sectional linear regression with homoskedastic errors, $MLE_{\hat{\sigma}^2} = N^{-1}\Sigma_i \hat{e}_i^2$. A less biased estimation can be derived by dividing with $N-K$ instead, where N is the number of observations and K is the number of parameters to be estimated. Such degree-of-freedom correction is offered by the (restricted) maximum likelihood estimation (Harville, 1977).

For robustness, we use an efficient REML algorithm implemented in an R package by (Smyth, 2002) to fit a heteroskedastic regression model. We find the results of our vote choice mean difference and the residual variance by household and regional productivity quintiles remain intact and robust. The variance function regression coefficient estimates (REML estimates of the mean differences and the residual variances) are slightly bigger than (unrestricted) ML estimates for the between-group variance regression, but smaller for the within-group variance regression. Please see Table B for details.

Table B. Restricted Maximum Likelihood Estimates of Vote Choice (Continuous Measure)

Variables	M[1]	M[2]	M[3]
First Stage: OLS Estimation			
Household Income Quintile	-0.035***	-0.036***	-0.037***
	[0.005]	[0.005]	[0.008]
Regional Productivity Quintile	0.031***		
	[0.004]		
2nd Quintile		0.112***	
		[0.022]	
3rd Quintile		0.079***	
		[0.023]	
4th Quintile		0.039*	
		[0.021]	
5th Quintile (Richest vs. Poorest)		0.175***	0.171***
		[0.020]	[0.008]
Self-placement: Right(0)-Left(10)	0.183***	0.183***	0.173***
	[0.003]	[0.003]	[0.004]
Age (16-95)	-0.002	-0.003	-0.001***
	[0.002]	[0.002]	[0.004]
Age Squared	0.000**	0.000**	0.000
	[0.000]	[0.000]	[0.000]
Male (1=Yes, 0=No)	-0.030**	-0.029**	-0.015
	[0.020]	[0.013]	[0.019]
Higher Education (1=Yes, 0=No)	0.134***	0.130***	0.133***
	[0.014]	[0.014]	[0.021]
Married (1=Yes, 0=No)	0.008	0.009	0.034
	[0.017]	[0.017]	[0.024]
Religiosity (1=Yes, 0=No)	0.005	0.000	0.004
	[0.015]	[0.015]	[0.023]
Constant	0.289***	0.311*	0.322***
	[0.062]	[0.004]	[0.089]
Second Stage: Variance Estimation			
Household Income Quintile	-0.046***	-0.046***	-0.038***
	[0.007]	[0.067]	[0.011]
Regional Productivity Quintile	-0.011*		
	[0.007]		
2nd Quintile		-0.099***	
		[0.035]	
3rd Quintile		-0.164***	
		[0.036]	
4th Quintile		-0.016	
		[0.033]	
5th Quintile (Richest vs. Poorest)		-0.105***	-0.116***
		[0.031]	[0.031]
Vote Turnout (%)	-0.036***	-0.036***	-0.037***
	[0.001]	[0.001]	[0.001]
Constant	2.418***	2.473***	2.551***
	[0.067]	[0.067]	[0.094]
Log Likelihood	-28,523	-28,493	-13,318
AIC	59,051	59,050	27,626
BIC	59,162	59,209	27,727
Sample	Full	Full	Subsample
Observations	20,595	20,595	9,646

Notes: Two-tailed test significance at *p<0.1, **p<0.05, ***p<0.01. Continuous dependent variable. P-values are in brackets. Calculated using REML algorithm (heteroskedastic regression) in R package "remlscore" to fit a heteroskedastic regression model (Smyth, 2002).

3.2 Variance Function Regression in a Bayesian Framework

As an additional check against small sample bias, we report Bayesian estimates of the residual variance coefficient (λ) expressed in Equation (1). Small samples can make the λ variance coefficient estimates skewed and thus normal distribution-based inference may be inaccurate. Our estimates (shown in Table C and Figure A) show our results are robust to this technique. Household income and regional productivity quintiles remain statistically meaningful (see Gibbs sampling percentage ranges shown in Table C) and in the same direction as in the results in the main text, in both the between-group variance and within-group variance.

Our Bayesian modeling for 'causal heterogeneity' borrows from Western and Bloome's programming in BUGS (Bayesian Inference Using Gibbs Sampling). Our empirical model integrates both a prior distribution for the mean difference coefficients (β) and a hierarchical prior for the variance coefficients (λ) to the normal likelihood:

$$\begin{aligned}
Vote\ Choice_i &\sim N(\widehat{Vote\ Choice}_i, \sigma_i^2) \\
\widehat{Vote\ Choice}_i &= x_i'\beta \\
log\sigma_i^2 &= z_i'\lambda \\
with\ prior\ distributions\ \beta &\sim N(prior\ mean\ vectors\ [b], prior\ covariance\ matrix\ [V]) \\
\lambda &\sim (prior\ mean\ vectors\ [g], covariance\ matrix\ [U]) \\
U &\sim Gamma^{-1}(hyperparameters\ u_0, hyperparameters\ u_1)
\end{aligned} \tag{1}$$

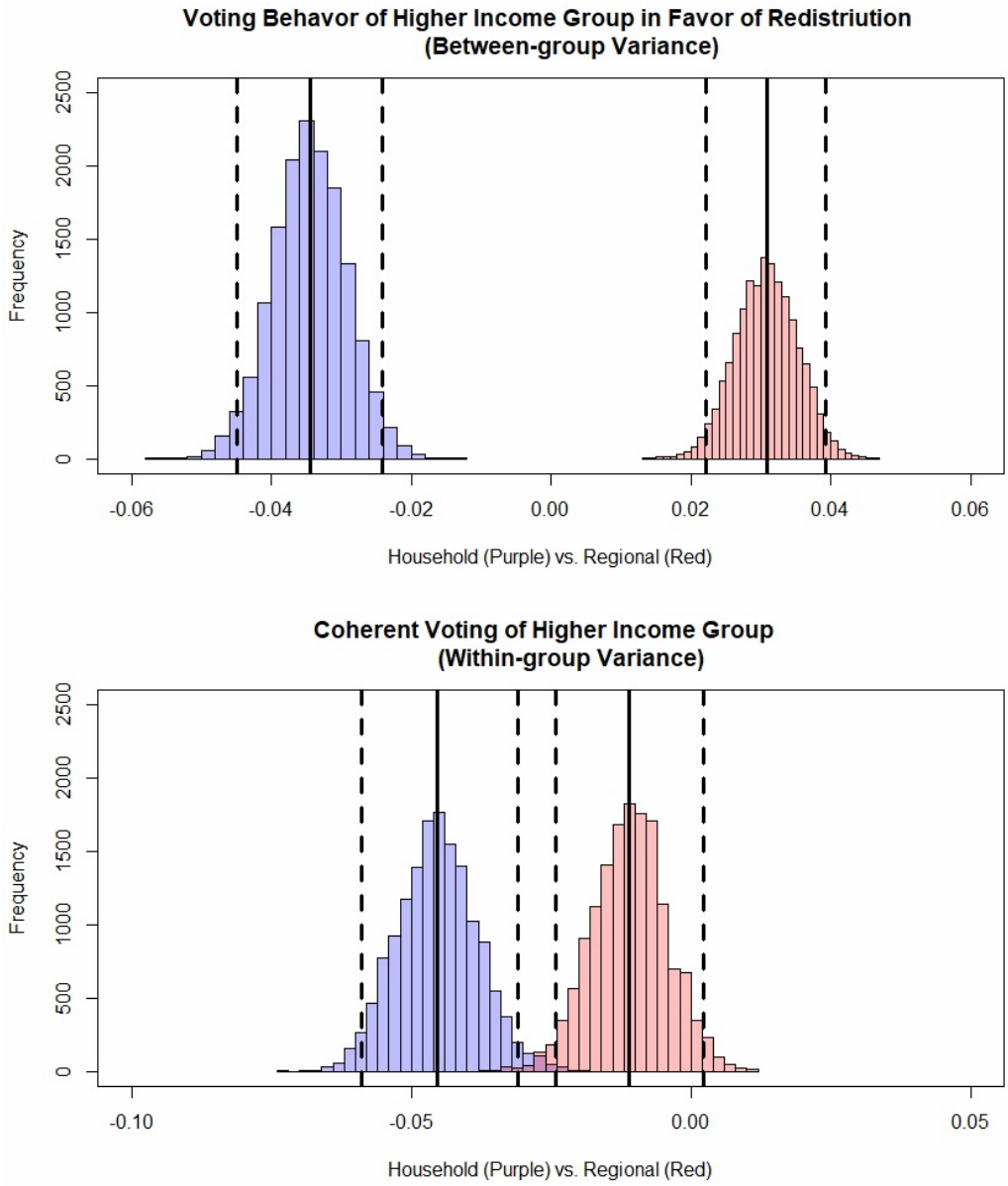
where the dependent variable, $votechoice_i$ with predictor x_i for the mean difference and z_i for the residual variance. To run this Bayesian estimation, we ensure that the non-informative priors (b, g) are all set to zero. For prior covariance (V), we began with a substantively large variance (10^6). The covariance matrix U also comes with the prior variance that assumes an inverse Gamma distribution with $u_0 = 0.001$ and $u_1 = 0.001$. We fit the model using WinBUGS software.

Table C. Bayesian Estimates of Vote Choice (Continuous Measure)

Variables: M[4]	Mean	Std.Dev.	MC Error	2.5%	Median	97.5%
First Stage: OLS Estimation						
Household Income Quintile	-0.035	0.005	4.209E-5	-0.045	-0.035	-0.024
Regional Productivity Quintile	0.031	0.004	3.64E-5	0.022	0.031	0.039
Self-placement (Right-Left)	0.183	0.003	2.35E-5	0.178	0.183	0.188
Age	-0.002	0.002	2.01E-5	-0.007	-0.002	0.002
Age Squared	4.89E-5	2.42E-5	1.94E-7	1.18E-6	4.89E-5	9.61E-5
Male	-0.030	0.013	1.02E-4	-0.056	-0.030	-0.004
Higher Education	0.135	0.014	1.38E-4	0.106	0.135	0.163
Married	0.008	0.017	1.23E-4	-0.025	0.008	0.041
Religiosity	0.005	0.015	1.31E-4	-0.025	0.005	0.036
Constant	0.289	0.063	4.96E-4	0.166	0.289	0.411
Second Stage: Variance Estimation						
Household Income Quintile	-0.045	0.007	2.61E-4	-0.059	-0.045	-0.030
Regional Productivity Quintile	-0.011	0.007	2.52E-4	-0.024	-0.011	0.002
Vote Turnout	-0.035	7.72E-4	4.44E-5	-0.037	-0.035	-0.034
Constant	2.405	0.061	0.004	2.281	2.410	2.515

Notes: The list of variables is all drawn in Table 1, Model 1 (Full) without fixed effects, available in the main text. The estimated values are Bayesian inferences using Markov Chain Monte Carlo (MCMC) methods via the WinBUGS that allow us to draw Bayesian inferences using Gibbs sampling. We use thinning (collecting every second part of randomly generated numbers) to remove auto-correlation. Out of 20,000 iterations, the first 10,000 numbers were burnt. This burning process also ensures that iterations are launched from three separate initial values and they converge.

Figure A. Histograms of Posterior Distributions



4 Robust to the Alternative Geographic Units

As described in the main text, we have carefully considered the theoretical geographic unit of analysis for a vote choice–the electoral district–and restricted the sample in our main analysis to countries in which we have regional productivity estimates at the district level. However, we cannot be certain that voters are aware of the theoretical unit–they might perceive other unit levels to be their region. Moreover, given statistical unpredictability associated with modifiable areal unit problems (Lee and Rogers, 2019; Soifer, 2019), we also test our results at different levels of geography (TL2 and TL3), with a larger sample in which the territorial unit does not in every case match the electoral district.

Table D shows results matching OECD regional productivity statistics with the CSES data for available country-election years: Australia (TL2), Austria (TL3 or TL2), Canada (TL2), Czech Republic (TL3), Estonia (TL3), Finland (TL3), France (TL2), Germany (TL3 or TL2), Great Britain (TL2), Greece (TL3), Hungary (TL3), Italy (TL3 or TL2), Norway (TL3 or TL2), Poland(TL3 or TL2), Spain (TL3 or TL2), Slovakia (TL3), Slovenia (TL3), Switzerland (TL3), and the United States (TL2). The resulting samples are larger than those featured in our main text.

Given the added number of country-election years extending to TL3 and TL2, we report the result in Table D for the binary dependent variable. This binary measure controls within its structure for country-election year effects, which reduces the risk of over-fitting and thus amplifies heteroskedasticity in the model. The statistical variance of the vote choice residuals tends to be smaller as the level of region income goes up (from TL3 to TL2).

Table D. Robust to Different Geographic Aggregations, TL2 and TL3

	M[5] TL2	M[6] TL2	M[7] TL2	M[8] TL3	M[9] TL3	M[10] TL3
First Stage: Logistic Estimation						
Household Income Quintile	-0.060***	-0.062***	-0.070***	-0.017†	-0.017†	-0.027*
	[0.011]	[0.011]	[0.019]	[0.011]	[0.011]	[0.016]
Regional Productivity Quintile	0.022**			0.010		
	[0.011]			[0.009]		
2nd Quintile		-0.143***			0.075†	
		[0.041]			[0.048]	
3rd Quintile		-0.095**			0.141***	
		[0.044]			[0.049]	
4th Quintile		-0.015			0.012	
		[0.043]			[0.045]	
5th Quintile (Richest vs. Poorest)		0.052	0.040		0.087**	0.082*
		[0.050]	[0.049]		[0.043]	[0.043]
Self-placement: Right(0)-Left(10)	0.310***	0.309***	0.278***	0.332***	0.334***	0.325***
	[0.007]	[0.007]	[0.011]	[0.006]	[0.006]	[0.009]
Age (16-95)	-0.004***	-0.004***	0.014†	0.019***	0.019***	0.017**
	[0.001]	[0.001]	[0.009]	[0.005]	[0.005]	[0.008]
Age Squared	0.000	0.000	-0.000*	-0.000***	-0.000***	-0.000**
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Male (1=Yes, 0=No)	-0.067**	-0.068**	-0.127***	-0.069**	-0.068**	-0.078*
	[0.027]	[0.027]	[0.047]	[0.028]	[0.028]	[0.040]
Higher Education (1=Yes 0=No)	-0.073**	-0.057**	-0.172***	-0.217***	-0.216***	-0.171***
	[0.028]	[0.029]	[0.050]	[0.029]	[0.000]	[0.043]
Married (1=Yes, 0=No)	-0.125***	-0.123***	-0.194***	-0.118***	-0.119***	-0.065
	[0.033]	[0.033]	[0.062]	[0.035]	[0.029]	[0.051]]
Religiosity (1=Yes, 0=No)	-0.209***	-0.206***	-0.074	0.033	0.029	0.016
	[0.031]	[0.031]	[0.054]	[0.032]	[0.032]	[0.047]
Constant	-1.140***	-1.028***	-1.293***	-2.084***	-2.210***	-1.990***
	[0.081]	[0.080]	[0.221]	[0.132]	[0.132]	[0.189]
Second Stage: Variance Estimation						
Household Income Quintile	-0.050***	-0.049***	-0.046***	-0.026***	-0.026***	-0.033***
	[0.004]	[0.004]	[0.007]	[0.005]	[0.005]	[0.006]
Regional Productivity Quintile	-0.008*			-0.015***		
	[0.005]			[0.004]		
2nd Quintile		0.026			-0.108***	
		[0.018]			[0.022]	
3rd Quintile		0.004			-0.058***	
		[0.019]			[0.022]	
4th Quintile		-0.032*			-0.097***	
		[0.019]			[0.021]	
5th Quintile (Richest vs. Poorest)		-0.002	-0.017		-0.087***	-0.088***
		[0.022]	[0.019]		[0.019]	[0.019]
Vote Turnout (%)	-0.013***	-0.013***	-0.013***	-0.012***	-0.013***	-0.014***
	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]	[0.001]
Constant	1.190***	1.142***	1.142***	0.948***	1.001***	1.111***
	[0.038]	[0.036]	[0.054]	[0.043]	[0.043]	[0.058]
Log Likelihood	-25,772	-25,721	-8,364	-25,530	-25,555	-11,823
AIC	51,552	51,457	16,735	51,068	51,124	23,654
BIC	51,585	51,513	16,763	51,100	51,181	23,864
Sample	Full	Full	Subsample	Full	Full	Subsample
Number of Countries (Election Years)	11(13)	11(13)	11(13)	14(16)	14(16)	14(16)
Observations	24,841	24,841	8,124	24,869	24,869	11,502

Notes: Two-tailed tests significant at †p<0.15, *p<0.1, **p<0.05, ***p<0.01. Binary dependent variable. p-values are in brackets. TL3 regions are nested within TL2 regions.

5 Robust to Restricted Models

The respondents' self-placement of political ideology can be endogenously correlated with education and income levels. We checked for the Pearson pairwise correlations across income, education level, and self-placement: -0.016 for corr (right-left self-placement, education level), -0.063 for corr (household income quintile, right-left self-placement), -0.054 for corr (regional productivity quintile, right-left self-placement), 0.177 for corr (education level, household income quintile), 0.099 for corr (education level, regional income quintile), and 0.06 for corr (household income quintile, regional income quintile). Although these values are not high, we rerun models excluding the self-placement and education level variables. Table E indicates that dropping self-placement and education variables does not alter the main finding significantly (See stable p-values in bracket, Table E).

We also test (but not report here) against the non-linear structure of the respondent's right-left ideological scale differences in economic dimension – vote preferences for the party advocating redistributive policies (Lachat, 2018). When adding a square term of the right-left self-placement measure to the existing model (Model 11), we find it positive and significant without altering our income variable estimates. This curvilinear effect also reveals that ideological self-placement is not endogenously correlated with income or education levels.

Table E. Robust to Restricted Models

Variables	M[11]	M[12]	M[13]
First Stage: OLS Estimation			
Household Income Quintile	-0.065***	-0.067***	-0.065***
	[0.003]	[0.003]	[0.000]
Regional Productivity Quintile	0.006**		
	[0.003]		
2nd Quintile		0.046***	
		[0.015]	
3rd Quintile		0.029*	
		[0.015]	
4th Quintile		0.037**	
		[0.015]	
5th Quintile (Richest vs. Poorest)		0.039***	0.027**
		[0.013]	[0.043]
Age (16-95)	0.014***	0.014***	0.017***
	[0.002]	[0.001]	[0.000]
Age Squared	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]
Male (1=Yes, 0=No)	-0.022**	-0.021**	-0.039***
	[0.009]	[0.009]	[0.001]
Married (1=Yes, 0=No)	-0.035***	-0.036***	-0.039***
	[0.001]	[0.011]	[0.010]
Religiosity (1=Yes, 0=No)	-0.299***	-0.297***	-0.299***
	[0.012]	[0.012]	[0.000]
Constant	1.377***	1.349***	1.329***
	[0.053]	[0.054]	[0.000]
Second Stage: Variance Estimation			
Household Income Quintile	-0.034***	-0.032***	-0.037***
	[0.008]	[0.008]	[0.002]
Regional Productivity Quintile	-0.021***		
	[0.007]		
2nd Quintile		-0.049	
		[0.037]	
3rd Quintile		-0.140***	
		[0.037]	
4th Quintile		0.031	
		[0.034]	
5th Quintile (Richest vs. Poorest)		-0.137***	-0.230***
		[0.033]	[0.000]
Vote Turnout (%)	-0.038***	-0.038***	-0.040***
	[0.001]	[0.001]	[0.000]
Constant	2.600***	2.596***	2.802***
	[0.070]	[0.069]	[0.000]
Log Likelihood	-20,682	-14,497	-9,353
AIC	41,374	29,007	18,714
BIC	41,406	29,062	18,743
First Stage County Election Year FE	Yes	Yes	Yes
Sample	Full	Full	Full
Observations	21,662	21,662	10,154

Notes: Two-tailed tests significant at *p<0.1, **p<0.05, ***p<0.01. Continuous dependent variable. P-values are in brackets. CSES samples are selected for electoral district matching cases only.

6 Robust to Party System Differences

A concern with our analysis could be that: 1) depending on the number of parties, voters may have a clearer vote choice on redistributive politics and be able to combine preferences across different policy dimensions better in multi-party systems; 2) the variance of vote choice would necessarily be higher in multi-party systems. We control for this possibility in the main analysis using the binary dependent variable approach from (Rueda and Stegmueller, 2019). This section also includes explicit controls for the "effective" number of political parties in a country's party system (Laakso and Taagepera, 1979).

We use the ENP (the effective number of parties) measure at the vote level from the Comparative Political Dataset (Armingeon et al., 2020). We also include a control for the absolute difference between the ENP on the votes (in elections) and the ENP on the seats (Table F, M[15]) as a measure of strategic voting incentives given the size of the ENP in elections is usually bigger than that of the ENP in parliaments (Laakso and Taagepera, 1979). We expect that the lower the strategic voting incentive (higher number of parties), the smaller the vote choice variability as voters are less cross-pressured with more party choices. We also include a measure of relative proportionality (Table F, M[16]) weighted by the total number of parties on the vote level.

In Table F, both household and personal income quintile estimates remain robust across the three different models (M[14] to M[16]). The ENP indicator coefficient estimates are all in the anticipated direction and statistically significant (at a p-value of 0.000 for M[14] through M[16]).

Table F. Robust to Party System Controls

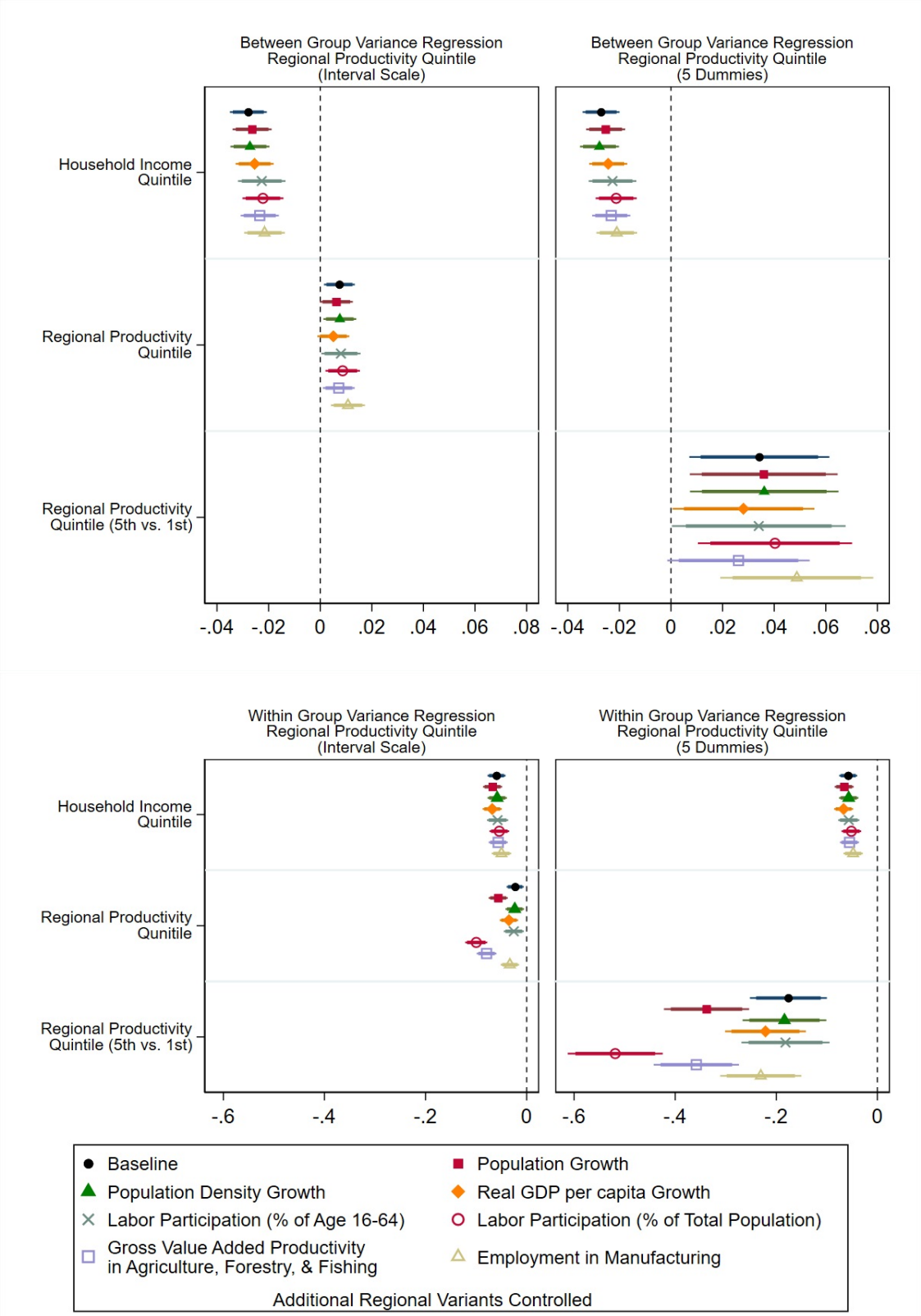
Variables	M[14]	M[15]	M[16]
First Stage: OLS Estimation			
Household Income Quintile	-0.027***	-0.030***	-0.029***
	[0.004]	[0.004]	[0.004]
Regional Productivity Quintile	0.010***	0.008***	0.008***
	[0.003]	[0.003]	[0.003]
Self-placement: Right(0)-Left(10)	0.179***	0.190***	0.190***
	[0.002]	[0.002]	[0.002]
Age (16-95)	0.006***	0.005***	0.005***
	[0.002]	[0.002]	[0.002]
Age Squared	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]
Male (1=Yes, 0=No)	0.006	0.010	0.011
	[0.009]	[0.009]	[0.009]
Higher Education (1=Yes, 0=No)	-0.042***	-0.049***	-0.050***
	[0.011]	[0.011]	[0.011]
Married (1=Yes, 0=No)	-0.040***	-0.026**	-0.027**
	[0.011]	[0.011]	[0.011]
Religiosity (1=Yes, 0=No)	-0.117***	-0.099***	-0.103***
	[0.011]	[0.011]	[0.011]
Constant	0.411***	0.354***	0.354***
	[0.055]	[0.056]	[0.056]
Second Stage: Variance Estimation			
Household Income Quintile	-0.059***	-0.055***	-0.056***
	[0.010]	[0.010]	[0.010]
Regional Productivity Quintile	-0.017*	-0.020**	-0.018**
	[0.009]	[0.009]	[0.009]
Vote Turnout (%)	-0.035***	-0.041***	-0.043***
	[0.001]	[0.001]	[0.001]
Effective Number of Parties (ENP) on *Votes*	0.219***	0.316***	0.150***
	[0.013]	[0.014]	[0.013]
Index of Absolute Proportionality			
ENP on *Votes* - ENP on *Seats*		-0.850***	
		[0.034]	
Index of Relative Proportionality			
(ENP on *Votes*- ENP on *Seats*) / (ENP on *Votes*)			-4.453***
			[0.194]
Constant	1.140***	1.689***	2.704***
	(0.133)	[0.133]	[0.146]
Log Likelihood	-14,161	-13,205	-13,323
AIC	28,331	26,423	26,659
BIC	28,370	26,470	26,707
First Stage County Election Year FE	Yes	Yes	Yes
Sample	Full	Full	Full
Observations	20,595	20,595	20,595

Notes: Two-tailed tests significant at *p<0.1, **p<0.05, ***p<0.01. Continuous dependent variable. P-values are in brackets.

7 Robust to Additional Socio-Economic Controls

We include results with a battery of additional controls for region-specific demographic and market pressures. In Figure B, we summarize the stability-regarding results for our main explanatory variables, given added socio-economic controls, including population growth rate, population density growth, real GDP per capita growth rate, labor market participation rate (as a share of the age between 16-64), labor utilization rate (as a share of the total population), industry-based productivity (esp., in agriculture, forestry, and fishing), and employment in manufacturing. In all cases, our main results remain intact.

Figure B. Income Effects on the Vote Choice (Continuous Measure) with Additional Controls



8 Robust to Weighted Income Quintiles

The household income quintile measures used in our analysis are calculated based on the national income distribution by the CSES. We do not have the underlying income information to calculate region-specific income distributions. While the national income distribution may be the theoretically relevant metric, a regional metric might be preferred given differences in cost of living or the importance of (locally) relative assessments of income (Ogorzalek et al., 2019). Those falling into the same national income quintile bracket may experience different perceived levels of income based on local prices and the local income distribution. Of course, controlling for the regional productivity quintile (one of our main explanatory variables included in all analyses) should account in large part for regional cost of living, with higher regional productivity quintiles generally having higher cost of living.

Given constraints with the CSES, we do our best to consider the cost of living differences and regional productivity conditions but weigh our individual observations by urban and rural residence. From a statistical point of view, living costs should serve as an inverse weight to how we think about the economic status of income quintile groups. In other words, those in the high-tier income quintile with high costs of living should receive less analytical weight (because they are not as rich in relative terms) than those in the same high-tier income quintile with low costs of living (Ogorzalek, Piston, and Puig, 2019). Accordingly, we divided income quintile groups by the degree of urbanization, our best proxy for the cost of living in the dataset. In this scheme, the largest analytical weight should go to those in the highest quintile group whose residential area is a rural base. This revised analytical setup makes them “more affluent" than those in more urban areas in the same quintile, in relative terms.

Table G presents the variance function regression results with our key findings intact. We find that the predicted voting choice behavior of high-income tiers with low costs of living (rural area residents) have larger coefficient values, both for vote choice and vote variance. This is what our theory predicts, especially their high degree of consistency in vote choice.

Table G. Effects of Costs-of-living Adjusted Quintiles on the Vote Choice

	M[17] More Weight on Rural Area Residents	M[18] More Weight on Rural Area or Middle-sized Town Residents	M[19] More Weight on Rural Area or Middle-sized Town or Suburban Residents
First Stage: OLS Estimation			
Household Income Quintile (Weighted)	-0.033***	-0.029***	-0.027***
	[0.005]	[0.004]	[0.004]
Regional Productivity Quintile (Weighted)	0.009**	0.007*	0.006*
	[0.004]	[0.004]	[0.004]
Self-placement: Right(0)-Left(10)	0.195***	0.195***	0.195***
	[0.002]	[0.002]	[0.002]
Age 16-95)	0.006***	0.006***	0.006***
	[0.002]	[0.002]	[0.002]
Age Squared	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]
Male (1=Yes, 0=No)	0.007	0.009	0.006
	[0.009]	[0.009]	[0.009]
Higher Education (1=Yes, 0=No)	-0.054***	-0.054***	-0.055***
	[0.011]	[0.011]	[0.011]
Married (1=Yes, 0=No)	-0.037***	-0.038***	-0.036***
	[0.012]	[0.012]	[0.012]
Religiosity (1=Yes, 0=No)	-0.120***	-0.119***	-0.122***
	[0.012]	[0.012]	[0.012]
Constant	0.355***	0.362***	0.352***
	[0.059]	[0.059]	[0.059]
Second Stage: Variance Estimation			
Household Income Quintile (Weighted)	-0.088***	-0.089***	-0.073***
	[0.012]	[0.010]	[0.009]
Regional Productivity Quintile (Weighted)	-0.042***	-0.075***	-0.044***
	[0.013]	[0.011]	[0.010]
Vote Turnout (%)	-0.045***	-0.044***	-0.045***
	[0.001]	[0.001]	[0.001]
Constant	2.871***	2.969***	2.964***
	[0.078]	[0.075]	[0.081]
Log Likelihood	-14,736	-14,605	-14,704
AIC	29,480	29,218	29,416
BIC	29,511	29,249	29,447
First Stage County Election Year FE	Yes	Yes	Yes
Sample	Full	Full	Full
Observations	19,995	19,995	19,995

Notes: Two-tailed tests significant *p<0.1, **p<0.05, ***p<0.01. Continuous dependent variable. P-values are in brackets. CSES samples are categorized into 4 urbanization groups: (1) rural areas, (2) middle-sized towns, (3) city suburbs, and (4) large cities. M[17] is based on the comparison groups with a revised assigned numeric value (rural area=1 vs. the rest=2); for M[18] (rural area or middle town = 1 vs. the rest = 2); and M[19] (rural area or middle town or suburb = 1 vs. large city = 2). Income quintiles of households and regions are then divided by their urbanization group level. Thus, the logic is that the larger the value in the denominator, the less the analytical weight to be assigned for weighted income quintiles.

9 Controls for Cultural Factors

Support for or against redistribution could also be driven by other cultural factors such as Euroscepticism, nationalism, and anti-multiculturalism. We include these additional measures to ensure our economic-focused results do not suffer from extreme omitted variable bias.

Drawing upon three indicators from the Manifesto Project dataset, we extract three cultural variables. We generate the Eurocepticism variable using the net value of the voter's party's policy stance on negative European Union/community mentions versus positive mentions. We also use the net value of positive mentions of nationalism (e.g. support for the established national ideas, pride of citizenship, and appeal to patriotism) and negative mentions of nationalism (e.g., opposition to the existing national state, national pride, national ideas). We additionally create the anti-multiculturalism variable based on the net value of negative multiculturalism (e.g., support for the enforcement or encouragement of cultural integration and homogeneity in society) and positive multiculturalism (e.g., favorable mentions of cultural diversity and plurality within the country). These cultural proxies are added to our model to capture non-economic reasons why some individuals are less supportive of redistribution. We offer the summary of these extended model estimates in OA Table I, wherein no substantial evidence was found to alter our previous findings.

Table H. Robust Checks to Sociocultural Controls

	M[20] Continuous Measure	M[21] Continuous Measure	M[22] Continuous Measure	M[23] Binary Measure	M[24] Binary Measure	M[25] Binary Measure
First Stage Estimation:		*OLS regression*			*logistic regression*	
Household Income Quintile	-0.035***	-0.029***	-0.031***	-0.055***	-0.054***	-0.044**
	[0.000]	[0.000]	[0.000]	[0.002]	[0.002]	[0.013]
Regional Productivity Quintile						
5th Quintile vs. 1st Quintile)	0.023*	0.024*	0.026*	0.211***	0.156***	0.192***
	[0.090]	[0.081]	[0.063]	[0.000]	[0.001]	[0.000]
Self-placement: Right(0)-Left(10)	0.161***	0.158***	0.172*	0.305***	0.266***	0.317***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Age (16-95)	0.008***	0.008***	0.010*	0.024***	0.025***	0.025***
	[0.000]	[0.001]	[0.000]	[0.004]	[0.003]	[0.003]
Age Squared	-0.000***	-0.000***	-0.000*	-0.000***	-0.000***	-0.000**
	[0.000]	[0.002]	[0.000]	[0.003]	[0.003]	[0.001]
Male (1=Yes, 0=No)	-0.008	0.009	-0.003	-0.016	-0.001	-0.018
	[0.496]	[0.471]	[0.787]	[0.720]	[0.980]	[0.691]
Higher Education (1=Yes 0=No)	-0.058***	-0.064***	-0.045***	-0.124***	-0.140***	-0.204***
	[0.000]	[0.000]	[0.004]	[0.010]	[0.003]	[0.000]
Married (1=Yes, 0=No)	-0.015	-0.022	-0.029*	-0.081	-0.086	-0.091***
	[0.343]	[0.165]	[0.074]	[0.147]	[0.124]	[0.103]
Religiosity (1=Yes, 0=No)	-0.111***	-0.113***	-0.119***	0.021	0.099*	0.027
	[0.000]	[0.000]	[0.000]	[0.699]	[0.065]	[0.605]
Euroskeptism	-0.062***			-0.041***		
	[0.000]			[0.000]		
Nationalism		-0.111***			-0.206***	
		[0.000]			[0.000]	
Anti-multiculturalism			-0.023***			-0.051***
			[0.000]			[0.000]
Constant	0.159**	0.786***	0.443***	-2.013***	-1.595***	-1.996***
	[0.039]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Second Stage: Variance Estimation		*gamma regression*			*gamma regression*	
Household Income Quintile	-0.054***	-0.067***	-0.062***	-0.032***	-0.043***	-0.034***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.012]	[0.000]
Regional Productivity Quintile						
5th Quintile vs. 1st Quintile)	-0.196***	-0.161***	-0.189***	-0.071***	-0.082*	-0.053***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.094]	[0.010]
Vote Turnout (%)	-0.042***	-0.051***	-0.049***	-0.013	-0.026***	-0.015***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Constant	2.652***	3.243***	3.176***	0.985***	2.047***	1.120***
[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Log Likelihood	-5,829	-5,950	-6,524	-9,825	-10,943	-9,924
AIC	11,666	11,908	13,057	19,659	21,893	19,855
BIC	11,695	11,937	13,086	19,688	21,922	19,884
First Stage County Election Year FE	Yes/Yes	Yes/Yes	Yes/Yes	Country-Election Year Embedded		
Sample	Subsample	Subsample	Subsample	Subsample	Subsample	Subsample
Number of Countries (Election Years)	11(13)	11(13)	11(13)	14(16)	14(16)	14(16)
Observations	9,646	9,646	9,646	9,646	9,646	9,646

Notes: Two-tailed tests significant at *p<0.1, **p<0.05, ***p<0.01. P-values are in brackets.

10 Regional Productivity Profiles

Our measure of regional productivity is regional GDP per capita at the territorial level 3 (TL3 equivalent to NUTS3) regions, except Italy for level 2. Figure C offers a map of regional GDP per capita in quintile distribution using 2014 observations for our sample across 9 countries, roughly showing the end of our sample period.

Additionally, about the CSES survey responses available to our analysis, Table H displays sectoral differences in regional productivity, measured in Gross Value Added (GVA) productivity, available from the Annual Regional Database of the European Commission (ARDECO) database. In Table H we use reds for service sectors and blues for non-service sectors. As can be seen, the 5th quintile of regional productivity in GVA displays the expansion of service sectors while non-service sectors are a smaller percentage of productivity compared to lower quintiles.

Figure C. Places and Economic Productivity

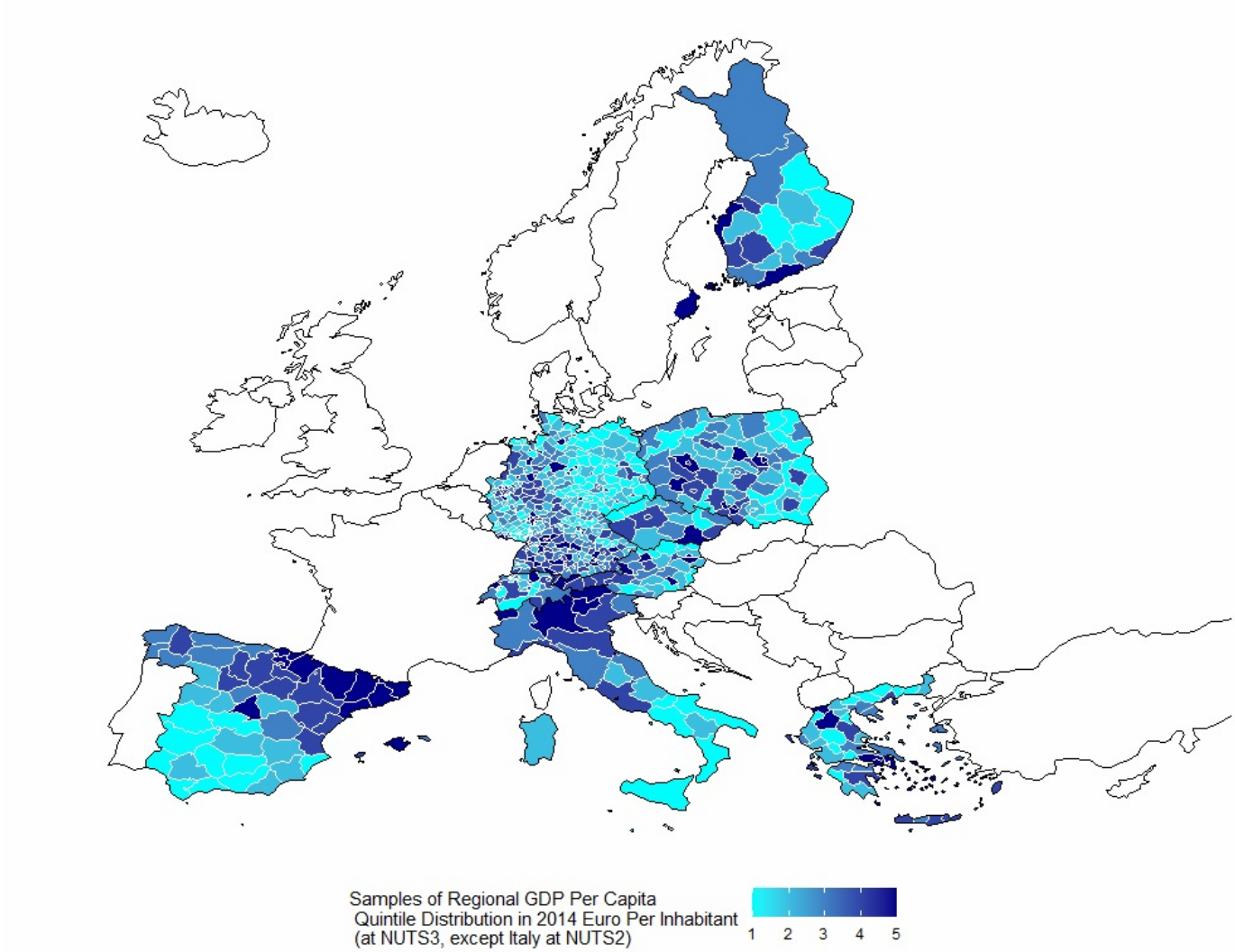
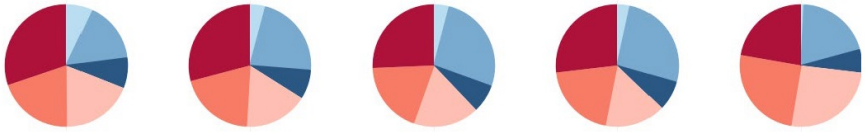
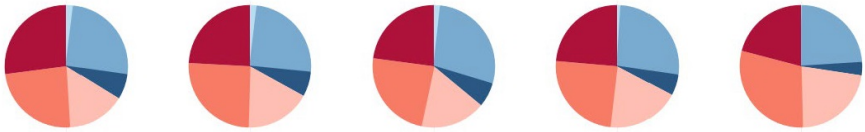
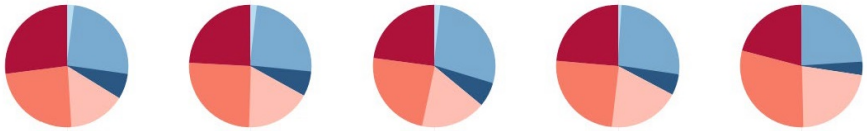
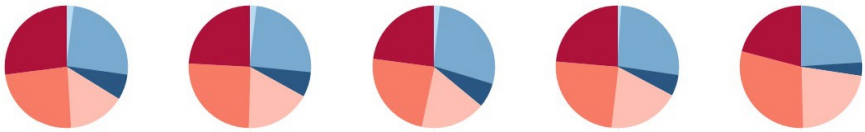
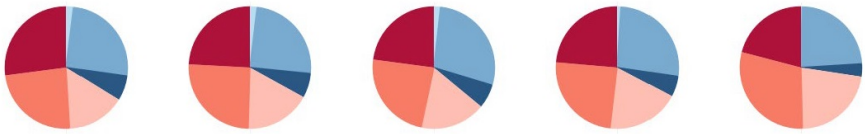
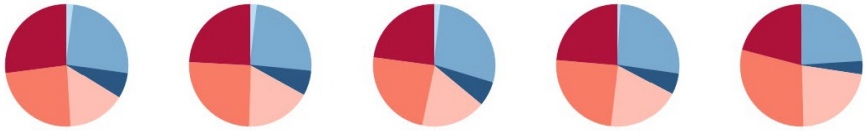


Table I. Gross Value Added Productivity by Sectors

Gross Value Added (Million EUR)	1ST Quintile (Lowest)	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile (Highest)
Austria (2013)					
Czech Republic (2013)					
Finland (2011)					
Germany (2013)					
Greece (2015)					
Italy (2006)					
Poland (2011)					
Spain (2008)					
Switzerland (2011)	NOT IN EUROZONE				

Source: The Annual Regional Database of the European Commission (ARDECO) database. The GVA measure by sector at constant prices in million euros (2015) is used for the NUTS3-level of regions only.

11 Sensitivity Analysis, Outliers Included

Our baseline analysis of the vote choice variation (Table 1, Model 1 of the main text) trims the data by excluding the first-stage regression residuals exceeding 1.5 standard deviations. This data cropping helps mitigate the potential impact of heteroskedasticity on the OLS model estimates for the continuous variable, as some election years with extreme residual variance in model residuals reveal patterns of heteroskedasticity (Lupu and Pontusson, 2011).

To ensure the robustness of our findings from Model 1 in Table 1, we re-estimated the model including observations with large regression residuals in the first stage OLS regression. Models [26] and [27] in OA Table J confirm that the key results are consistent.

Table J. Robustness Check for Analyses With Outliers

	M[26] With Outliers	M[27] With Outliers
First Stage: OLS Estimation		
Household Income Quintile	-0.016***	-0.024***
	[0.005]	[0.005]
Regional Productivity Quintile	0.010**	0.009**
	[0.004]	[0.004]
Self-placement: Right(0)-Left(10)	0.019***	0.019***
	[0.000]	[0.000]
Age 16-95)	0.002	0.005**
	[0.002]	[0.002]
Age Squared	-0.000	-0.000*
	[0.000]	[0.000]
Male (1=Yes, 0=No)	-0.017	-0.008
	[0.012]	[0.012]
Higher Education (1=Yes, 0=No)	-0.070***	-0.050***
	[0.015]	[0.014]
Married (1=Yes, 0=No)	-0.022	-0.025*
	[0.015]	[0.015]
Religiosity (1=Yes, 0=No)	-0.131***	-0.134***
	[0.015]	[0.015]
Constant	[0.874***	0.430***
	[0.062]	[0.062]
Second Stage: Variance Estimation		
Household Income Quintile	-0.056***	-0.056***
	[0.008]	[0.008]
Regional Productivity Quintile	-0.022***	-0.015**
	[0.007]	[0.008]
Vote Turnout (%)	-0.042***	-0.044***
	[0.001]	[0.001]
Constant	2.761***	2.805***
	[0.071]	[0.073]
Log Likelihood	-15,944	-14,977
AIC	31,895	29,962
BIC	31,927	29,994
First Stage County FE	Yes	Yes
First Stage Election Year FE	No	Yes
Sample	Full	Full
Observations	20,595	20,595

Notes: Results are based on Table 1, Model [1], from the main text. Two-tailed tests significant *p<0.1, **p<0.05, ***p<0.01. This uses a continuous measure for the dependent variable. P-values are in brackets. In Model [27], our election year fixed effects for the European Union parliamentary election held in 2004, 2009, and 2014 and years preceding each EU parliamentary election year.

References

Aitkin, Murray. 1987. "Modelling Variance Heterogeneity in Normal Regression Using GLIM". *Journal of the Royal Statistical Society: Series C (Applied Statistics) 36*(3), 332–339.

Armingeon, Klaus, Virginia Wenger, Fiona Wiedemeier, Christian Isler, Laura Knöpfel, David Weisstanner, and Sarah Engler. 2020. *Comparative Political Dataset 1960-2018*. Bern: Institute of Political Science, University of Berne.

Benoit, Kenneth and Michael Laver. 2007. "Estimating Party Policy Positions". *Electoral Studies 26*(1), 90–107.

Budge, Ian and Michael J Laver. 2016. *Party Policy and Government Coalitions*. New York: ST. Martin Press.

Harville, David A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems". *Journal of the American Statistical Association 72*(358), 320–338.

Laakso, Markku and Rein Taagepera. 1979. ""Effective" Number of Parties: A Measure with Application to West Europe". *Comparative Political Studies 12*(1), 3–27.

Lachat, Romain. 2018. "Which Way From Left to Right?". *International Political Science Review 39*(4), 419–435.

Lee, Dong Wook and Melissa Rogers. 2019. "Measuring Geographic Distribution for Political Research". *Political Analysis 27*(3), 263–280.

Lowe, Will, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. "Scaling Policy Preferences from Coded Political Texts". *Legislative Studies Quarterly 36*(1), 123–155.

Lupu, Noam and Jonas Pontusson. 2011. "The Structure of Inequality and the Politics of Redistribution". *American Political Science Review*, 316–336.

Ogorzalek, Thomas, Spencer Piston, and Luisa Godinez Puig. 2019. "Nationally Poor, Locally Rich". *Electoral Studies 67*.

Rueda, David and Daniel Stegmueller. 2019. *Who Wants What?: Redistribution Preferences in Comparative Perspective*. New York: Cambridge University Press.

Smyth, Gordon K. 2002. "An Efficient Algorithm for REML in Heteroscedastic Regression". *Journal of Computational and Graphical Statistics 11*(4), 836–847.

Soifer, Hillel David. 2019. "Units of Analysis in Subnational Research". In: *Inside Countries: Subnational Research in Comparative Politics*. Eds. A. Giraudy, E. Moncada, and R. Snyder eds. . New York: Cambridge University Press, 92–112.

Volkens, Andrea, Pola Lehmann, Nicolas Merz, Sven Regel, Annika Werner, Onawa Lacewell, and Henrike Schultze. 2013. "The Manifesto Data Collection". *Manifesto Project (MRG/CMP/MARPOR). Version 2013a. Berlin: Wissenschaftszentrum Berlin fü r Sozialforschung (WZB)*.

Western, Bruce and Deirdre Bloome. 2009. "Variance Function Regressions for Studying Inequality". *Sociological Methodology 39*(1), 293–326.