Università Commerciale "Luigi Bocconi"

PhD School

PhD program in: Public Policy and Administration

Cycle: 36°

Disciplinary Field (code): SPS/04

Essays on the Material Origins of Political Change

Advisor: Prof. Dr. Massimo Anelli

Co-Advisor: Prof. Dr. Catherine De Vries

PhD Thesis by

Paolo Agnolin

ID number: 3087769

Year: 2025

Abstract

This dissertation explores how automation, technological change, and economic transformations shape political dynamics in post-industrial societies. It examines how shifts in material conditions influence both the demand and supply sides of politics, with evidence drawn from Western Europe and the United States. The first essay addresses methodological issues in studying the determinants of the globalization backlash, highlighting the bias introduced by post-treatment variables in regressions comparing economic and cultural drivers of voting behavior and providing additional evidence on the culture-economy nexus. The second essay investigates the impact of automation on trade unions in Western Europe, finding that regions more exposed to automation experience a decrease in union density, primarily driven by a broader labor market shift away from unionized industries. The third essay offers a structural explanation for the changing composition of political elites, examining how automation influences the likelihood of different social groups pursuing political office. Drawing on multiple sources of data on political candidates and occupational backgrounds in the US, the analysis reveals that areas with greater exposure to automation experience a decline in candidates from occupations most impacted by technological change. This trend contributes to the underrepresentation of working-class and automatable workers in politics.

Contents

Introduction

The ongoing transformations driven by automation, technological change, and globalization are reshaping the structure of advanced economies and presenting new challenges for communities. As these forces reshape labor markets, they also alter the foundations of political life. This dissertation investigates the political consequences of these structural economic shifts, with a focus on how technological change affects political preferences, intermediate organizations, and political representation in post-industrial societies.

A longstanding premise in political economy holds that individuals' political preferences are shaped, at least in part, by their economic condition. From this perspective, the distribution of resources in a society and the material well-being of citizens influence their attitudes toward policy and their political activity. The idea that economic transformation can drive political change has deep roots in the social sciences and has informed debates on the political consequences of, *inter alia*, inequality, unemployment, and economic insecurity. Economic shocks, such as recessions or structural shifts in labor demand, can act as catalysts for political realignment, especially when they erode the economic position of large segments of the population. While early accounts traditionally focused on broad class conflict and systemic instability, more recent research has examined how specific forms of economic dislocation reshape political attitudes and behavior. At the core of this tradition lies the idea that changes in individuals' material conditions can act as powerful triggers for political engagement, realignment, or withdrawal.

Building on this tradition, a growing body of work in economics and political science has deepened our understanding of how technological change is reshaping not only economies, but also modern societies down to their very deepest roots. Evidence suggests that the automation of routine tasks and the exposure to international trade in some sectors has contributed to changing earnings distribution and declining employment shares of some traditional occupations across a broad range of advanced economies.

The core contribution of this dissertation is to advance our understanding of how material transformations in the economy translate into political change. It does so by integrating insights from comparative politics, political economy, and labor studies. The dissertation is structured as a cumulative project comprising three empirical studies, each addressing a distinct aspect of the political consequences of structural changes: (1) the demand side of politics, through changes in public opinion and voting behavior; (2) the role of intermediate organizations, in particular trade unions; and (3) the supply side of politics, through changes in the composition of political elites. The first empirical study primarily examines the political consequences of economic shocks, using economic globalization as a key case, whereas the other two focus on the impact of technological change, particularly automation and robotization. While each essay addresses a specific research question and stands on its own, together they speak to a common concern: understanding how economic dislocation caused by automation and other structural forces reverberates through the political system.

The essays in this dissertation offer both methodological and theoretical contributions. Each chapter employs quantitative analysis grounded in a causal inference framework, emphasizing methodological rigor and innovation while introducing new and valuable data. The theoretical approach combines individual-level analysis with a comprehensive exploration of contextual factors, such as labor market dynamics and institutions. A central theme of this research is its emphasis on understanding how the technological, economic and political domains are deeply interrelated, rather than treating them as isolated forces. Its comparative scope encompasses post-industrial, high-income societies, drawing evidence from Western Europe and the United States.

The first central theme of this dissertation is the study of the material origins of voting behavior, particularly in the context of structural economic changes. The first essay, coauthored with Italo Colantone and Piero Stanig, contributes both methodologically and empirically to the debate on the determinants of the “globalization backlash”. This phenomenon, marked by the rise of nationalist, isolationist, and radical right parties, has typically been explained through two broad lenses: one emphasizing economic distress, and the other focusing on non-merely economic factors such as immigration, identity, and demographic change. Rather than being mutually exclusive, these two explanatory frameworks are deeply intertwined. Their interaction poses important methodological challenges for scholars seeking to isolate causal effects. In this paper, we study the implications of post-treatment bias in the context of voting behavior. By replicating previous studies on voting behavior, we show how pitting cultural and economic explanations of the globalization backlash against each other may lead to biased estimates of causal effects. Additionally, we use observational data from 15 Western European countries to explore the causal relationship between economic shocks and cultural attitudes. We find that regional exposure to import competition leads individuals to express lower support for democracy and liberal values, more conservative views, and heightened concern over immigration.

An important implication of our findings for the broader academic community studying voting behavior is that the pursuit of a single, overarching model to explain complex social phenomena, such as voting, is unlikely to be fruitful when the underlying explanatory factors are deeply interrelated. Instead, a more productive approach, grounded in a principled causal framework, involves making separate, well-identified claims about the effect of each factor in turn.

A second major theme of this dissertation is that intermediate organizations and labor market institutions play a pivotal role in shaping political change, particularly in the context of structural economic transformations. Historically, labor unions were crucial in liberal democracies by hindering the increasing wage inequality, by channeling political demands and discontent into an organized voice, and by linking blue-collar constituencies to mainstream left parties. However, the past decades have witnessed a progressive weakening of unions, coinciding with an atomization of political demands. In the second essay, coauthored with Massimo Anelli, Italo Colantone and Piero Stanig, we investigate the impact of technological advancements —specifically industrial robot adoption—on trade unions in Western Europe. First, we estimate novel data on union density by region and sector across 15 Western European countries. This new data reveals a broad and systematic decline in unionization across both regions and industries over recent decades. This trend carries significant political implications, as higher levels of union density have traditionally been associated with greater support for mainstream left parties. A weakening of union presence may therefore contribute to broader political realignments.

The empirical analysis shows that automation, and industrial robotization in particular, is a major driver of declining unionization. Our findings reveal that regions more exposed to automation experience a significant decline in union density. Notably, this decline is not primarily due to a weakening of unions within automation-affected sectors but rather stems from a broader transformation in the labor market, as employment shift from traditionally unionized industries toward sectors with lower union presence. These findings contribute to explaining why technologically driven economic grievances increasingly manifest as support for right-wing movements rather than bolstering pro-redistribution left-wing parties.

Overall, our results highlight the importance of unionization as a contextual factor that may influence electoral dynamics at the local level, shaping the political repercussions of automation. As the share of employment in traditionally unionized sectors like manufacturing declines, and job growth concentrates in less-unionized areas such as the private service sector, the ability of trade unions to mobilize workers and exert political influence may be increasingly jeopardized.

Finally, in the last essay, I provide a structural explanation for the changing composition of political elites, focusing on the role of technological advancements. The shift from an industrial to a post-industrial society has significantly reshaped the political landscape, with studies documenting a decline in working-class representation. I argue that the distributional effects of automation play a key role in shaping political representation. Specifically, I theorize that automation alters the distribution of economic resources and opportunities across occupations, influencing who runs for office. I propose that those most negatively affected by automation, particularly from working-class or automatable backgrounds, are less likely to pursue political careers.

This research examines how material changes impact political candidacy, using data from U.S. state legislators, lower-level office candidates in California, and a representative survey of the U.S. population. The findings show that in regions with higher exposure to automation, there is a notable decrease in candidates from occupations most affected by robotization. By shifting the focus to the supply side of politics, this study offers new insights into how economic and technological transformations shape political representation and inequality.

While most existing research on automation emphasizes demand-side responses, such as changes in voting behavior, this paper highlights the importance of supply-side dynamics. A central motivation behind this work is the idea that the composition of the candidate pool constrains voters' choices and shapes their political perceptions, while also influencing the kinds of policies ultimately enacted. At the same time, the paper shows that even supply-side factors, such as who chooses to run for office, can themselves be shaped by structural economic transformations. This reveals a new interconnection between structuralist approaches to political change and the study of candidate emergence and elite composition.

Taken together, the three essays in this dissertation aim to deepen our understanding of how structural economic transformations - particularly those driven by globalization and technological change - reshape the foundations of democratic politics. A joint reading of the dissertation suggests that voter preferences, the role of intermediate organizations such as trade unions, and the composition of political candidates are not isolated dimensions, but interconnected channels through which material shocks influence political life. While each essay focuses on a distinct mechanism, collectively they show how economic and technological forces reverberate across the political system: shifting political support and individual attitudes, eroding collective representation, and transforming patterns of who is represented and who is in a position to influence the democratic process. In doing so, the project highlights the need for an integrated perspective that combines structuralist insights with attention to both institutional context and political agency.

Chapter 1

In Search of the Causes of the Globalization Backlash: Methodological Considerations on Post-Treatment Bias

In Search of the Causes of the Globalization Backlash: Methodological Considerations on Post-Treatment Bias*

Paolo Agnolin† Italo Colantone‡ Piero Stanig§

Abstract

We study the implications of post-treatment bias in the context of the globalization backlash. We discuss whether horse-race regressions can inform about the relative role of economic vs. cultural drivers. We make three methodological points: (1) if and insofar as cultural variables are post-treatment with respect to economic factors, the estimates of the effect of economic shocks on voting are biased in regressions that include cultural controls (and vice versa); (2) for the same reason, such horse-race regressions do not allow to accurately estimate the relative role of economic vs. cultural factors; (3) one cannot infer mediation effects from changes in regression coefficients for a given factor of interest before and after including post-treatment controls. We accompany the methodological discussion with empirical evidence on the relevance of post-treatment bias in studies of the globalization backlash, both by replicating and expanding on earlier studies, and by presenting novel cross-country results on the culture-economy nexus.
Keywords: Globalization backlash; populism; radical right; post-treatment bias; causal mediation analysis

*We thank Francesco Bellogi for excellent research assistance. We thank Martin Elff, Jeff Frieden, Peter Hall, Edward Mansfield, Lanny Martin, Diana Mutz, Dani Rodrik, John Sides, Chris Way, as well as seminar participants at Harvard University, the MPSA 2018 Meeting in Chicago, the EPSA 2018 meeting in Vienna, the 2019 Workshop on Populism: Causes and Responses of the London School of Economics, the 2022 Globalization Backlash Workshop of the Niehaus Center for Globalization and Governance of Princeton University, and the 2023 Workshop on Resilience of Democracy of Mainz University for helpful comments. The usual disclaimer applies. This paper is part of the project "Class Politics in the XXI Century". An antecedent version of this study has appeared in 2018 as a Bocconi Baffi-Carefin working paper, with the title "The Economic Roots of the Cultural Backlash: Global Competition and Attitudes in Western Europe".

†Bocconi University, Duke University and Dondena Research Centre, Via Roentgen 1, 20136 Milan, Italy. Contact: paolo.agnolin@unibocconi.it.

‡Bocconi University, Baffi Research Centre, GREEN Research Centre, CESifo and FEEM, Via Roentgen 1, 20136 Milan, Italy. Contact: italo.colantone@unibocconi.it.

§Department of Social and Political Sciences, Bocconi University, Italy; Division of Social Sciences, Yale-NUS, Singapore; Department of Political Science, NUS, Singapore. Via Roentgen 1, 20136 Milan, Italy. Contact: piero.stanig@unibocconi.it.

1 Introduction

The success of nationalist, isolationist, and radical-right parties and candidates in Western democracies has stimulated a lively academic debate. Scholars are far from reaching a consensus regarding the determinants of this phenomenon, which is widely referred to as "globalization backlash" (Colantone et al., 2022; Walter, 2021).

Several studies have documented how the backlash is linked to economic distress, especially as driven by globalization (e.g., Autor et al., 2020; Colantone and Stanig, 2018a, 2018b; Milner, 2021) and technological progress (e.g., Anelli et al., 2021; Gallego and Kurer, 2022; Milner, 2021). The main idea behind these studies is that structural economic changes create winners and losers, and the ensuing economic grievances foster anti-establishment voting. Scholars pointing to a different family of explanations emphasize how recent political developments result from a "cultural backlash". According to this view, a prominent role is played by the status threat posed by international migration, changing race and gender relations, and demographic trends (e.g., Hangartner et al., 2019; Norris and Inglehart, 2019; Mutz, 2018).

Some observers see economic and cultural explanations of the backlash as fundamentally alternative: it has to be one or the other. Instead, in our view there is sufficient empirical evidence to conclude that *both* economic *and* cultural explanations are crucial to account for contemporary trends in electoral outcomes. Our stance is not isolated: several contributions in the literature emphasize how economic and cultural factors may interplay with each other (e.g., Franzese, 2019; Frieden, 2022; Ferrari et al., 2021). A growing body of evidence documents the impact of economic factors on individual attitudes such as nativism and authoritarianism (e.g., Anelli et al., 2021; Ballard-Rosa et al., 2021, Ballard-Rosa et al., 2022; Carreras et al., 2019; Ferrara, 2022; Hays et al., 2019). At the same time, scholars have highlighted how cultural concerns may increase the political salience of economic shocks (e.g., Gidron and Hall, 2017; Margalit, 2019).

The interplay between economic and cultural factors poses methodological challenges. These translate directly into matters of research design. In fact, a common approach to studying the causal structure of the globalization backlash relies on a sort of "horse-race" empirical strategy. In this approach, economic and cultural factors are included jointly as explanatory variables in vote regressions. We refer to these specifications as "long" regressions, as opposed to "short" regressions that include one factor at the time. Loss of statistical significance for the economic factor in a long regression, for instance, as compared to a short regression that omits cultural variables, is interpreted as evidence that "the economy does not matter". From a methodological point of view, we make three main points in this paper. First, we underline that, if and insofar as cultural variables are post-treatment with respect to economic shocks, the estimates of the effect of economic shocks are going to be biased in long regressions. Second, by the same token, it is impossible to accurately estimate the relative role of economic vs. cultural factors from the same specifications. Third, one cannot infer mediation effects—i.e., mechanisms—from changes in regression coefficients for a given economic factor before and after including controls for culture. Mutatis mutandis, exactly the same considerations can be made for studies of cultural factors.

We proceed in three steps. First, we provide evidence on the substantive empirical relevance of post-treatment bias in studies of the globalization backlash. To this purpose, we replicate some results from three published papers that study the Brexit referendum (Colantone and Stanig, 2018a and Chan et al., 2020) and the US presidential election of 2016 (Mutz, 2018).[1] We then show how these results change when including vs. excluding some controls that are arguably post-treatment. For instance, the main finding of Colantone and Stanig (2018a) is that individuals living in regions more exposed to import competition from China are more likely to support the Leave option in the Brexit referendum. This result does not survive the inclusion of a control for individual attitudes as to whether

[1] Replication materials for all the analyses presented in this paper can be found at Agnolin et al. (2024).

immigration is good or bad for Britain's cultural life. One could interpret this evidence as suggesting that the import shock does not matter, and that cultural concerns with respect to immigration are what really drives support for Brexit. However, we show that such cultural concerns are actually post-treatment with respect to the import shock, as higher trade exposure is associated to worse attitudes about immigration. Loss of significance for trade exposure in the long regression, as compared to the short regression that does not condition for culture, is then likely due to post-treatment bias.

We make a very similar point with respect to the Brexit analysis by Chan et al. (2020), who find that the main result of Colantone and Stanig (2018a) on the China shock is not robust to controlling for individual cultural consumption traits. Following the data description in the paper, we construct a dataset very close to theirs, and we show that the cultural characteristics included in their analysis are post-treatment to trade exposure. Omitting these post-treatment controls allows to replicate the main result of Colantone and Stanig (2018a), despite the use of a different data source. Finally, in the case of Mutz (2018), we use the replication dataset to show that individuals holding more positive evaluations of their personal economic situation are less likely to support Trump in the 2016 presidential election. This result is lost once we include the stance on immigration policy as a control. Evidence of this type leads Mutz (2018) to conclude that status threat, not economic hardship, explains the 2016 presidential vote. Yet, also in this case we show that the immigration stance is arguably post-treatment to the individual economic situation, making strong conclusions on the irrelevance of economic factors ultimately controversial.

In the second part of the paper, we illustrate this methodological issue with a contrived example on the globalization backlash. This is complemented by regressions based on simulated data. We do not have the pretense of being very innovative in this exercise. Indeed, an ample literature addresses post-treatment bias, going back to Rosenbaum (1984), and even earlier to seminal contributions such as Frisch and Waugh (1933). This methodological issue is also covered in major textbooks (e.g., Angrist and Pischke, 2009;

Gelman and Hill, 2007). Yet its substantive and practical consequences do not seem to be fully appreciated in some of the recent voting behavior debate. In this respect, Acharya et al. (2016) show that 40% of observational studies published from 2010 to 2015 in three of the top journals in political science explicitly condition on a post-treatment variable, with an additional 27% conditioning on a plausibly post-treatment variable. The problem might even be more serious in experimental settings: Montgomery et al. (2018) estimate that 47% of experimental studies engage in post-treatment conditioning.

Overall, the current debate seems to be based on excessively optimistic expectations regarding how much one can learn about causal ordering from observational voting behavior data based on regressions that condition on many variables. Crucially, causal ordering is very hard to infer empirically, and for that matter, experimental methods are not superior under this respect, in particular when one is interested in the ordering that obtains naturally, and therefore questions of external validity of experiments are of the foremost importance. We aim at providing some structure to the debate. The first main message of our methodological illustration is that controlling for a post-treatment variable leads to biased estimates on the main factor of interest. Besides that, we also recognize that the aim of long regressions is not always to run a horse race, but also to better understand mechanisms. Even if this is the case, we show how comparing coefficients across short and long specifications is not going to provide valid answers. Specific methods to explore mechanisms have been proposed in the recent causal mediation literature, and the assumptions required for them to yield valid answers have been spelled out (see, e.g., Imai et al., 2011). We clarify, in the context of our contrived example, how these assumptions are potentially very demanding, being rich in substance and far from merely technical. The appropriateness of causal mediation techniques should then always be carefully evaluated on a case-by-case basis.

Finally, in the third part of the paper, we provide novel observational evidence on how pervasive the issue of post-treatment bias can be in studies of the globalization backlash that investigate the role of trade exposure as an economic factor. We focus on fifteen western European countries over 1995-2018, employing individual-level survey data from the European Social Survey (ESS) and the European Values Study (EVS). Expanding on the analysis presented in the first section, and along the lines of earlier studies by Ballard-Rosa et al. (2022), Ballard-Rosa et al. (2021), Ferrara (2022), Ferrari et al. (2021) and Hays et al. (2019), we show that exposure to import competition from China at the regional level triggers individual reactions in terms of an array of cultural attitudes, which should then be considered post-treatment controls in long vote regressions. Specifically, more trade-exposed individuals are systematically less supportive of democracy and liberal values, more in favor of unconstrained strong leaders, less permissive with respect to abortion, and particularly concerned with immigration, especially with the cultural threat posed by it.

Overall, we provide further evidence pointing to the existence of economic roots for the cultural shifts observed in Western democracies. Obviously, cultural shifts need not be necessarily related to the economic context. It is important to clarify that we are not claiming that all (or most) of the variation in these cultural or attitudinal variables is driven by economic factors, nor that these cultural aspects do not play an important independent role in shaping voting behavior. What our findings suggest is that cultural attitudes are *at least partly* endogenous to trade exposure, which is enough to make them post-treatment.

The main practical implication of our analysis is that a study of the globalization backlash that uses a plausibly-identified strategy for estimating the effect of an economic shock, but does not "control for culture", might be better than one that does condition on it. The same applies, symmetrically, for studies that focus on the effects of cultural causes, exploiting plausibly exogenous variation of variables bearing a cultural meaning. Notable examples in this respect are: Barone et al. (2016), Tabellini (2019), and Clayton et al. (2021) on immigrants; Hangartner et al. (2019) and Dustmann et al. (2019) on refugee arrivals; Anduiza and Rico (2023) on sexism; and Cavaille and Marshall (2019) on education.

The search for a single model explaining intricate social phenomena such as voting behavior is fraught with theoretical and methodological shortcomings. These are inherently related to the complex interplay of different factors, which plagues horse-race approaches. Echoing Gelman and Imbens (2013), the big question regarding the causes of the globalization backlash is worth asking, but knowing that, ultimately, it cannot be answered in that form in a single empirical analysis. A thorough answer in a principled causal framework will take the form of separate claims about the role of one given factor at the time. Building a cumulative body of such results might be the best way forward, also in view of informing policy action. The risk involved is that the individual studies are to some extent inherently non-cumulative. But, crucially, one could also argue that policy intervention can be driven by non-cumulative results. For instance, dismissing the economic roots of the backlash based on questionably-specified empirical analyses may be very dangerous; conversely, recognizing the causal role of deindustrialization in the success of radical right parties and candidates might suggest that policy interventions are needed to address economic distress. Having said that, it is also interesting to investigate the interplay of economic and cultural factors, and their relative role in determining voting behavior. In the last section of the paper, we discuss possible ways forward in this direction, encompassing both mediation analysis and structural equation modeling.

In closing, we note that we cast our discussion in the framework of the globalization backlash. However, the methodological points we put forward are more generally relevant for other political science applications characterized by similar features. Arguably, these are likely to constitute the majority of contexts when it comes to studies of voting behavior.

2 The perils of the horse race

In this section, we provide evidence on the relevance of post-treatment bias in studies of the globalization backlash. Specifically, we show how including vs. excluding post-treatment controls may affect the main findings of three published studies.

We begin by considering the paper by Colantone and Stanig (2018a), which studies the link between globalization and Brexit. The study finds that higher exposure to import competition from China, measured between 1990 and 2007, leads to higher support for the Leave option in the Brexit referendum of 2016. Trade exposure is measured at the regional level. Higher shocks are attributed to regions that were historically specialized in industries in which Chinese imports have subsequently grown more. The paper shows that higher trade exposure is related to long-run regional economic decline, which is in turn politically consequential. The analysis is carried out both at the regional level and at the individual level, with equivalent results across a large number of different models.

For the purpose of our study, we focus on the individual-level analysis of Colantone and Stanig (2018a), which is based on data from Wave 8 of the British Election Study (Evans et al., 2016). Specifically, in column 1 of Table 1 we reproduce their main result, using the published replication dataset. The outcome variable is an indicator taking value one if the individual reports supporting the Leave option in the referendum. This is regressed on exposure to Chinese imports in the region of residence. Trade exposure is measured at the fine-grained NUTS-3 level. The hierarchical model we estimate, as in the original paper, includes random intercepts for NUTS-3 regions, as well as fixed effects for coarser NUTS-1 regions, along with controls for age, gender, and education of individuals. The coefficient on the China shock is positive and statistically significant. In terms of magnitude, a unit increase in the China shock (i.e., one thousand euros per worker) is associated with an increase in the probability of Leave support by 8.2 percentage points.

In column 2 of Table 1 we augment the model including a control for the individual stance about whether immigration is good for Britain's cultural life. This control is sourced from the replication dataset of Colantone and Stanig (2018a). It is measured on a 7-point scale, with higher values denoting more positive views. The cultural stance on immigration is significantly associated with support for Leave. In particular, the negative sign of the coefficient indicates that individuals holding more negative views on immigration are more likely to support the Leave option. Importantly, the coefficient on the China shock in this model is no longer statistically significant, and very close to zero.

One could read this result as evidence of omitted variable bias in the short model of column 1. That is, once controlling for immigration attitudes in the long model, import competition is not a significant determinant of vote. It is culture, not the economy, that explains support for Brexit. Yet this interpretation is unwarranted because immigration attitudes are post-treatment with respect to the import shock. This is what we show in column 1 of Table 2, where we regress the individual cultural stance about immigration on trade exposure, using the same specification as in column 1 of Table 1. The coefficient on the China shock is positive and statistically significant, indicating that, even conditional on all the other controls, respondents living in areas more exposed to import competition tend to display less favorable attitudes on immigration. The effect of the import shock on immigration attitudes is actually modest: one standard deviation increase in the import shock is associated with worsening immigration attitudes by 4% of a standard deviation, and a move from minimum to maximum trade exposure is associated with a move in attitudes by less than half of a point on a 7-point scale. This is not surprising, as cultural attitudes about immigration are affected by many factors other than trade exposure. Yet, even this relatively weak endogeneity of attitudes to the import shock may invalidate inferences from the long regression approach.

One could also be tempted to read the results just described as indicating that culture fully mediates the effect of the economic variable. In particular, a worsening of immigration attitudes is the (only) mechanism through which higher import shocks translate into higher Leave support. In the methodological section of the paper we show that this conclusion, too, would be unwarranted. Proper mediation analysis requires much more than just a comparison of short vs. long regression results, and in addition it hinges upon assumptions which may not hold in this empirical context.

Table 1: Short vs. long regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Source:	**Colantone & Stanig (2018)**		**Chan et al. (2020)**		**Mutz (2018)**	
Dep. var.:	Leave	Leave	Leave	Leave	Trump th.	Trump th.
Economic factor:	China shock	China shock	China shock	China shock	Family fin.	Family fin.
	0.082*	-0.004	0.445*	0.380	-0.204**	-0.114
	[0.040]	[0.033]	[0.210]	[0.210]	[0.078]	[0.075]
Cultural attitude:	-	Immigration	-	Cons. omnivore	-	Immigration
		-0.132**		-1.074**		-1.306**
		[0.002]		[0.081]		[0.079]
	-	-	-	Cons. paucivore	-	-
				-0.398**		
				[0.040]		
Observations	15,819	15,819	18,909	18,909	2,888	2,888
Model	Hierarchical	Hierarchical	Logit	Logit	Linear	Linear

Note: ** p<0.01; * p<0.05

Next, we consider the paper by Chan et al. (2020), who use data from the UK Understanding Society (UKHLS) survey (University of Essex, Institute for Social and Economic Research, 2023) to study the vote in the Brexit referendum. In particular, their explicit aim is to evaluate the relative strength of two different narratives about the social bases of Brexit. One is economic, as related to the China shock, and more generally to deindustrialization and regional economic decline. The other is more cultural in nature, relating to a resurgence of nationalism and cultural insularity. As in Colantone and Stanig (2018a), the China shock is assigned to each individual based on the NUTS-3 region of residence. In terms of cultural variables, respondents are classified in three categories based on cultural consumption patterns: (1) cultural "omnivores", who consume many different genres of music and visual arts; (2) cultural "univores", who consume only popular genres; and (3) cultural "paucivores", who are in between omnivores and univores. Moreover, Chan et al. (2020) employ controls for self-reported strength of British identity, and for self-identified national identities (English, Scottish, etc.).

Table 2: Evidence of culture-economy nexus

	(1)	(2)	(3)	(4)
Source:	**Colantone & Stanig (2018)**	**Chan et al. (2020)**		**Mutz (2018)**
Dep. var.:	Immigration	Cons. omnivore	Cons. paucivore	Immigration
Economic factor:	China shock	China shock	China shock	Family finances
	-0.627**	-1.071**	-0.412*	0.069**
	[0.173]	[0.403]	[0.196]	[0.020]
Observations	15,819	18,909	18,909	2,888
Model	Hierarchical	Mult. logit	Mult. logit	Linear

Note: ** p<0.01; * p<0.05

Since no official replication material is available for Chan et al. (2020), we independently reconstruct a replication database from the same sources. We provide full details on this exercise in Section A of the Online Appendix.[2] In column 3 of Table 1, we present the estimates of a "short" logit regression in the spirit of Chan et al. (2020) focusing on the role of the China shock, i.e., excluding cultural variables. The outcome variable is an indicator equal to 1 if the respondent reports supporting the Leave option. This is regressed on the China shock in the region of residence of the respondent. We include the same pre-treatment controls as in Chan et al. (2020): ethnicity, education, marriage status, family size, gender, and a quadratic polynomial for age. In addition, as in their paper, we include fixed effects for NUTS-1 regions, and controls for social class and for poverty (i.e., household income below 60% of the median). Like in various specifications of the original paper, the coefficient on the China shock is positive and statistically significant. Using the crude approximation of dividing the logit coefficient by 4 to obtain a magnitude on probability scale, a unit increase in the China shock is associated with an 11 percentage point increase in the probability of supporting Leave. This is quite close to the result in column 1 of Table 1. Overall, notwithstanding the differences in survey data and specification, the result on

[2]Our exercise is not meant as a criticism, or a direct re-evaluation, of the findings in Chan et al. (2020). In particular, we do not have access to some predictors—like social status—that play an important role there. In addition, their findings seem to be sensible, and, overall, they have the merit of highlighting the complexity of public opinion in the run-up and the wake of the Brexit referendum.

the China shock based on the data and the basic conditioning set of Chan et al. (2020) is actually substantially equivalent to the one in Colantone and Stanig (2018a).

In column 4 of Table 1 we augment the specification of column 3 with cultural variables analogous to those used in Chan et al. (2020), that we independently estimate from the same data, and with the same methodology, used in the original paper. Being a cultural omnivore, and to a lesser extent a paucivore, as opposed to a univore, is significantly associated with lower Leave support.[3] At the same time, the coefficient on the China shock is 15% smaller in magnitude, and no longer statistically significant. Very similar findings in Chan et al. (2020) are interpreted as indicating that the China shock result is "not as robust as reported" (p. 480) and "quite different to that reported by Colantone and Stanig (2018a)" (p. 484). However, also in this case the long regression is actually plagued by post-treatment bias, since cultural consumption is endogenous to trade exposure.

This is what we show in columns 2-3 of Table 2, where we regress cultural consumption classes on the China shock, conditioning on the full set of covariates employed in column 3 of Table 1. Specifically, we report estimates of a multinomial logit model where the outcome is the cultural class in which a respondent is classified. The two columns report, respectively, the estimates for the omnivore and for the paucivore class, with the univore class being the reference category. In both equations, the coefficient on the China shock is negative and highly statistically significant. This indicates that, even conditional on a rich set of demographic, income, and class characteristics, and on the NUTS-1 region of residence, respondents in areas more exposed to Chinese import competition are much less likely to be cultural omnivores, and somewhat less likely to be cultural paucivores, compared to being univores. In other words, otherwise identical respondents, who reside in the same NUTS-1 region, but in NUTS-3 regions that differ in terms of import competition, display different cultural consumption patterns. In terms of magnitudes, one standard deviation increase

[3] Strength of British identity, and identifying as English, are significantly associated with higher support for Leave. Full results are reported in Table A.3 of the Online Appendix.

in the import shock is associated to a lower probability of being omnivore by around 3.5 percentage points, and to a lower probability of being paucivore by around 1.3 percentage points. These results are not suprising in light of the available evidence on broad dynamics of socio-economic decline related to the China shock (see, e.g., Autor et al., 2021), and call for caution in the comparison of short vs. long regression results in this context.

As a third example, we focus on some results based on the analysis by Mutz (2018), who studies support for Trump in the US presidential election of 2016. Specifically, in column 5 of Table 1 we consider as outcome variable the Trump thermometer advantage rating employed in the original paper as one of the proxies for Trump support.[4] This is regressed on an indicator of pocketbook economic evaluation, namely the subjective perception of family finances, measured on a 5-point scale. The specification also includes controls for education, ethnicity, partisanship, gender, age, religion, income, unemployment, median income in the place of residence, and sociotropic economic evaluation, i.e., the perception concerning the state of the national economy. Full details, and full regression results, are reported in Section B of the Online Appendix.

Here we highlight that the coefficient on pocketbook economic perceptions is negative and statistically significant, indicating that, conditional on all the other predictors, respondents with more favorable views of their personal finances were less supportive of Donald Trump. A similar finding is obtained for respondents with more favorable views of the national economy. This evidence is ultimately compatible with standard economic vote expectations, and, one could argue, points to the fact that, to some extent, Trump's election was not *completely* out of the ordinary: sociotropic and pocketbook evaluations, along with demographic variables and party identification, are highly predictive of candidate preferences.[5]

[4]This variable, like in the original paper, is the difference between the Republican and Democratic thermometers, then coarsened to a 20-point scale.

[5]We do not discuss here the potential problems related to the endogeneity of economic perceptions with respect to candidate preferences (Bartels, 2002). These might actually compound, and not ameliorate, the issues we are highlighting in this paper.

In column 6 of Table 1 we augment the specification in column 5 with a cultural control: the stance on immigration policy. This is measured, like in the original paper, as the average of three items on a 5-point scale, with higher values denoting more pro-immigration attitudes. The negative and significant coefficient on this variable suggests that individuals holding more positive views of immigration are less supportive of Trump. At the same time, in this augmented model the coefficient on pocketbook economic evaluations is no longer statistically significant. Patterns like this ultimately lead Mutz (2018) to conclude substantively that status threat, rather than economic discontent, is behind Trump's victory in 2016. Yet, also in this case there is a post-treatment bias concern, as immigration stances could be causally downstream with respect to pocketbook economic evaluations.

In this respect, in column 4 of Table 2 we report the coefficient on pocketbook economic perceptions from a regression of immigration attitudes on all the variables included in the model of column 5 in Table 1. The coefficient on the evaluation of family finances is positive and statistically significant, indicating that respondents with more positive economic perceptions are also more in favor of immigration. That is, the immigration stances of American respondents who are otherwise identical in terms of a rich set of political, economic and demographic controls, are significantly related to pocketbook economic evaluations. Specifically, a one standard deviation improvement in personal economic perceptions (approximately equal to one point in the 5-point scale) is associated with an improvement in immigration stances by around 6% of a standard deviation. Analogously to the first replication exercise of Colantone and Stanig (2018a), even such a relatively weak endogeneity may invalidate inferences from the long regression.

The three examples discussed in this section highlight how post-treatment bias may play a relevant role in empirical studies of the globalization backlash. In particular, they show how the inclusion of post-treatment cultural variables in regressions of vote choice on economic factors might render the coefficient on these factors insignificant. This happens even in cases of relatively weak endogeneity. In this respect, we want to state very clearly that we are not claiming that all (or most) of the variation in cultural variables is driven by economic factors, nor that culture does not play an important *independent* role in affecting voting behavior. Yet, neither of these claims needs to be satisfied for cultural variables to be post-treatment, and therefore act as “bad controls” in regressions of voting behavior on economic factors.

3 Post-treatment bias and the culture-economy nexus

In general, practices that raise concerns of post-treatment bias are often driven by the desire to arrive at a causal effect estimate of a given factor “net” of an alternative explanation. Post-treatment bias emerges when the second explanation is not really an alternative, being itself affected by the initial factor. In this case, “either-or” questions are fundamentally ill-formed and the inclusion of a post-treatment variable does not allow to properly back out any “net” effect. By the same token, one cannot infer mediation effects from changes in regression coefficients for the main factor before and after including a control for a possible mediator. In fact, the desire to assess *how* a cause affects an outcome, and thus the role of one or more mediating factors, is another typical motivation leading to analysis plagued by post-treatment bias.

Importantly, while the idea of post-treatment bias within the potential outcomes framework dates back to the work of Rosenbaum (1984), the pioneers of multiple regression half a century earlier were very aware of one set of intuitions that are central in our discussion. For instance, Frisch and Waugh (1933) are very clear that, in a sense, coefficients *are called into existence* in the moment the regression is specified. Hence the coefficients within a given specification can only be interpreted substantively in the context of the other regressors included. In simple terms, there is no coefficient for “the effect of economic shocks”; there are coefficients for “the effect of economic shocks conditional on whatever else is included in the regression”.

In political science, early criticism of overconditioning in multiple regression was prominently proposed by Achen (2005). We emphasize that in the case of well-identified—e.g., instrumental variable or natural experiment-based—observational studies, controlling for variables that are causally downstream with respect to the main variable of interest implies not estimating well-defined causal quantities. Conversely, the "raw" coefficient from a shorter specification yields an estimate with a causal (albeit obviously still debatable) interpretation. In the context of the globalization backlash, this applies for instance to the inclusion of post-treatment cultural variables in studies of economic drivers of voting behavior.

We illustrate this methodological point through a contrived example based on the machinery of principal stratification (Frangakis & Rubin, 2002), close in spirit to the one presented in Gelman and Hill (2007). In our example we consider a binary "culture" variable that can take two values: every individual in the population can be classified as being either libertarian or authoritarian. We also consider a binary "economic distress" variable: every individual in the population can be classified as either being hit by an economic shock or not. We assume that the economic shock hits half of the population randomly, so that each individual has the same probability of being hit by the shock.[6]

In our hypothetical set-up, for every individual we observe the value of culture (libertarian vs. authoritarian) and whether the individual received or not the economic shock. We also observe a binary variable equal to one if the individual supports a radical-right party. This information allows for the typical horse-race analysis aimed at understanding to what extent the economy and culture "explain" vote for the radical right. In our example, both the economy and culture "matter". Specifically: (1) authoritarian individuals are more likely than libertarians to support the radical-right party; and (2) irrespective of the cultural type, being hit by the economic shock makes all individuals more likely to support the radical-right party. We also allow the economic shock to have an impact on the cultural traits

[6]We use binary variables to make the example tractable. Yet, it can be reproduced with multi-valued and continuous variables, only to the detriment of clarity in terms of intuition.

of individuals; specifically, being hit by the economic shock turns a fraction of otherwise libertarian individuals into authoritarian. The existence of this subset of individuals makes culture post-treatment (or an "intermediate outcome") with respect to the economic shock. Given this set-up, which we consider realistic in light of the available empirical evidence, we show how estimating the effect of the economic shock conditioning on observed culture leads to biased estimates.

Table 3: Hypothetical set-up

	Culture		**Radical-Right Vote**			**Observable**	
	M(0)	M(1)	Y(0)	Y(1)		Y(0, M(0))	Y(1, M(1))
Type of Voter	no shock	shock	no shock	shock	number	no shock	shock
Genuine libertarian	lib	lib	0.2	0.3	α	Y(0,0)	Y(1,0)
Impressionable libertarian	lib	auth	0.3	0.4	γ	Y(0,0)	Y(1,1)
Genuine authoritarian	auth	auth	0.7	0.8	β	Y(0,1)	Y(1,1)

Table 3 illustrates our example in full detail. The first three columns from the left describe the three types of voters in our hypothetical population: (1) genuine libertarians, who remain libertarian even if hit by the economic shock; (2) "impressionable" libertarians, who become authoritarian if hit by the economic shock; and (3) genuine authoritarians, who are always authoritarian irrespective of whether or not they are hit by the economic shock. In technical wording, these types are called "principal strata" and are defined based on the joint potential values of the intermediate variable with and without the treatment. In our set-up, the treatment is exposure to the economic shock. This is denoted by $T \in \{0, 1\}$, a dummy equal to 1 if the individual is hit by the shock. The intermediate variable, which is partially endogenous to the shock, is the cultural trait. This is denoted by $M \in \{lib, auth\}$, a binary variable capturing whether the individual is observed to be libertarian vs. authoritarian. For each individual, we can define a pair $(M(0), M(1))$, where $M(0)$ is the value of culture if the individual is not hit by the shock (i.e., $T = 0$), and $M(1)$ is the value of culture if the individual is hit by the shock (i.e., $T = 1$). These are presented in columns 2-3 of Table 3.[7]

[7]Readers familiar with the Angrist et al. (1996) framework for instrumental variables will notice the parallel

The next two columns report the propensity to vote for the radical-right party for each type of voter, when not exposed to the shock (column 4) and when exposed to the shock (column 5). These two potential outcomes are denoted as $Y(0)$ and $Y(1)$. In line with the assumption that culture matters, without the shock genuine authoritarians have a higher propensity to vote for the radical right than genuine libertarians (0.7 vs 0.2), while impressionable libertarians lay between the two (0.3). In line with the assumption that the economy matters, being hit by the economic shock raises the propensity to support the radical right for all voters. Specifically, we assume that the effect of the shock is equal to 10 percentage points and is constant across the three strata.

The sixth column of Table 3 reports the (normalized) number of each type of voters in the population: α, γ and β. The sum of their shares out of the total population is equal to 1. Importantly, we shall stress that stratum membership at the individual level is always unobserved to the researcher, and so are stratum shares. We only observe whether individuals are libertarian vs. authoritarian, and whether they are shocked or not, but we do not know which stratum they belong to. For instance, if an individual is observed to be authoritarian and shocked, we do not know whether she is a genuine authoritarian or an impressionable libertarian. Similarly, if an individual is observed to be libertarian and non-shocked, we do not know whether she is a genuine or an impressionable libertarian.

Having assumed that the economic shock hits individuals randomly, we can compute the causal effect of the shock on radical-right vote by taking the difference in means between shocked and non-shocked individuals. Given the assumed data-generating process, this would be equal to 0.1. Analogously, we can estimate a regression on a dummy variable for whether the individual was shocked to back out an unbiased estimate of the effect of the shock on radical-right vote. Problems arise when one tries to estimate the effect of the

between the three principal strata here and the notions of "never-takers", "compliers", and "always-takers" in the discussion of IV estimation of Local Average Treatment Effects (LATE). The parallel is not a coincidence (see Frangakis and Rubin, 2002). In principle, one could also think of a fourth stratum: individuals who are authoritarian when not shocked and libertarian when shocked. The assumption by which this stratum is empty is a monotonicity assumption analogous to the one made in the framework of LATE estimation.

economic shock while controlling for culture. In what follows, we show that this does not lead to an unbiased estimate of the effect of the economic shock. At the same time, it does not deliver an estimate of the role of the economy "net" of culture, nor it can be informative of the possible mediation effect of culture.

Given the binary nature of the explanatory variables, in our set-up "controlling for culture" involves three steps: (1) computing the difference in means for shocked vs. non-shocked individuals who display authoritarian orientations, i.e., the effect of the shock among observed authoritarians; (2) computing the difference in means for shocked vs. non-shocked among individuals who display libertarian orientations, i.e., the effect of the shock among observed libertarians; and then (3) averaging these two effects using the shares of observed authoritarians and libertarians in the population.

Let us start with step 1, focusing on individuals with observed authoritarian attitudes. Authoritarians who are not hit by the shock can only be genuine authoritarians. According to column 4 in Table 3, their propensity to vote for the radical-right party, denoted as $Mean^0_{auth}$, is equal to 0.7. Observed authoritarians who are hit by the shock can be either genuine authoritarians or impressionable libertarians. In other words, the group of treated among observed authoritarians includes individuals from two different strata. The average support for the radical right we could empirically back out from observed data is equal to the weighted average of the vote propensity for treated individuals in the two groups, where weights are given by the group sizes. Using information in columns 5-6 of Table 3, this can be expressed as $Mean^1_{auth} = \frac{0.8\beta+0.4\gamma}{\beta+\gamma}$. The estimated effect of the economic shock among authoritarians, denoted as θ_{auth}, is then equal to the difference of means between treated and control authoritarians: $Mean^1_{auth} - Mean^0_{auth}$. Hence: $\theta_{auth} = \frac{0.8\beta+0.4\gamma}{\beta+\gamma} - 0.7$.

From the assumptions that we made in our hypothetical set-up, we know that the true effect of the economic shock is equal to 0.1 for all individuals in the population, irrespective of their cultural traits. Hence, it is easy to see how θ_{auth} provides an unbiased estimate of the true effect for observed authoritarians only if γ is equal to zero, i.e., there are no impressionable individuals whose observed culture is affected by the shock. When γ is greater than zero—and therefore culture is post-treatment with respect to the economic shock—θ_{auth} departs from the true value and can become negative if the number of impressionable voters is high enough (i.e., greater than $\frac{\beta}{3}$ in our example). For instance, assuming that there are 400 impressionable libertarians and 200 genuine authoritarians in the population, of which, respectively, 200 and 100 are treated, $\theta_{auth} = \frac{0.8*100+0.4*200}{100+200} - 0.7 = 0.53 - 0.7 = -0.17$. This is very different from the true effect of the economic shock (0.1). The bias emerges because this comparison of means among individuals who display authoritarian attitudes is not estimating any well-defined causal quantity, as it mixes individuals in different strata.

Analogous considerations apply to step 2, where we focus on observed libertarians. Observed libertarians who are hit by the economic shock can only belong to the stratum of genuine libertarians. According to column 5 in Table 3, their propensity to vote for the radical-right party, denoted as $Mean^1_{lib}$, is equal to 0.3. Observed libertarians who are not hit by the shock can belong either to the stratum of genuine libertarians or to the stratum of impressionable libertarians. Hence, this group of individuals spans two different strata. The average support for the radical right that would be empirically backed out from observed data is equal to: $Mean^0_{lib} = \frac{0.2\alpha+0.3\gamma}{\alpha+\gamma}$. The estimated effect of the economic shock among libertarians, denoted as θ_{lib}, is equal to $Mean^1_{lib} - Mean^0_{lib}$. Hence: $\theta_{lib} = 0.3 - \frac{0.2\alpha+0.3\gamma}{\alpha+\gamma}$. As in the case of observed authoritarians, θ_{lib} is equal to the true value of the effect (0.1) only if γ is equal to zero, i.e., there are no impressionable individuals in the population. In the numerical example introduced above, if there are 400 impressionable libertarians and 200 genuine libertarians, of which, respectively, 200 and 100 are hit by the shock, $\theta_{lib} = 0.3 - \frac{0.2*100+0.3*200}{100+200} = 0.3 - 0.27 = 0.03$. Also for the group of observed libertarians, we obtain a biased estimate.

Step 3 involves taking the average of the effects obtained for authoritarians and libertarians, using their respective shares in the population as weights. Having assumed that

50% of individuals are randomly hit by the shock, the number of observed libertarians is equal to $\alpha + \frac{\gamma}{2}$, and the number of observed authoritarians is equal to $\beta + \frac{\gamma}{2}$. Weights are then obtained as their ratios over total population $(\alpha + \beta + \gamma)$. Since the effects obtained at steps 1 and 2 are biased, their weighted average is also going to be biased. In our numerical example, with 200 genuine libertarians, 400 impressionable libertarians, and 200 genuine authoritarians, there would be 400 observed libertarians and 400 observed authoritarians. The overall estimate of the economic shock, denoted by θ_{all}, would then be the simple average of the effects for the two groups. That is: $\theta_{all} = \frac{\theta_{lib}+\theta_{auth}}{2} = \frac{0.03-0.17}{2} = -0.07$. This is biased—and with the opposite sign compared to the true effect of 0.1.

The final outcome of step 3, -0.07, is equivalent to the estimated coefficient that we would obtain from a regression of voting on the economic shock while controlling for culture, i.e., the estimated effect of the economic shock in a horse-race regression. To show this point, we complement the above computations with regressions on simulated data. Specifically, we generate 1,000 different samples with 800 individuals each. The data-generating process is based on the assumptions of our hypothetical set-up, as summarized in Table 3. The partition of individuals across the three strata is as in the numerical example outlined above (i.e., 200 genuine libertarians, 400 impressionable libertarians, and 200 genuine authoritarians). For each sample, we estimate two regressions. In the first one, we regress the dummy for radical-right vote on the dummy for the economic shock, without controlling for culture. In the second regression, we add the control for culture, i.e., a dummy equal to one if the individual is observed to be authoritarian. Table 4 reports the average estimated coefficients out of the 1,000 regressions for both specifications, along with the average standard errors. The average estimated coefficient on the economic shock variable is equal to the true effect (0.1) in column 1, where we do not control for culture. It is instead biased in column 2, where the control for culture is included. The average point estimate in this case is in fact equal to the value obtained at step 3 in the above computations: -0.07.

In a horse race approach, these empirical results would be read as evidence that "economic factors do not matter". Yet, drawing this conclusion would be incorrect, as we know from the data-generating process that the economic shock has a positive effect on the propensity to vote for the radical-right party. Since culture is post-treatment, conditioning on it in the regression leads to biased estimates of the effect of the economic shock.

Table 4: Regressions on simulated data

	(1)	(2)
Dep. var.:	**Radical-Right Vote**	
Import Shock	0.10	-0.07
	[0.03]	[0.04]
Culture		0.34
		[0.04]
Obs.	800	800
N. of samples	1,000	1,000

Notes: Coefficients and standard errors (in brackets) are averages across 1,000 samples, each containing 800 observations.

3.1 Mediation

There is a second type of conclusion that researchers tend to draw in a situation like the one described above. That is that culture, the post-treatment variable, fully "mediates" or "channels" the effect of the economic shock on voting behavior. According to this view, full mediation explains why, once the control for the mediator is included in the specification, one cannot retrieve a significant positive effect for the economic shock. Drawing such a conclusion in our set-up would be incorrect. In fact, we know from the data-generating process that exposure to the economic shock changes the cultural trait only for a stratum of the population, i.e., the impressionable libertarians. For all others, i.e., genuine libertarians and genuine authoritarians, exposure to the shock raises the propensity to vote for the radical-right party without any change in culture. Hence, we can rule out that the effect of the shock on voting is fully mediated by culture.

How much of the overall effect of the shock is then mediated by culture in our example? In what follows, we address this question in the framework of causal mediation analysis. In particular, relying on the framework of Frangakis and Rubin (2002) and Forastiere et al. (2018), we discuss how quite demanding assumptions are required not only to estimate, but even to just define mediation effects. In this respect, the same underlying problems that plague the horse race approach also constitute a threat to valid mediation analysis.

Generally speaking, it is tempting to believe that the machinery of causal mediation analysis (e.g., Imai et al., 2011) can be leveraged to shed light on the complex causal structure underlying voting behavior, getting around the obstacle posed by post-treatment bias. Indeed, mediation analysis allows in principle to disentangle direct and indirect effects of a given treatment, with the latter being "transmitted by a mediator". However, "modern" mediation analysis relies on quite demanding assumptions whose plausibility may often be problematic when dealing with voting behavior. To be clear, the proponents of causal mediation analysis are very transparent on the importance of these assumptions, with Imai et al. (2011) being a notable example. The older literature on mediation in the psychometrics tradition (Baron & Kenny, 1986; Wu & Zumbo, 2008) also asked many important questions about causal order in relation to valid mediation analysis (e.g., Smith 1982). Yet, for practitioners, the substantive implications of these assumptions may be somewhat difficult to visualize. Key assumptions may then be perceived as being merely technical, leading to applications of the causal mediation approach in contexts where it is not warranted. In our view, the search for the causes of radical-right and populist parties' success, with the interplay between economics and culture discussed above, may be a case in point.

We develop our discussion within the framework of *principal ignorability*, as in Forastiere et al. (2018). In this framework, the total effect of the treatment (TE) can be decomposed into a Natural Direct Effect (NDE) and a Natural Indirect Effect (NIE). The natural direct effect is the average treatment effect fixing the mediator at the level it would have taken in the absence of the treatment. The natural indirect effect is the average effect of the change in the mediator that would be induced by the treatment, but holding treatment status constant.[8] In our example, the natural direct effect refers to the effect of being exposed to the economic shock while fixing culture to its stratum-specific value in the absence of the shock; the natural indirect effect is the change in the voting behavior of an individual who adopts the culture she would display if she were hit by the economic shock, although she is in fact not hit by the economic shock.

To illustrate, let us introduce some additional notation. Observed and potential outcomes are expressed as functions of two arguments: the value of the treatment, and the value of the mediator. We write $Y(T, M)$ to indicate the value of the outcome under treatment T and mediator M. In our example, for instance, $Y(1, 1)$ is the propensity to support the radical right for someone who received the economic shock and displays authoritarian attitudes; $Y(1, 0)$ is the propensity to support the radical right for someone who received the economic shock and does not display authoritarian attitudes. The last two columns of Table 3 illustrate them in our example. In the second last column, for non-shocked individuals, the value of the treatment is always zero, while the indicator for culture is equal to zero in the first two strata (i.e., observed libertarians), and to one in the third stratum (i.e., observed authoritarians). In the last column, for shocked individuals, the value of the treatment is always one, while the indicator for culture is equal to zero in the first stratum, and to one in the other two strata.

The total effect of the treatment can be written as $TE = E\left(Y(1, M(1)) - Y(0, M(0))\right)$, where $M(1)$ is the value of the mediator under treatment, and $M(0)$ is the value of the mediator under no treatment, and the expectation is taken over the total population. The natural direct effect is defined as $NDE = E\left(Y(1, M(0)) - Y(0, M(0))\right)$, where the mediator is fixed to its value under no treatment. The natural indirect effect is the difference between the two: $NIE = TE - NDE = E\left(Y(1, M(1)) - Y(1, M(0))\right)$. This simple formulation of

[8]We make a “no-interaction” assumption throughout to keep the exposition simpler.

the approach reveals the fundamental difficulty in the estimation of natural effects. That is, it involves comparing quantities that are not observed and, in some cases, are inherently unobservable.

For instance, consider the natural direct effect. This is the causal effect of the treatment on the outcome if the mediator stays at the level it would have assumed in the absence of the treatment. In our substantive example, this has to do with the change in the propensity to support the radical right when exposed to the economic shock, but retaining the cultural orientations of individuals who are not exposed to the shock. The crucial term is therefore $Y(1, M(0))$. In our example, $M(0) = M(1)$ for genuine libertarians (i.e., first stratum) and genuine authoritarians (i.e., third stratum). Indeed, their culture stays the same irrespective of whether they are hit or not by the economic shock. Frangakis and Rubin (2002) define such strata as "disassociative". For them, we can rewrite the natural direct effect as $NDE = E\left(Y(1, M(1)) - Y(0, M(0))\right)$. Clearly, we then have that $NDE = TE$, and therefore $NIE = 0$. Very intuitively, if culture does not change with the treatment, there is no mediation effect, thus the total effect is just equal to the direct effect.

In our example, mediation can only operate via the stratum of impressionable libertarians, whose culture changes depending on whether they are hit or not by the economic shock. Frangakis and Rubin (2002) define such a stratum as "associative". For these individuals we know that $M(0) = 0$, as they are observed as libertarian in the absence of the shock. Conversely, $M(1) = 1$, as they become authoritarian when shocked. Based on this information, we can retrieve the total effect as $TE = E\left(Y(1, M(1)) - Y(0, M(0))\right)$. However, in order to decompose this effect into natural direct and indirect effects, we still need the term $Y(1, M(0))$. That is, for this stratum, the outcome with the economic shock but libertarian cultural traits $Y(1, 0)$. This is something we can never observe for individuals belonging to this stratum. Indeed, once they are hit by the shock, all these individuals are observed as authoritarian. Frangakis and Rubin (2002) refer to such objects as *a priori* counterfactuals, which differ from the standard counterfactuals involved in casual estimation because they are *never* observable. The only way to make progress in terms of causal mediation analysis is by making assumptions on such *a priori* counterfactuals. Specifically, in our example, we need to make an assumption on $Y(1,0)$ for impressionable libertarians. Depending on the assumption we make, we are going to *define* mediation in terms of a specific counterfactual, and we are going to obtain measures of direct and indirect effects that crucially hinge upon the specific underlying assumption.

We develop our analysis along what Forastiere et al. (2018) call generalized *weak* principal ignorability. This hinges on assumptions of homogeneity of the potential outcomes across principal strata. Intuitively, these assumptions make it possible to define *a priori* counterfactuals based on observable outcomes of other strata. Focusing on our example, weak principal ignorability involves an assumption with two components. The first component is that $Y(1,1)$, the outcome under treatment for observed authoritarians, is the same for genuine authoritarians (i.e., stratum 3) and for impressionable libertarians (i.e., stratum 2).[9] Note that both values are defined in our example, as they are $Y(1,M(1))$ for the two strata. The assumption is needed because we are not able to differentiate, in the data, which $Y(1,1)$ observations (i.e., observed authoritarian and exposed to the economic shock) come from the genuine authoritarian stratum and which belong to the impressionable libertarian stratum. In our example, such homogeneity requirement would be violated. In fact, in the data-generating process, impressionable libertarians, conditional on receiving the shock, show a lower propensity to vote for the radical right than genuine authoritarians: 0.4 vs. 0.8, respectively. This is far from unrealistic as a choice for the data-generating process, yet it would make causal mediation analysis problematic.

The second component of the assumption is also potentially problematic. It requires that $Y(1,0)$, the outcome under the treatment when the mediator is equal to zero, is the same for genuine libertarians (i.e., stratum 1) and for impressionable libertarians (i.e., stratum

[9] In general, what is required is that the *distribution* of potential outcomes is the same across principal strata. For simplicity, and without loss of generality, in our example we assume that potential outcomes—in the form of propensity to support the radical right—are constant within strata.

2). For genuine libertarians, who never change culture, $Y(1,0)$ is simply the observed outcome under treatment: $Y(1,M(1))$. For impressionable libertarians, $Y(1,0)$ is never observed, as they all switch culture when exposed to the shock. As discussed above, this is an *a priori* counterfactual. The weak principal ignorability assumption entails attributing to this counterfactual an observable value taken from a different stratum, that of genuine libertarians. This is crucial as it defines the counterfactual against which mediation is evaluated.

In substantive terms, defining mediation requires us to answer the following question for impressionable libertarians: how would an individual in this stratum vote if she had been exposed to the shock but for some reason kept the culture of someone in the same stratum who was not exposed to the shock? Under the weak principal ignorability assumption, the voting behavior of individuals exposed to the shock, and who became authoritarian for this reason, would be assumed to be the same as that of individuals who were exposed to the shock but, being "genuine libertarians", did not become authoritarian. This is arguably not a wildly heroic assumption, but it is not an innocuous or merely technical assumption either: it is in fact loaded with substance. Indeed, it is not hard to imagine that those voters who do not become authoritarian in the face of an economic shock are a "different type of people" in terms of political behavior than those who do.

If we are ready to make this assumption, we can compute the natural direct effect for the stratum of impressionable libertarians. In formula, $NDE = E(Y(1,M(0)) - Y(0,M(0))$. From Table 3, we know that $Y(0,M(0))$, which is equal to $Y(0,0)$ for this stratum, is 0.3. If principal ignorability holds, $Y(1,M(0)) = Y(1,0)$ for the second stratum is equal to $Y(1,0)$ for the first stratum, which is 0.3 in the example. Hence, the natural direct effect is equal to zero. By the same token, the natural indirect effect is equal to the total effect. In other words, under the assumption of weak principal ignorability, for the group of impressionable libertarians in our example all the effect of the economic shock is mediated by the change in culture. In aggregate, then, the larger the share of impressionable libertarians in the population, the larger the share of the overall effect of the economic shock that is attributed to mediation by culture. It is important to underline, though, that this result heavily hinges upon the assumption made on the *a priori* counterfactual. To give an idea of what it entails, under this assumption impressionable libertarians would be the only group in the population that do not show a response to the economic shock independently of culture.

The bottom line of this illustration is that causal mediation analysis is not an obvious option for studies of voting behavior in contexts where post-treatment bias is an issue. Researchers who intend to engage in this type of analysis should be aware of the implications of the underlying assumptions, that are fundamentally relevant for how mediation itself is defined in the first place. For instance, the data-generating process in our example is constructed in a reasonable way, consistent with the available empirical evidence, yet we have seen how causal mediation analysis would be problematic in this context. This calls for caution among researchers working on economic vs. cultural drivers of populism. Beyond that, we suspect that in many settings applied scholars might be uncomfortable with the assumptions underlying causal mediation, once their substantive implications are more carefully spelled out. Indeed, heterogeneous potential outcomes across strata of the population are likely to be quite common in political science applications. In this respect, a promising way forward is suggested by Ferrari et al. (2021), who apply, in the context of economic vs. cultural drivers of extremism, a recently developed Bayesian methodology to identify latent clusters (hdpGLM, Ferrari, 2020). This allows to estimate heterogeneous causal effects across different clusters of individuals, for which the role of the mediating factor may be more or less important.

For ease of exposition, we have developed our discussion in the framework of principal ignorability. In Section C of the Online Appendix, we show how similar conclusions can be reached when working within the sequential ignorability framework of Imai et al. (2011).[10]

[10]Casting the main discussion in terms of principal ignorability has two main advantages: (1) it is arguably easier to visualize and evaluate against substantive knowledge; and (2) it makes it possible to work through our example without invoking additional unobserved confounders. The strong version of principal ignorability (along with a monotonicity assumption like the one we have made throughout the example) implies sequential

The latter contribution suggests strategies involving sensitivity analysis to check whether violations of the sequential ignorability assumption lead to invalid inference about mediation effects. In our example, given the way in which it has been constructed, sensitivity analysis would certainly not provide encouraging answers.

As a final remark, connecting the two parts of the methodological section, we underline how the problems with causal mediation analysis stem from the very same reason that impedes the unbiased estimation of the effect of the economic shock while controlling for culture. That is that principal strata are defined by how their members respond—in terms of the mediator—to the treatment. Post-treatment bias may be seen as a consequence of heterogeneity of potential outcomes. For instance, when we "control for culture" in our example regression, we are acting as if homogeneity of the potential outcomes across strata held. The potential outcomes, though, are not homogeneous, and this leads to biased estimates even if the causal effect is constant across strata. If we observed the stratum membership of each individual (which is a stable feature, unaffected by the treatment), we could control for it and back out the correct causal effect of the treatment. Within each stratum, in fact, the mean difference between treated and controls would provide a valid estimate of the treatment effect for the stratum. Yet, stratum membership is unobserved. Moreover, even if it was observed, causal mediation analysis would still require the demanding second assumption of weak principal ignorability.

4 Novel observational evidence

In this section, we provide novel observational evidence on the potential pervasiveness of post-treatment bias in studies of the globalization backlash that investigate the role of trade exposure as an economic factor. Specifically, based on individual-level survey data, we study the effect of exposure to import competition from China on a large array of cultural

ignorability; the weak version of principal ignorability that we employ is less stringent than sequential ignorability, albeit sufficient to back out mediation effects (Forastiere et al., 2018).

attitudes. Expanding on earlier findings by Ballard-Rosa et al. (2022), Ballard-Rosa et al. (2021), Ferrara (2022), Ferrari et al. (2021) and Hays et al. (2019), we show that trade exposure triggers individual reactions in terms of several cultural attitudes, which should then be considered post-treatment controls in long vote regressions.

We focus on fifteen western European countries, over the period 1995-2008.[11] We provide full details on the empirical exercise in Section D of the Online Appendix. Exposure to Chinese imports is computed at the regional level, and instrumented using Chinese exports to the United States, as in Colantone and Stanig (2018b). Individual-level data are sourced from the European Social Survey (ESS, 2002, 2004, 2006, 2008) and the European Value Study (EVS, 2020). As outcome variables we consider ten cultural attitudes belonging to four main groups: (1) "meta-political" attitudes about liberal democracy; (2) private authoritarian attitudes; (3) traditional conservatism; and (4) immigration attitudes. All variables are coded so that higher values correspond to more undemocratic, authoritarian, conservative, and nativist stances. Some of these variables are available in the ESS sample, others in EVS. Thus we run separate regressions for the two samples.

We estimate specifications of this form:

$$Cultural\ Attitude_{icrt} = \alpha_{ct} + \beta_1\ Import\ Shock_{cr(i)t} + \boldsymbol{Z_{it}}\gamma' + \epsilon_{icrt} \tag{1}$$

where i indexes individuals, c countries, r regions and t years. $r()$ is a function that maps individual i to her NUTS-2 region of residence r, allowing to attribute to each individual the relevant import shock ($Import\ Shock_{cr(i)t}$). This is computed over the two years prior to the year of the interview. Finally, Z_{it} is a vector of (plausibly) pre-treatment controls for individual characteristics, containing age, gender, and dummies for educational attainment.

Figure 1 reports IV estimates of Equation 1, in which the ten different cultural attitudes are regressed on import competition. The majority of estimated coefficients are statistically

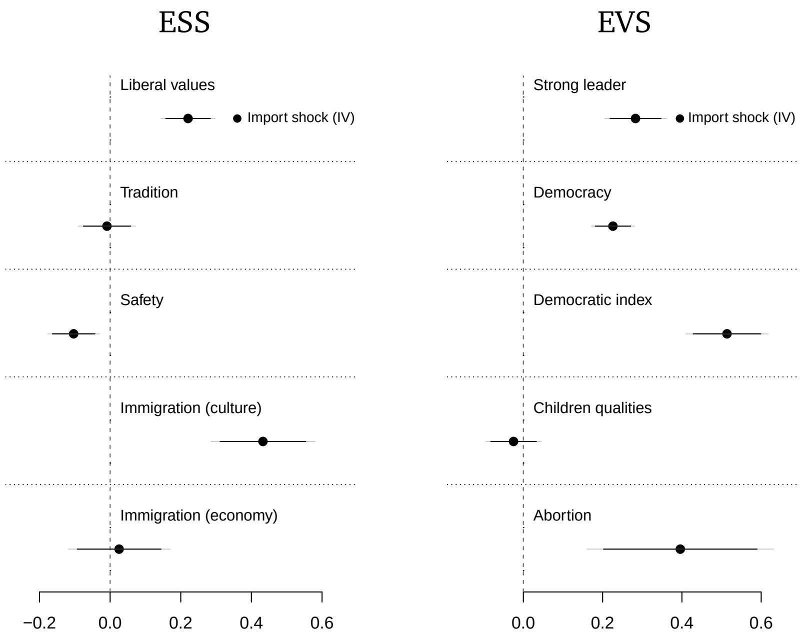
[11] The sample includes: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

significant in the expected direction. In particular, concerning the group of meta-political attitudes, respondents residing in regions more exposed to the import shock tend to be more sympathetic with the idea of unconstrained strong leaders, and less unequivocally supportive of democracy and liberal values than otherwise similar individuals residing in areas less exposed to the shock. As for conservative and private authoritarian traits such as attitudes about child rearing, and the importance of traditions and living in safe surroundings, there is no detectable positive association with the import shock. The only form of conservatism that seems to be significantly reinforced by economic distress captured by the import shock is the stance on abortion. In fact, individuals from areas more exposed to Chinese competition tend to be less permissive with respect to abortion than similar individuals from less exposed regions.

Finally, the relation between trade exposure and attitudes toward immigration depends on the dimension considered. The estimated coefficient on the effect of import competition on the assessment of whether immigration is good for the national economy is not statistically different from zero. Conversely, the estimated coefficient on the import shock on the assessment of whether immigration is good for the national cultural life is positive and significant. All else equal, respondents who reside in regions that received stronger import shocks tend to be more concerned with the "cultural threat" posed by immigrants. This suggests that economic distress, as driven by trade exposure, does not necessarily affect only attitudes toward "material" dimensions, but may lead individuals to perceive a threat to their social status from a cultural point of view.

To sum up, our results indicate that trade exposure may have significant effects on an array of cultural attitudes. These findings add to a growing body of evidence that documents the impact of economic factors on cultural traits (e.g., Anelli et al., 2021; Ballard-Rosa et al., 2022; Ballard-Rosa et al., 2021; Carreras et al., 2019; Ferrara, 2022; Ferrari et al., 2021 and Hays et al., 2019). Overall, this evidence suggests that demands for cultural protectionism, as well as appeals to ethnic or racial superiority, cannot be interpreted at

Figure 1: Cultural attitudes are affected by the import shock



Note: The bars correspond to 95% and 90% confidence intervals. If the interval crosses the dashed vertical line, the null hypothesis of no relationship cannot be rejected. Full results can be found in Tables D.1 and D.2 of the Online Appendix.

face value as consequences of a concern for "culture" however defined. Rather, they seem to be, at least partly, the cultural manifestation of grievances that are driven by situations of economic distress. From a methodological perspective, this body of available evidence warns about the pervasiveness of post-treatment bias in vote regressions that include jointly both economic and cultural factors.

5 Conclusion

We address the methodological issue of post-treatment bias in the context of studies of voting behavior, with specific focus on the globalization backlash. According to the available empirical evidence, both economic and cultural factors contribute to determine the backlash, and they significantly interplay with each other. That is, exposure to economic shocks may affect individual cultural attitudes, and cultural concerns may raise the political salience of economic shocks. Such an interplay poses methodological challenges in terms of post-treatment bias.

We make three main methodological points. First, if and insofar as cultural variables are post-treatment with respect to economic factors, the estimates of the effect of economic shocks on voting are biased in regressions that include cultural controls (and vice versa). Second, for the same reason, such horse-race regressions do not allow to accurately estimate the relative role of economic vs. cultural factors. Third, one cannot infer mediation effects from changes in regression coefficients for a given factor of interest before and after including post-treatment controls. We accompany the methodological discussion with empirical evidence on the relevance of post-treatment bias in studies of the globalization backlash, both by replicating and expanding on earlier studies, and by presenting novel cross-country results on the culture-economy nexus.

Crucially, we show how even relatively weak endogeneity of cultural variables with respect to economic factors may invalidate inferences from long regression approaches where both economic and cultural variables are included jointly. In our view, there is a sharp discontinuity here. The issue is less about the strength of "post-treatmentness" in a specific empirical application than it is about research design and model specification. Without post-treatment controls, a plausibly-identified piece of evidence might have a causal interpretation. With some amount of influence of the "treatment" variable on an intermediate outcome included as control, that interpretation is no longer warranted. In other words, if the "treatment" is (plausibly) exogenous, the short regression design yields a credible causal estimate. Conversely, a long model including post-treatment controls is *ex ante* not causally identified.

Importantly, we develop our analysis around identifying the effect of economic factors, with cultural variables being post-treatment. However, exactly the same considerations can be made for studies of cultural factors, where economic controls may be post-treatment. To make an example, assume that low-level personality traits can be considered, conditional on some background variables, as good as randomly assigned. A regression of vote choice on these traits would then plausibly identify the effect of personality on voting behavior. Now, imagine augmenting the regression with a control for an economic variable, say the individual's occupation (e.g., through dummies corresponding to each occupational code). Given that people might select into occupations also based on their personality (Graziano et al., 2012; Kitschelt & Rehm, 2014; McKay & Tokar, 2012), this would be a post-treatment control. The augmented regression would then not estimate the effect of personality "net of occupation"; conversely, it would estimate quantities that do not have any well-defined causal interpretation.

How should researchers proceed, then, when confronted with a situation in which various relevant factors interplay with each other? For instance, how should we frame studies of the globalization backlash when considering economic vs. cultural drivers? The main indication emerging from our study is that controlling for a post-treatment variable leads to biased estimates on the main factor. Hence, a study that focuses on the causal effect of an economic factor, without controlling for culture, might be *better* than one that does control for culture, and the same applies symmetrically for the study of cultural drivers. In other words, in the logic of causal empiricism (Samii, 2016), the best way to further our understanding of the phenomenon might be to study, in a principled causal framework, one potential cause at a time. As highlighted by Gelman and Imbens (2013), Franzese (2019) and Frieden (2022), this may also be a sensible way to inform policy action. For instance, dismissing the economic roots of the backlash—and leaving them unaddressed by policy—based on empirical analyses that are questionably-specified may be very dangerous.

Yet, the question concerning the interplay, and the relative importance, of economic vs. cultural factors is also relevant, both theoretically and empirically. In this respect, possible inroads may be made by triangulating the results of different specifications, by deploying sensitivity analysis, and through causal mediation approaches (Imai et al., 2011). Yet, we have discussed how these approaches hinge on demanding assumptions whose tenability may be problematic in analytical contexts such as the one we consider in this paper. In any case, when these assumptions are transparently spelled out and theoretically reasonable, proper causal mediation analysis is strictly superior to analyses that try to infer about mechanisms by comparing coefficients in short vs. long regressions. The recent contribution by Ferrari et al. (2021) is also proposing ways to study the heterogeneity of mediation effects across different types of individuals in the population, which may be crucial in this context.

Alternatively, if one wanted to allocate the effects of culture and economic drivers within the same model, a promising option is provided by structural equation modeling. This involves specifying a full theoretical model with possibly demanding, but transparent, assumptions about functional forms, exclusions, and inclusions (e.g., Achen, 2002). Equations directly derived from the theoretical model, via assumptions about the sources of randomness, are then fit to the data. This approach might be a useful way forward to address "big picture" questions like the relative weights of different factors in vote choice. What is key to keep in mind, though, is that the simple long regression approach, where all factors are included jointly, is not, in general, an approximation to a structural model (Reiss & Wolak, 2007). To the contrary, its estimates might be highly misleading regarding the theoretical quantities of interest.

References

Acharya, A., Blackwell, M., & Sen, M. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review*, *110*(3), 512–529.

Achen, C. H. (2002). Toward a new political methodology: Microfoundations and art. *Annual Review of Political Science*, *5*(1), 423–450.

Achen, C. H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, *22*(4), 327–339.

Agnolin, P., Colantone, I., & Stanig, P. (2024). *Replication Data for: In search of the Causes of the Globalization Backlash*. https://doi.org/10.7910/DVN/MKJYV3

Anduiza, E., & Rico, G. (2023). Sexism and the far-right vote: The individual dynamics of gender backlash. *American Journal of Political Science*, *68*(2), 478–493. https://doi.org/https://doi.org/10.1111/ajps.12759

Anelli, M., Colantone, I., & Stanig, P. (2021). Individual vulnerability to industrial robot adoption increases support for the radical right. *Proceedings of the National Academy of Sciences*, *118*(47). https://doi.org/10.1073/pnas.2111611118

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444–455.

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. *Princeton University Press*.

Autor, D., Dorn, D., Hanson, G., & Majlesi, K. (2020). Importing political polarization? The electoral consequences of rising trade exposure. *American Economic Review*, *110*(10), 3139–83. https://doi.org/10.1257/aer.20170011

Autor, D., Dorn, D., & Hanson, G. H. (2021). *On the persistence of the China shock* (tech. rep.). Brookings Papers on Economic Activity.

Ballard-Rosa, C., Jensen, A., & Scheve, K. (2022). Economic decline, social identity, and authoritarian values in the United States. *International Studies Quarterly*, *66*(1). https://doi.org/10.1093/isq/sqab027

Ballard-Rosa, C., Malik, M. A., Rickard, S. J., & Scheve, K. (2021). The economic origins of authoritarian values: Evidence from local trade shocks in the United Kingdom. *Comparative Political Studies*, *54*(13), 2321–53. https://doi.org/10.1177/00104140211024296

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173.

Barone, G., D'Ignazio, A., de Blasio, G., & Naticchioni, P. (2016). Mr. Rossi, Mr. Hu and politics: The role of immigration in shaping natives' voting behavior. *Journal of Public Economics*, *136*(100), 1–13.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*, 117–150.

Carreras, M., Carreras, Y. I., & Bowler, S. (2019). Long-term economic distress, cultural backlash, and support for Brexit. *Comparative Political Studies*, *52*(9), 1396–1424. https://doi.org/10.1177/0010414019830714

Cavaille, C., & Marshall, J. (2019). Education and anti-immigration attitudes: Evidence from compulsory schooling reforms across western Europe. *American Political Science Review*, *113*(1), 254–263.

Chan, T. W., Henderson, M., Sironi, M., & Kawalerowicz, J. (2020). Understanding the social and cultural bases of Brexit. *The British Journal of Sociology*, *71*(5), 830–851.

Clayton, K., Ferwerda, J., & Horiuchi, Y. (2021). Exposure to immigration and admission preferences: Evidence from France. *Political Behavior*, *43*(1), 175–200.

Colantone, I., Ottaviano, G. I., & Stanig, P. (2022). The backlash of globalization. *In G. Gopinath, E. Helpman, and K. S. Rogoff (Eds), Handbook of International Economics (Vol. V)*, 405–477.

Colantone, I., & Stanig, P. (2018a). Global competition and Brexit. *American Political Science Review*, *112*(2), 201–218.

Colantone, I., & Stanig, P. (2018b). The trade origins of economic nationalism: Import competition and voting behavior in western Europe. *American Journal of Political Science*, *62*(4), 936–953.

Dustmann, C., Vasiljeva, K., & Piil Damm, A. (2019). Refugee migration and electoral outcomes. *The Review of Economic Studies*, *86*(5), 2035–2091.

ESS. (2002). Ess round 1: European social survey round 1 data (2002) (6.6) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS1-2002].

ESS. (2004). Ess round 2: European social survey round 2 data (2004) (3.6) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS2-2004].

ESS. (2006). Ess round 3: European social survey round 3 data (2006) (3.7) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS3-2006].

ESS. (2008). Ess round 4: European social survey round 4 data (2008) (4.5) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS4-2008].

Evans, G., Fieldhouse, E., Green, J., Schmitt, H., van der Eijk, C., Mellon, J., & Prosser, C. (2016). British election study internet panel wave 8 (2016 eu referendum study, daily campaign survey). https://doi.org/10.5255/UKDA-SN-8202-2

EVS. (2020). EVS (2020): European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008) [ZA4804 Data file Version 3.1.0, https://doi.org/10.4232/1.13486].

Ferrara, F. M. (2022). Why does import competition favor Republicans? Localized trade shocks and cultural backlash in the US. *Review of International Political Economy*, *115*, 1–24.

Ferrari, D. (2020). Modeling context-dependent latent effect heterogeneity. *Political Analysis*, *28*(1), 20–46.

Ferrari, D., Franzese, R., Jackson, H., Wai, R., Wu, P., Kim, B., Kim, W., Pollack, E., & Sanders, H. (2021). *How socioeconomic malaise & decline fuel xenophobic nationalist extremism* [Mimeo, University of Michigan].

Forastiere, L., Mattei, A., & Ding, P. (2018). Principal ignorability in mediation analysis: Through and beyond sequential ignorability. *Biometrika*, *105*(4), 979–986.

Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, *58*(1), 21–29.

Franzese, R. J. J. (2019). The comparative and international political economy of anti-globalization populism. *Oxford Research Encyclopedia of Politics*.

Frieden, J. (2022). Attitudes, interests, and the politics of trade: A review article. *Political Science Quarterly*, *137*(3), 569–588.

Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, 387–401.

Gallego, A., & Kurer, T. (2022). Automation, digitalization, and artificial intelligence in the workplace: Implications for political behavior. *Annual Review of Political Science*, *25*(1), 463–484. https://doi.org/10.1146/annurev-polisci-051120-104535

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. *Cambridge University Press*.

Gelman, A., & Imbens, G. (2013). *Why ask why? Forward causal inference and reverse causal questions* (Working Paper No. 19614). National Bureau of Economic Research. https://doi.org/10.3386/w19614

Gidron, N., & Hall, P. A. (2017). The politics of social status: Economic and cultural roots of the populist right. *The British Journal of Sociology*, *68*(S1), S57–S84. https://doi.org/https://doi.org/10.1111/1468-4446.12319

Graziano, W. G., Habashi, M. M., Evangelou, D., & Ngambeki, I. (2012). Orientations and motivations: Are you a “people person,” a “thing person,” or both? *Motivation and Emotion*, *36*, 465–477.

Hangartner, D., Dinas, E., Marbach, M., Matakos, K., & Xefteris, D. (2019). Does exposure to the refugee crisis make natives more hostile? *American Political Science Review*, *113*(2), 442–455. https://doi.org/10.1017/S0003055418000813

Hays, J., Lim, J., & Spoon, J.-J. (2019). The path from trade to right-wing populism in Europe. *Electoral Studies*, *60*, 1020–38. https://doi.org/https://doi.org/10.1016/j.electstud.2019.04.002

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, *105*(4), 765–789.

Kitschelt, H., & Rehm, P. (2014). Occupations as a site of political preference formation. *Comparative Political Studies*, *47*(12), 1670–1706.

Margalit, Y. (2019). Economic insecurity and the causes of populism, reconsidered. *Journal of Economic Perspectives*, *33*(4), 152–70. https://doi.org/10.1257/jep.33.4.152

McKay, D. A., & Tokar, D. M. (2012). The HEXACO and five-factor models of personality in relation to RIASEC vocational interests. *Journal of Vocational Behavior*, *81*(2), 138–149.

Milner, H. V. (2021). Voting for populism in Europe: Globalization, technological change, and the extreme right. *Comparative Political Studies*, *54*(13), 2286–2320. https://doi.org/10.1177/0010414021997175

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, *62*(3), 760–775.

Mutz, D. C. (2018). Status threat, not economic hardship, explains the 2016 presidential vote. *Proceedings of the National Academy of Sciences*, *115*(19), E4330–E4339. https://doi.org/10.1073/pnas.1718155115

Norris, P., & Inglehart, R. (2019). Cultural backlash: Trump, Brexit, and authoritarian populism. *Cambridge University Press*. https://doi.org/10.1017/9781108595841

Reiss, P. C., & Wolak, F. A. (2007). Structural econometric modeling: Rationales and examples from industrial organization. *Handbook of econometrics*, *6*, 4277–4415.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series A (General)*, *147*(5), 656–666. https://doi.org/https://doi.org/10.2307/2981697

Samii, C. (2016). Causal empiricism in quantitative research. *The Journal of Politics*, *78*(3), 941–955. https://doi.org/10.1086/686690

Smith, E. R. (1982). Beliefs, attributions, and evaluations: Nonhierarchical models of mediation in social cognition. *Journal of Personality and Social Psychology*, *43*(2), 248.

Tabellini, M. (2019). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *The Review of Economic Studies*, *87*(1), 454–486. https://doi.org/10.1093/restud/rdz027

University of Essex, Institute for Social and Economic Research. (2023). Understanding society: Waves 1-13, 2009-2022 and harmonised bhps: Waves 1-18, 1991-2009. [data collection] (18th) [SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-19].

Walter, S. (2021). The backlash against globalization. *Annual Review of Political Science*, *24*, 421–442.

Wu, A. D., & Zumbo, B. D. (2008). Understanding and using mediators and moderators. *Social Indicators Research*, *87*, 367–392.

In Search of the Causes of the Globalization Backlash: Methodological Considerations on Post-Treatment Bias

Online Appendix

Paolo Agnolin, Italo Colantone, and Piero Stanig

Replication materials for all the analyses are available at Agnolin et al. (2024)

Contents

A Replication of Chan et al. (2020)

Given that their analysis is based mostly on a publicly available survey, Chan et al. (2020) do not distribute replication files. They provide sufficient details to replicate the data preparation, with some exceptions that we detail in what follows. Our aim here in any case is not to re-evaluate the results of that paper, but to show how the inclusion of cultural variables might affect otherwise robust results regarding the role of economic distress in voting behavior. Here we detail how we constructed the variables that ultimately enter the analysis.

The indicator for Leave support—from wave 8 of the UKHLS panel survey (University of Essex, Institute for Social and Economic Research, 2023)—takes the value 1 if the respondent supports Leave, 0 if the respondent supports Remain, and missing otherwise. This is the same outcome variable used in the original paper.

The UKHLS does not have a simple vote eligibility variable, nor a citizenship indicator. To sort out citizenship of the respondent, we rely on information from all the waves (as this information is not reported for all respondents in every wave) up to the one from which we obtain the outcome variable. First, we identify all respondents that are not born in the UK, and then all those who are naturalized. We define as "not citizens" those that were born abroad and not naturalized. The age at the time of the referendum is calculated based on the information about the birth year. Given that the specific day and month of birth is not reported, we exclude all respondents who are not 19 or older in 2016, as some might have turned 18 only after the referendum.

The dummy for poverty is based on the wave 8 household-level dataset of the UKHLS. We calculate the equivalized income by dividing the variable `h_fihhmngrs_dv` by the square root of the sum of the number of children and the number of adults in the household. We then calculate the median household equivalized income, and generate a dummy equal to one for all households with an equivalized income lower than 60% of the the median (which is 1,891 British pounds). This procedure follows closely the approach described in the original paper.

For the strength of British identity variable, we rely on information from waves 1, 3 and 6, and assign to each respondent the average of the three responses if more than one is available. To construct the indicators for English and other national identities, we rely on information scattered in all waves. The national identity is asked with multiple items (natid_1 about English identity, natid_2 for Welsh) that can take values “mentioned”, “not mentioned”, or missing. These are asked to different subsets of respondents in different waves. We rely on the most recent non-missing response, as described in the original paper.

After having obtained binary information about the responses to all the national identity items, we create the dummies based on a five-fold typology: (1) British only; (2) English only; (3) Welsh, Scottish, or (Northern) Irish only; (4) British and English; and (5) all other combinations. The residual category includes all respondents that do not fall into categories 1-4 (hence, for instance, British and Welsh, British and Scottish, or neither British nor English nor Welsh/Scottish/Northern Irish, etc.)

We recode the marital status so that: “living as couple” and “married” count as “couple”; “widowed/surviving civil partner”, “divorced/dissolved civil partner” and “separated (incl. from civil partner)” are counted as “divorced/widowed"; and “never married” as “single”. We topcode the number of children to 3, and create a three-category variable for no children, 1 or 2 children, and 3 or more.

For the race of the respondent, we try to approximate the definition as described summarily in the original paper. Respondents with missing value (a total of 664) are coded as missing.[12]

We recode the categories “british/english/scottish/welsh/northern irish (white)”, “gypsy or irish traveller (white)” and “any other white background” into the White category; “indian (asian or asian british)”, “pakistani (asian or asian british)”, “bangladeshi (asian or asian british)”, “chinese (asian or asian british)” and “any other asian background” into the Asian category; “caribbean (black or black british)”, “african (black or black british)”, and “any other

[12]In the text here we reproduce the categories as listed in the value labels in the public Stata files of the UKHLS; this includes the use of lowercase for ethno-geographic categories.

black background” into the Black category; and finally “white and black caribbean (mixed)”, “white and black african (mixed)”, “white and asian (mixed)”, “any other mixed background (mixed)”, “arab (other ethnic group)” and “any other ethnic group” into the Other category. Notice that the marginals of our ethnic categorization differ from those reported in Chan et al. (2020). Specifically, of the 14,670 respondents included in the descriptive statistics in the online appendix to Chan et al. (2020), 92.6% are White, 3.7% Asian, 1.6% Black, with 2.1% in the Other category. Conversely, our estimation sample—that includes 18,909 respondents—is 85% White, 9.2% Asian, 3.5% Black, and 2.5% Other. Notice that these are the unweighted frequencies.

For education, we create a full set of dummies for all the possible values taken by the variable `h _qfhigh _dv`. We do not try to reclassify education in a sparser number of categories, but include fixed effects for each of the values that the education variable might take, including the residual “None of the above” category.

The social status variable used in Chan et al. (2020) is reportedly based on the scheme developed in Chan and Goldthorpe (2004). We were unable to find a crosswalk from occupational information in UKHLS to social status as defined there, hence we do not include this variable in the analysis. In Chan et al. (2020) social class is measured with a six-fold version of National Statistics Socio-Economic Classification (NS-SEC). We reproduce this by coarsening the variable `h_jbnssec8_dv` (from wave 8 of the UKHLS), which has eight categories in the public use data: “Large employers & higher management”, “Higher professional”, “Lower management & professional”, “Intermediate”, “Small employers & own account”, “Lower supervisory & technical”, “Semi-routine”, and “Routine”. Two of six categories in Chan et al. (2020) are “higher management and professional” and “semi-routine/routine”, which are plausibly derived by collapsing, respectively, “Large employers & higher management” with “Higher professional”, and “Semi-routine” with “Routine”. We proceed in this way to arrive at a six-fold social class categorization. Notice that more than 40% of respondents of wave 8 have the value “Inapplicable” in the NS-SEC variable: these are plausibly people not in the labor force.

Chan et al. (2020) seem to treat these as missing in the main analysis; in fact, there is no estimate for “Inapplicable” in the tables, nor it is the reference category. This said, we do not see a justification for this, also given the aim of having results comparable to Colantone and Stanig (2018a), where respondents not in the labor force are included in the analysis. Hence we create a class variable that has a separate category for “Inapplicable”. Treating people not in the labor force as having a missing social class value leads to their exclusion from the regressions; conversely, treating the “inapplicable” as a specific category means that they enter the analysis with their own intercept.

A.1 Latent class analysis

To create cultural consumption classes in the spirit of those in Chan et al. (2020), we rely on information contained in waves 2 and 5 of the UKHLS. Note that the original paper mentions waves 3 and 5; yet, in the public use UKHLS files there are no cultural consumption items in wave 3, hence we suspect this is a typo. We clean the missing values, combine the answers from the two waves into one, turn the variables into dummies, and export the data.

In the R environment, we estimate the latent class model—using the package poLCA (Linzer & Lewis, 2011)—on the full set of respondents with complete observations on the cultural consumption items. Chan et al. (2020) seem to hint to a restriction to people in the working age population (20 to 64), but as this might make a small difference, and it is not formally explained in the appendix to the original paper, we prefer to use all the available data.

The model has three classes, and we request three replications using different starting values for the estimation algorithm. This automates the search for the global maximum of the log-likelihood function: poLCA returns the parameter estimates corresponding to the model with the greatest log-likelihood (Linzer & Lewis, 2011).

The marginals of the three classes are in line with those reported in Chan et al. (2020). Strictly speaking, the results of this type of latent class models are not fully a deterministic function of the data (even if in principle the multiple-estimates approach with different starting values that we adopt should get the global maximum). In addition, Chan et al. (2020) perform some correction which includes six residual local dependence terms, that we are unable to implement given that a full explanation is not provided. Hence this analysis is in the spirit of the original paper, but does not purport to constitute an exact replication. In any case, conceptually our exercise yields a classification of respondents into three latent classes based on the same exact consumption variables, hence it is equivalent from the substantive point of view.

Importantly, the three classes we obtain are qualitatively similar to those in Chan et al. (2020) in terms of consumption patterns. Table A.2 shows the conditional probabilities of engaging in each of the cultural activities—the same as used in the original paper—conditional on class membership. Class 1 has overall higher probabilities of engaging in all types of consumption, and class 3 has very low probabilities of engaging in any of them. The intermediate class 2 has relatively high probabilities of visiting a museum and visiting a visual arts exhibition. The median number of consumption types for class 1 is five; it is four for class 2, and it is zero for class 3. Following the terminology in the original paper, we label the class with lowest average consumption "univore", the one with highest average consumption "omnivore", and the intermediate one "paucivore".

Table A.1: Summary of the LCA estimation

Class	1	2	3
Estimated class population shares	0.0729	0.273	0.6541
Predicted class memberships	0.0542	0.278	0.6678

Number of observations	:	56515
Number of estimated parameters	:	26
Residual degrees of freedom	:	229
Maximum log-likelihood	:	-155634.1
AIC(3)	:	311320.3
BIC(3)	:	311552.8
G^2(3)	:	2784.414 (LR/deviance)
X^2(3)	:	3971.627 (Chi-square)

Table A.2: Conditional item response probabilities by outcome variable for each class

Class	**arts2b10**	**arts2b11**	**arts2b12**	**arts2a2**	**arts2a3**	**arts2a5**	**arts2a6**	**mla3**
1	0.1976	0.3615	0.6678	0.9835	0.4797	0.8119	0.3988	0.9547
2	0.0680	0.1388	0.3538	0.6221	0.0775	0.2235	0.1698	0.7768
3	0.0058	0.0170	0.1307	0.0173	0.0111	0.0173	0.0688	0.1230

A.2 Models

Columns 1-2 of Table A.3 report the full estimates of the "short" and "long" vote regressions, respectively. These regressions correspond to columns 3-4 of Table 1. All the estimations use the cross-sectional weights `h_indinui_xw`.

The coefficient on the China shock in column 2 (0.380, with standard error 0.210) is de facto indistinguishable from the one Chan et al. (2020) estimate in a very similar specification (model 9 of their Table 2) that includes both the national identities and the cultural consumption variables: 0.388, with standard error 0.208.

The coefficients on the paucivore and omnivore dummies are similar to those reported in the original paper. We estimate -1.074 (s.e. 0.081) for omnivores (comparable to -0.88, with s.e. 0.07, in the original paper) and -0.398 (s.e. 0.040) for paucivores (comparable to -0.312, s.e. 0.051, in the original paper). This similarity is reassuring regarding the fact that our independent replication of the latent class analysis is ultimately tapping the same conceptual space as the one used in the original paper.

The results for the conditioning variables are overall very similar to those in the original paper. Women, minorities, and younger voters tend to be less supportive of Leave, while individuals with more children are more supportive; the dummy for poverty and those for marital status do not enter the regression with a significant coefficient. In the model in which they are included, "English only" and "British and English" identities predict Leave support, with the coefficient on the former around twice as large as the one on the latter: these are respectively 0.442 in our estimate (vs. 0.463 in the original paper) and 0.233 in our estimate (vs. 0.213 in the original).

Table A.3: Results

Dep. var.:	(1) Leave	(2) Leave	(3) Omnivore	(4) Paucivore	(5) Omnivore	(6) Paucivore
China shock	0.445*	0.380	-1.071**	-0.412*	-0.866*	-0.252
	[0.210]	[0.210]	[0.403]	[0.196]	[0.376]	[0.181]
Omnivore		-1.074**				
		[0.081]				
Paucivore		-0.398**				
		[0.040]				
British & English		0.233**				
		[0.063]				
English		0.442**				
		[0.055]				
Other		-0.137				
		[0.084]				
Scot/Welsh/ (N)Irish		0.003				
		[0.103]				
Strength Brit ID		0.108**				
		[0.009]				
Inc. <60% median	0.033	0.013	-0.288*	-0.319**	-0.158	-0.303**
	[0.058]	[0.059]	[0.143]	[0.067]	[0.135]	[0.062]
Higher prof./manager	-0.617**	-0.633**	0.165	0.320**	-0.015	0.280**
	[0.082]	[0.081]	[0.172]	[0.084]	[0.167]	[0.081]
Low salariat	-0.405**	-0.411**	0.253	0.208**	0.145	0.152*
	[0.063]	[0.064]	[0.134]	[0.062]	[0.128]	[0.059]
Intermediate	-0.134	-0.159*	-0.102	0.053	-0.124	0.064
	[0.077]	[0.075]	[0.174]	[0.091]	[0.165]	[0.086]
Self-employed	0.032	0.069	0.325	0.191*	0.228	0.127
	[0.069]	[0.073]	[0.168]	[0.095]	[0.153]	[0.086]
Manual superv.	0.125	0.010	-0.488	-0.192	-0.425	-0.166
	[0.105]	[0.107]	[0.275]	[0.126]	[0.267]	[0.121]
Routine	-0.016	-0.042	-0.320*	-0.201**	-0.255	-0.190**
	[0.067]	[0.068]	[0.159]	[0.067]	[0.153]	[0.065]
Asian	-0.415**	-0.330**	-1.598**	-0.877**	-1.178**	-0.632**
	[0.111]	[0.113]	[0.204]	[0.097]	[0.189]	[0.088]
Black	-0.703**	-0.571**	-1.515**	-0.793**	-1.127**	-0.542**
	[0.133]	[0.131]	[0.229]	[0.139]	[0.213]	[0.131]
Other race	-0.788**	-0.597**	-0.075	-0.051	-0.043	-0.028
	[0.169]	[0.171]	[0.242]	[0.151]	[0.240]	[0.150]
1-2 children	0.132*	0.100	-0.787**	-0.259**	-0.646**	-0.128*
	[0.058]	[0.060]	[0.104]	[0.058]	[0.097]	[0.054]
3+ children	0.383**	0.325**	-1.521**	-0.609**	-1.239**	-0.435**
	[0.118]	[0.117]	[0.333]	[0.124]	[0.329]	[0.122]
Married/cohab	0.024	-0.015	-0.071	0.074	-0.106	0.092
	[0.068]	[0.067]	[0.106]	[0.069]	[0.108]	[0.070]
Divorced/widowed	0.066	0.030	-0.037	-0.173*	0.059	-0.156*
	[0.086]	[0.085]	[0.134]	[0.079]	[0.133]	[0.078]
Female	-0.233**	-0.227**	0.051	0.187**	-0.049	0.173**
	[0.034]	[0.034]	[0.071]	[0.046]	[0.067]	[0.044]
Age	0.046**	0.057**	0.104**	0.036**	0.086**	0.021**
	[0.007]	[0.007]	[0.014]	[0.007]	[0.014]	[0.007]
$(Age)^2$	-0.000**	-0.000**	-0.001**	-0.000**	-0.001**	-0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
NUTS-1 FE	Y	Y	Y	Y	Y	Y
Education FE	Y	Y	Y	Y	Y	Y
Constant	-2.318**	-3.041**	-3.034**	-0.762**	-3.400**	-0.886**
	[0.200]	[0.212]	[0.452]	[0.274]	[0.401]	[0.251]
Observations	18,909	18,909	18,909	18,909	18,909	18,909

Note: Standard errors clustered by NUTS-3 region in brackets.
** $p<0.01$, * $p<0.05$

The coefficient on British identity we recover (0.108, with standard error 0.009) is de facto identical to the one estimated in the original paper (0.104, with standard error 0.008).

Columns 3-4 of Table A.3 report the estimates of a multinomial logit model where the categorical variable for the latent class membership is regressed on the China shock and all the controls included in the specification of column 1. These estimates correspond to those in columns 2-3 of Table 2. The China shock has a negative and statistically significant effect on membership in the two classes denoting higher cultural consumption, after controlling for demographics. The results for the control variables are highly intuitive: poor people, those with more children, ethnic minorities, and those in routine occupations are less likely to belong to the higher-consumption classes. In columns 5-6 we report the estimates of separate binary logit models where the outcome is an indicator for membership in, respectively, the omnivore and the paucivore classes, which takes the value of zero if a respondent is not a member of that latent class. The results are substantively analogous according to this simpler (but less adequate) approach.

B Replication of Mutz (2018)

For the empirical exercise based on Mutz (2018), we rely on the published replication package. We focus on the cross-sectional analysis, that is based on a survey from October 2016. The survey includes a representative national probability sample collected by the National Opinion Research Center (NORC) at the University of Chicago.

Column 1 of Table B.1 reports estimates from the "short" vote regression, corresponding to column 5 of Table 1. The dependent variable is the Trump thermometer advantage rating, measured on a 20-point scale as in the original paper.[13] Subjective perception of family finances is used to assess pocketbook economic evaluation. This is measured on a five-point scale index, with higher values indicating more positive perceptions. We include controls for partisanship (Democrat), education (dummy variable indicating college graduation), ethnicity (dummy variable indicating white respondents), gender (female), religiosity, age (categorical variable with seven categories), family income, unemployment status, median income in the place of residence, and sociotropic economic perceptions (1-5 scale index measuring perceptions of the national economy, with higher values indicating more positive perceptions). Unlike Mutz (2018), who treated the 7-category age variable as continuous, we utilize a set of seven dummy variables for the age categories. Also, we exclude the measures of local economic context that were deleted from the replication files for data-privacy reasons (i.e., share of unemployed and share of manufacturing workers).

What we estimate in column 1 of Table B.1 is a shorter version of the original paper's extensive specification (reported in Appendix Table S4 of Mutz, 2018). In fact, our aim is not to re-evaluate the results in the original paper, but to show how the exclusion vs. inclusion of cultural variables might lead to different conclusions about the role of economic distress in

[13] Respondents are asked to rate each presidential candidate (Donald Trump/Hillary Clinton) on a thermometer that runs from 0° to 100°. Rating above 50° indicates favorable attitudes toward the candidate, while rating below 50° represents unfavorable attitudes. In order to compute Trump's thermometer advantage rating, Clinton candidate ratings were subtracted from Trump thermometer ratings. The difference thus ranges from – 100 to 100 and is later collapsed into 20 evenly spaced categories.

voting behavior. To this purpose, in column 2, we augment the specification with a variable measuring the stance on immigration policy. This is measured, like in the original paper, as the average of three items on a 5-point scale. In particular, respondents were asked about their approval or opposition to the following proposals addressing immigration: (1) provide a path to citizenship for some illegal aliens who agree to return to their home country for a period of time and pay substantial fines; (2) increase border security by building a fence along part of the US border with Mexico; (3) return illegal immigrants to their native countries. The three items are on 5-point scales. Higher values denote more pro-immigration attitudes. The specification of column 2 in Table B.1 corresponds to the one in column 6 of Table 1.

Finally, in column 3 of Table B.1 we regress the stance on immigration policy on pocketbook economic perceptions and all the other controls employed in column 1. This specification corresponds to column 4 of Table 2.

Table B.1: Replication of Mutz (2018)

	(1)	(2)	(3)
Dep. var.:	Trump th.	Trump th.	Immigration
Family finances (perception - better)	-0.204**	-0.114	0.069**
	[0.078]	[0.075]	[0.020]
Support for immigration	-	-1.306**	-
		[0.079]	
Party identification (Democratic)	-3.324**	-2.673**	0.499**
	[0.094]	[0.099]	[0.023]
Education (not college graduate)	0.816**	0.470**	-0.265**
	[0.152]	[0.141]	[0.039]
Race (white)	1.077**	1.064**	-0.010
	[0.171]	[0.162]	[0.042]
Gender (female)	-0.929**	-0.741**	0.144**
	[0.140]	[0.134]	[0.035]
Religiosity	0.050	0.047	-0.003
	[0.027]	[0.025]	[0.007]
Income	0.043*	0.049**	0.005
	[0.020]	[0.018]	[0.005]
Looking for work	-0.034	-0.090	-0.043
	[0.315]	[0.299]	[0.071]
Median income	-0.000	-0.000	0.000
	[0.000]	[0.000]	[0.000]
National economy (better)	-1.493**	-1.146**	0.266**
	[0.081]	[0.080]	[0.020]
Age dummies	Y	Y	Y
Observations	2,888	2,888	2,888
R-squared	0.598	0.642	0.392

Note: ** p<0.01; * p<0.05

C Sequential ignorability in our example

In this section, we discuss our contrived example in the framework of "sequential ignorability", which is often adopted for causal mediation analysis in political science (Imai et al., 2011). As for the case of principal ignorability, the assumption underlying sequential ignorability is composed of two parts. The first requires that the treatment variable (in our example, the economic shock) is randomly assigned, at least conditional on a set of pre-treatment covariates. This assumption is not particularly problematic, as it is required in any case to make claims about the causal effect of the treatment on the outcome. As a matter of fact, we have made this assumption throughout the example.

The second part of the assumption states that, conditional on the treatment, the observed value of the mediator is independent of the potential outcomes. In our example, the potential outcomes are the two propensities to support the radical-right party under, respectively, "no shock" or "shock". Sequential ignorability then requires that, among those who did not receive the shock, displaying libertarian or authoritarian cultural traits (i.e., the observed value of the mediator) is unrelated to the pair of propensities to vote for the radical right. The same applies symmetrically to shocked individuals. As can be seen in Table 3, this requirement is clearly not satisfied in our set-up. In fact, both under "no shock" and "shock", the value of the mediator—i.e., being observed as authoritarian or libertarian (columns 2-3)—is related to the individual's stratum (column 1), which determines both the baseline predisposition to support the radical right under no shock and the probability to support it having received the shock (columns 4-5). Sequential ignorability would instead require that being observed as displaying libertarian or authoritarian attitudes, conditional on having received or not the shock, is independent of one's propensities to support the radical right.

In a more formal way, one can also see how the violation emerges in the setting of the structural equations by Imai et al. (2011), p. 787. In our case, the causal effect of the treatment on the mediator is heterogeneous: zero for genuine libertarians and genuine authoritarians, and positive for the impressionable stratum.

Formally,

$$M_i(T_i) = \alpha_{2i} + \beta_{2i} T_i + \epsilon_{2i} \tag{2}$$

and

$$Y_i(T_i, M_i) = \alpha_{3i} + \beta_3 T_i + \gamma M_i + \epsilon_{3i} \tag{3}$$

As in Imai et al. (2011), we can rewrite the (observation-specific) effect of the treatment on the mediator in equation 2 as $\beta_{2i} = \beta_2 + \eta_i$, and analogously the intercept in equation 3 as $\alpha_{3i} = \alpha_3 + \xi_i$.

In our example, the effect of the treatment on the mediator is heterogeneous, and its variation is correlated with the group-specific intercept in the equation for the outcome. In particular, observations with the highest baseline propensity to support the radical right (i.e., highest α_{3i}) are the genuine authoritarians, for which the effect of the treatment on the mediator is zero. Regressions that do not allow for heterogeneity have the form:

$$M_i(T_i) = \alpha_2 + \beta_2 T_i + \epsilon^*_{2i} \tag{4}$$

and

$$Y_i(T_i, M_i) = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon^*_{3i} \tag{5}$$

where the new error terms are, respectively, $\epsilon^*_{2i} = \eta_i T_i + \epsilon_{2i}$ and $\epsilon^*_{3i} = \xi_i + \epsilon_{3i}$. If ξ_i and η_i covary, the error terms are correlated across equations. Hence sequential ignorability does not hold.

D Observational study

We focus on fifteen western European countries, over the period 1995-2008. The sample includes: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. As an economic shock we consider exposure to import competition from China at the regional level, and we relate it to various individual-level measures of cultural attitudes, based on survey data.

D.1 Individual-level data

Individual-level data are sourced from the European Social Survey (ESS) and the European Value Study (EVS). For ESS we use the first four waves (ESS, 2002, 2004, 2006, 2008), spanning the period 2002-2008. As for EVS (EVS, 2020), we employ waves 3-4, covering the period 1995-2008. Throughout the analysis, we run separate regressions for the ESS and the EVS sample. In fact, these samples contain data on different cultural attitudes.

We focus on four main groups of individual cultural traits: (1) "meta-political" attitudes; (2) private authoritarian attitudes; (3) traditional conservatism; and (4) immigration attitudes. Some of these variables are available in the ESS sample, others in EVS.

The set of meta-political attitudes contains three variables: (1) importance of liberal values (ESS); (2) support for unconstrained strong leaders (EVS); and (3) support for democracy (EVS). The indicator for liberal values is an index corresponding to the first principal component of four individual items on the importance of: equality and equal opportunities; understanding different people; being free; and following rules. We regard these attitudes as underpinning the foundations of liberal democracy. Data on each item are available on a 6-point scale. The items are positively correlated, with the importance of following rules showing the weakest correlation with the others. The variable on strong leaders captures preferences regarding a strong leader free from the control of parliament and elections, while support for democracy captures individual opinions about the desirability of democracy as a form of government. Both variables are on a 4-point scale, with higher values denoting more authoritarian attitudes. We also employ an overall "democratic" index, computed as the sum of these two items, to capture general support for the democratic system.

Our second set of attitudes addresses the concept of authoritarianism as a personality trait, as discussed in political psychology. In particular, we focus on child-rearing as a dimension of private authoritarian stances. Specifically, we use the first principal component of four EVS items on the importance assigned to the following qualities of children: manners, imagination, obedience, and independence. These items are often used as proxies for an authoritarian personality (Feldman & Stenner, 1997). The four original items are coded as binary variables equal to 1 when a given quality is considered to be "especially important". While assigning importance to manners and obedience is considered as an authoritarian stance, assigning it to imagination and independence has the opposite interpretation, and the two pairs of attitudes are indeed negatively correlated. The overall index, named "children qualities", is coded in such a way that higher values denote a more authoritarian attitude.

To measure traditional conservatism we use two ESS items about the importance assigned to following traditions ("tradition") and living in safe surroundings ("safety"). These are measured on a six-point scale, with higher values denoting higher assigned importance. We also consider an EVS item on attitude about abortion, a central marker of traditionalism and conservatism in western countries (e.g., Fiorina and Abrams, 2008; Engeli et al., 2012). Specifically, we employ a measure on a ten-point scale with higher values indicating a less permissive stance, from "always" to "never justifiable".

The last set of cultural variables contains attitudes about immigration. In particular, we consider two ESS items asking about: (1) whether the country's cultural life is undermined or enriched by immigrants ("immigration culture"); and (2) whether immigration is bad or good for the country's economy ("immigration economy"). Both variables are measured on a 10-point scale, with higher values denoting more negative views of immigration. That is, individuals with higher scores tend to believe that their country's cultural life is undermined by immigration, and that immigration has a negative impact on the economy.

More details on all the cultural variables are provided in Table D.3.

D.2 The China shock

The main explanatory variable in our empirical analysis is exposure to import competition from China. The surge of China as a global exporter can be viewed as an exogenous source of structural change, with different regions being more or less vulnerable depending on their ex-ante industry specialization. The literature has documented how exposure to the so-called "China shock" has caused significant and long-lasting economic grievances at the regional level, which have in turn been related to political consequences (for a review, see Colantone et al., 2022).

Following Autor et al. (2013) and Colantone and Stanig (2018b), we measure exposure to Chinese import competition at the region-year level using the following indicator:

$$Import\ Shock_{crt} = \sum_{j} \frac{L_{rj}^{pre-sample}}{L_{r}^{pre-sample}} * \frac{\Delta IMPChina_{cjt}}{L_{cj}^{pre-sample}} \tag{6}$$

where c indexes countries, r regions, j industries and t years. $\Delta\text{IMPChina}_{cjt}$ is the change in (real) imports from China over the past n years, in country c and manufacturing industry j. This is normalized by the pre-sample number of workers in the same country and industry ($L_{cj}^{pre-sample}$). To retrieve the regional shock, we compute a weighted summation of all the industry-level changes in imports. The weights capture the relative importance of each manufacturing industry out of total employment in each region pre-sample ($L_{rj}^{pre-sample}$ / $L_{r}^{pre-sample}$). This index is based on a theoretical model developed by Autor et al. (2013). It is meant to capture the displacement generated by Chinese imports on the supply side of importing countries. Intuitively, the import shock is stronger in years in which the surge in Chinese imports scaled up, and for regions in which relatively more workers were historically employed in industries most affected by the subsequent import growth.

We measure the import shock at the NUTS-2 level of regional disaggregation.[14] Overall, our sample spans 143 regions. We source import data from Eurostat Comext and the CEPII-BACI database. To compute the employment shares at the regional level, we rely on data from Eurostat and a number of national sources. Full details are provided in Table D.4. We work at the NACE Rev 1.1 sub-section level of industry disaggregation, which cuts the manufacturing sector into 14 industries (details in Table D.5).

To deal with potential endogeneity concerns, we follow the same approach as in Autor et al. (2013) and Colantone and Stanig (2018b). Specifically, we employ the following instrumental variable:

$$Instrument\ Import\ Shock_{crt} = \sum_j \frac{L_{rj}^{pre-sample}}{L_r^{pre-sample}} * \frac{\Delta IMPChinaUSA_{jt}}{L_{cj}^{pre-sample}} \quad (7)$$

The difference with respect to Equation (6) is in the numerator of the second term, where we consider imports from China to the US instead of each European country. This instrument is designed to capture the variation in Chinese imports to Europe that is driven by exogenous changes in supply conditions in China, rather than by domestic factors specific to each European country, that could correlate with individual cultural attitudes and electoral outcomes.

D.3 Results

Tables D.1-D.2 report full estimation results based on the specification outlined in Equation 1. The coefficients on the import shock correspond to those reported in Figure 1.

[14]The only exceptions are France, Germany, and the UK, for which either individual data or employment shares data are only available at the NUTS-1 level.

Table D.1: Cultural attitudes on import shock - ESS sample

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	**Liberal values**	**Tradition**	**Safety**	**Immig. (culture)**	**Immig. (econ.)**
Import shock	0.221**	-0.009	-0.103**	0.433**	0.026
	[0.038]	[0.041]	[0.036]	[0.074]	[0.072]
Female	-0.137**	0.099**	0.248**	-0.009	0.329**
	[0.011]	[0.012]	[0.011]	[0.021]	[0.020]
Age	-0.001**	0.020**	0.007**	0.014**	0.004**
	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]
Estimator	2SLS	2SLS	2SLS	2SLS	2SLS
Education dummies	yes	yes	yes	yes	yes
Country-year effects	yes	yes	yes	yes	yes
Obs.	95,806	97,060	97,085	100,243	99,934
R-squared	0.04	0.12	0.09	0.13	0.10
First-stage results					
US imports from China	0.078**	0.077**	0.077**	0.076**	0.076**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Kleibergen-Paap F-Stat.	5,995	6,019	6,022	6,052	6,057

Note: ** p<0.01, * p<0.05.

D.4 Information on variables

Table D.2: Cultural attitudes on import shock - EVS sample

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	**Strong leader**	**Democracy**	**Democratic index**	**Children qualities**	**Abortion**
Import shock	0.283**	0.226**	0.514**	-0.025	0.396**
	[0.039]	[0.027]	[0.052]	[0.035]	[0.118]
Female	-0.019*	0.042**	0.020	-0.029**	0.069*
	[0.009]	[0.006]	[0.012]	[0.006]	[0.027]
Age	0.001**	-0.002**	-0.002**	0.003**	0.028**
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]
Estimator	2SLS	2SLS	2SLS	2SLS	2SLS
Education dummies	yes	yes	yes	yes	yes
Country-year effects	yes	yes	yes	yes	yes
Obs.	46,867	47,104	44,805	38,986	49,423
R-squared	0.10	0.07	0.12	0.16	0.16
First-stage results					
US imports from China	0.093**	0.092**	0.092**	0.070**	0.092**
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Kleibergen-Paap F-Stat.	2,477	2,486	2,398	1,664	2,557

Note: ** p<0.01, * p<0.05

Table D.3: Cultural attitudes

Variable	Survey	Survey item	Range Min.	Range Max.	Mean	Std. Dev.
Liberal values	ESS	First principal component of the four items:			0	1.257
		She/he thinks it is important that every person in the world should be treated equally. She/he believes everyone should have equal opportunities in life.	1 Very much like me	6 Not like me at all		
		It is important to her/him to listen to people who are different from her/him. Even when she/he disagrees with them, she/he still wants to understand them.	1 Very much like me	6 Not like me at all		
		It is important to her/him to make her/his own decisions about what she/he does. She/he likes to be free and not depend on others.	1 Very much like me	6 Not like me at all		
		She/he believes that people should do what they're told. She/he thinks people should follow rules at all times, even when no-one is watching.	1 Very much like me	6 Not like me at all		
Tradition	ESS	Tradition is important to her/him. She/he tries to follow the customs handed down by her/his religion or her/his family.	0 Not al all like me	6 Very much like me	4.178	1.370
Safety	ESS	It is important to her/him to live in secure surroundings. She/he avoids anything that might endanger her/his safety.	0 Not al all like me	6 Very much like me	4.541	1.232
Immigration (culture)	ESS	Would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries?	0 Cultural life enriched	10 Cultural life undermined	4.248	2.465
Immigration (economy)	ESS	Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries?	0 Good for the economy	10 Bad for the economy	4.933	2.355
Strong leader	EVS	Would you say it is a very good, fairly good, fairly bad or very bad way of governing this country? Having a strong leader who does not have to bother with parliament and elections	1 Very bad	4 Very good	1.821	0.927
Democracy	EVS	Would you say it is a very good, fairly good, fairly bad or very bad way of governing this country? Having a democratic political system	1 Very good	4 Very bad	1.550	0.669
Democratic index	EVS	Sum of the variables: strong leader, democratic	2	8	3.350	1.277
Children qualities	EVS	Here is a list of qualities that children can be encouraged to learn at home. Which, if any, do you consider to be especially important? First principal component of the four items about:			0	0.563
		Good manners	0 Not mentioned	1 Important		
		Imagination	0 Important	1 Not mentioned		
		Obedience	0 Not mentioned	1 Important		
		Independence	0 Important	1 Not mentioned		
Abortion	EVS	Please tell me for each of the following statements whether you think it can always be justified, never be justified, or something in between. Abortion	1 Always justifiable	10 Never justifiable	6.161	2.991

Table D.4: Import shock data

	Employment Data		Trade Data	
Country	**Initial Year**	**Source**	**Availability**	**Source**
Austria	1995	Eurostat	1995 - 2007	Eurostat Comext
Belgium	1995	National Bank of Belgium	1988 - 2007	Eurostat Comext
Finland	1995	Statfin	1995 - 2007	Eurostat Comext
France	1989	INSEE	1988 - 2007	Eurostat Comext
Germany	1993	Federal Employment Agency	1988 - 2007	Eurostat Comext
Greece	1988	HSA Statistics Greece	1988 - 2007	Eurostat Comext
Ireland	1995	Eurostat	1988 - 2007	Eurostat Comext
Italy	1988	ISTAT	1988 - 2007	Eurostat Comext
Netherlands	1988	CBS Statistics Netherlands	1988 - 2007	Eurostat Comext
Norway	1994	Statistics Norway	1995 - 2007	CEPII - BACI
Portugal	1990	INE Portugal	1988 - 2007	Eurostat Comext
Spain	1993	INE Spain	1988 - 2007	Eurostat Comext
Sweden	1993	SCB Statistics Sweden	1995 - 2007	Eurostat Comext
Switzerland	1995	SFSO Swiss Statistics	1995 - 2007	CEPII - BACI
United Kingdom	1989	ONS	1988 - 2007	Eurostat Comext

Table D.5: NACE Rev. 1.1 industries

Code	**Industry description**
DA	Manufacture of food products, beverages and tobacco
DB	Manufacture of textiles and textile product
DC	Manufacture of leather and leather products
DD	Manufacture of wood and wood products
DE	Manufacture of pulp, paper and paper products; publishing and printing
DF	Manufacture of coke, refined petroleum products and nuclear fuel
DG	Manufacture of chemicals, chemical products, and man-made fibres
DH	Manufacture of rubber and plastic products
DI	Manufacture of other non-metallic mineral products
DJ	Manufacture of basic metals and fabricated metal products
DK	Manufacture of machinery and equipment n.e.c.
DL	Manufacture of electrical and optical equipment
DM	Manufacture of transport equipment
DN	Manufacturing n.e.c. (furniture, toys etc.)

References

Agnolin, P., Colantone, I., & Stanig, P. (2024). *Replication Data for: In search of the Causes of the Globalization Backlash*. https://doi.org/10.7910/DVN/MKJYV3

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, *103*(6), 2121–68. https://doi.org/10.1257/aer.103.6.2121

Chan, T. W., & Goldthorpe, J. H. (2004). Is there a status order in contemporary British society? Evidence from the occupational structure of friendship. *European Sociological Review*, *20*(5), 383–401.

Chan, T. W., Henderson, M., Sironi, M., & Kawalerowicz, J. (2020). Understanding the social and cultural bases of Brexit. *The British Journal of Sociology*, *71*(5), 830–851.

Colantone, I., Ottaviano, G. I., & Stanig, P. (2022). The backlash of globalization. *In G. Gopinath, E. Helpman, and K. S. Rogoff (Eds), Handbook of International Economics (Vol. V)*, 405–477.

Colantone, I., & Stanig, P. (2018a). Global competition and Brexit. *American Political Science Review*, *112*(2), 201–218.

Colantone, I., & Stanig, P. (2018b). The trade origins of economic nationalism: Import competition and voting behavior in western Europe. *American Journal of Political Science*, *62*(4), 936–953.

Engeli, I., Green-Pedersen, C., & Larsen, L. T. (2012). *Morality politics in Western Europe: Parties, agendas and policy choices*. Springer.

ESS. (2002). Ess round 1: European social survey round 1 data (2002) (6.6) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS1-2002].

ESS. (2004). Ess round 2: European social survey round 2 data (2004) (3.6) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS2-2004].

ESS. (2006). Ess round 3: European social survey round 3 data (2006) (3.7) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS3-2006].

ESS. (2008). Ess round 4: European social survey round 4 data (2008) (4.5) [Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS4-2008].

EVS. (2020). EVS (2020): European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008) [ZA4804 Data file Version 3.1.0, https://doi.org/10.4232/1.13486].

Feldman, S., & Stenner, K. (1997). Perceived threat and authoritarianism. *Political Psychology*, *18*(4), 741–770. https://doi.org/https://doi.org/10.1111/0162-895X.00077

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*(1), 563–588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, *105*(4), 765–789.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, *42*(10), 1–29.

Mutz, D. C. (2018). Status threat, not economic hardship, explains the 2016 presidential vote. *Proceedings of the National Academy of Sciences*, *115*(19), E4330–E4339. https://doi.org/10.1073/pnas.1718155115

University of Essex, Institute for Social and Economic Research. (2023). Understanding society: Waves 1-13, 2009-2022 and harmonised bhps: Waves 1-18, 1991-2009. [data collection] (18th) [SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-19].

Chapter 2

Robots Replacing Trade Unions: Novel Data and Evidence from Western Europe

Robots Replacing Trade Unions: Novel Data and Evidence from Western Europe*

Paolo Agnolin† Massimo Anelli‡ Italo Colantone§ Piero Stanig¶

Abstract

Labor unions play a crucial role in liberal democracies by influencing labor market and political dynamics, organizing workers' demands and linking them to parties. However, their importance has progressively diminished in the last decades. We suggest that technological change—and industrial robotization in particular—has contributed to weakening the role of unions. We produce novel granular data on union density at the sub-national and industry level for 15 countries of western Europe over 2002-2018. Employing these data, we estimate the impact of industrial robot adoption on unionization rates. We find that regions more exposed to automation experience a decrease in union density. The decline in unionization occurs via a compositional effect, i.e., a reallocation of employment away from traditionally unionized industries towards less unionized ones. On the other hand, there is no clear evidence of a systematic reduction in union density within industries more exposed to automation.

*We thank Catherine de Vries, Pablo Beramendi, Tito Boeri, Silja Häusermann, Herbert Kitschelt, Alessia Matano, Paolo Naticchioni, and seminar participants at the 2024 APSA Comparative Labor Workshop, Bank of Italy, Bocconi University, Duke University, California Institute of Technology, EPSA 2023, European Society for Population Economics Conference 2024, International Political Economy Society Conference 2024, Joint Research Centre, MPSA 2022, Scuola Normale Superiore, SPSA 2024, Turin University, University of North Carolina at Chapel Hill, Verona University, Yale University and Zurich University for insightful comments. We gratefully acknowledge funding from the Fondazione CARIPLO Science and Technology Grant. Andrea Cancellieri, Giovanni Della Seta, Simone Foresti, Juan David Garcia Gonzalez, Maria Elena Lasiu, Mariadolores Schiavone, Fabrizio Viti and Jakob Wall provided excellent research assistance. The usual disclaimer applies.

†Bocconi University, Department of Social and Political Sciences, Dondena Centre for Research on Social Dynamics. Contact: paolo.agnolin@unibocconi.it.

‡Bocconi University, Department of Social and Political Sciences, CESifo, IZA, Dondena Centre for Research on Social Dynamics and Baffi Research Centre. Contact: massimo.anelli@unibocconi.it.

§Bocconi University, Department of Social and Political Sciences, GREEN Research Centre, Baffi Research Centre, CESifo and FEEM. Contact: italo.colantone@unibocconi.it.

¶Bocconi University, Department of Social and Political Sciences, GREEN Research Centre and Dondena Centre for Research on Social Dynamics, Contact: piero.stanig@unibocconi.it.

1 Introduction

Throughout the 20th century, trade unions have played a central role in liberal democracies, influencing labor market dynamics and channeling workers' political demands into an organized voice (Ahlquist, 2017). By negotiating wages and working conditions, unions influenced how the benefits of economic progress were distributed, contributing to limit the rise in inequalities (e.g., Farber et al., 2021). By mobilizing workers in support of redistributive policies, unions served as a link between workers' constituencies and pro-redistribution parties, especially of the mainstream left (e.g., Häusermann and Kitschelt, 2023; Przeworski and Sprague, 1986). Yet, in recent decades, unions have experienced a decline in membership, paired with diminished relevance in the democratic process (Rosenfeld, 2014). This phenomenon is consequential for electoral dynamics (e.g. Kitschelt, 2012; Rennwald and Pontusson, 2021).

In this article, we provide the first evidence on unionization dynamics at the granular region and industry level covering 15 countries of western Europe, over 2002-2018. Using these novel data, we show that technological change is a key determinant of the decline in unionization. Specifically, we focus on the surge of industrial robot adoption in manufacturing, which has been identified as a main dimension of structural change over the period of analysis (e.g., Boix, 2019; Gallego and Kurer, 2022).

Automation through robots has been found to produce aggregate welfare gains with significant distributional effects. Workers that are more vulnerable to substitution by robots, and regions where these workers are more concentrated, are relative losers of this phenomenon (e.g., Acemoglu and Restrepo, 2020; Dauth et al., 2021). The unequal economic consequences of automation, in turn, have been shown to be politically consequential. In particular, automation losers have been found to turn towards radical parties, especially of the right (e.g., Anelli et al., 2021; Milner, 2021). The impact of automation on unionization has instead remained largely underexplored, with the notable exception of Balcazar (2024).

Yet trade unions are potentially key actors shaping the implications of technological change, both in terms of distributional effects on the labor market, and in terms of the ensuing political consequences. By investigating the link between automation and unionization, this article aims at furthering the general understanding of the economic and political implications of structural change.

Thus far, comparative research on trade unions has been hindered by data limitations. In fact, the only available source of comparable data across countries, the database by Visser (2019), only provides information on the overall unionization rate at the country-year level. For most European countries, in fact, these are also the only available data on unionization. Aggregate figures obscure critical sub-national and sectoral variation, and the lack of granular data makes it essentially impossible to study the determinants of unionization in a causally identified way (Ahlquist, 2017; Lipset, 1983).

In light of this, the first contribution of this paper is to assemble a novel database on region- and industry-level unionization rates across 15 western European countries.[1] To this purpose, we combine data from the European Social Survey (ESS) with country censuses over two decades (2002-2018), and use dynamic multilevel regressions with post-stratification (MrP, Park et al., 2004; Gelman et al., 2019). Intuitively, we first estimate the predicted probability of unionization for different types of individuals, defined as combinations of age, gender, education, occupation, industry, and region of residence, based on ESS data. Then, we compute the union density at, for instance, the regional level, by taking a weighted average of the unionization rates for different types. The weights represent the relative share of each type of individuals out of the population of the region, based on census data. This is an innovative application of the methodology initially proposed by Park et al. (2004), which has mostly been used thus far for predicting public opinion at the sub-national level in single-country contexts.

[1]The complete database containing granular unionization data will be made publicly available for the research community upon publication of this paper, at the following dedicated website: www.uniondata.info.

This approach allows us to obtain union density estimates that are representative at the sub-national and industry level, and comparable across countries. We validate these unionization estimates with administrative data from the handful of countries where such granular data are available. Moreover, when aggregating our estimates at the country-year level for all sample countries, we retrieve patterns that are highly consistent with those obtained from the Visser (2019) database.

We show that automation is a significant determinant of unionization decline. In particular, regions that are more exposed to robot adoption witness a decrease in union density. The effect of automation is not only statistically but also substantively significant: a one standard deviation increase in robot exposure leads to a decrease in union density by around 34.4% of a standard deviation, accounting for country-year and region fixed effects. The decline in unionization occurs via a compositional effect, i.e., a reallocation of employment away from traditionally unionized industries—that are on average more exposed to automation—towards less unionized ones. Conversely, there is no clear evidence of a systematic reduction in union density within industries where robot adoption is higher. Overall, our results speak to the importance of unionization as a contextual factor that may influence electoral dynamics at the local level, shaping the political repercussions of automation.

2 Background and conceptual framework

In this section we present the conceptual framework of the analysis. First, we discuss the political significance of trade unions and their decline over recent decades, which has been shown to be politically consequential. Then, we propose technological change, and chiefly robotization, as a driver of this decline. Finally, we connect our work to existing literature that has investigated other factors of unionization decline, clarifying and highlighting our novel contribution.

2.1 The political significance of trade unions

From an economic perspective, trade unions have historically played an important role in terms of wage setting and, more generally, in terms of bargaining over working conditions (e.g., Freeman and Medoff, 1984). Unions influence distributional outcomes by shaping how the welfare gains of economic progress are distributed. Unionization is generally associated with reduced wage inequality among unionized workers (Rosenfeld, 2014). Moreover, unions' influence on labor compensation extends beyond their immediate membership (Western and Rosenfeld, 2011). In fact, public policies often expand the reach of union-negotiated wage agreements to non-unionized workers ("union coverage effects"), and employers may independently increase wages to deter further unionization ("union threat effect"). Overall, through a combination of direct and indirect channels, rising unionization tends to reduce inequality (Farber et al., 2021), and declining unionization tends to increase it (Western and Rosenfeld, 2011).

Beyond their economic significance, labor unions also play a central role from a political perspective, in a number of ways. At a first level of analysis, unions enhance political participation by providing information and fostering engagement. Union members are more politically knowledgeable and engage more in political discussion (Iversen and Soskice, 2015). In line with that, unions have a positive influence on voter turnout (Becher and Stegmueller, 2019; Leighley and Nagler, 2007; Radcliff and Davis, 2000), an empirical regularity known as the "union vote premium". Unions also enable working-class citizens to exercise political power (Lipset, 1983; Marks, 1982). In fact, they provide workers with political training and support pathways to careers in politics. Union members and leaders frequently transition into elected or appointed government roles, and serving as a union officer is often a stepping stone to public office. This is an important factor fostering the presence of workers in politics (Mach et al., 2024).

Besides the impact on political participation, and related to that, unions organize workers' demands and frame the political discourse around distributional issues. Trade unions shape workers' policy views, making them think in terms of class interest and class conflict rather than cultural or ethnic conflict (Frymer and Grumbach, 2021; Kim and Margalit, 2017). At the same time, they foster the relevance of workers' interests in the electoral arena. On the one hand, they do so by inducing political parties to focus on distributional economic issues that are relevant to workers (Pontusson and Rueda, 2010). On the other hand, they mobilize workers in support of parties that adopt pro-redistribution stances, especially within the mainstream left (Chang, 2001; Dark, 2018; Schlozman, 2015). Labor unions also significantly influence the behavior of elected politicians, raising political responsiveness and legislative support for redistributive policies. In particular, strong union presence increases responsiveness to lower-income constituents, narrowing disparities in representation across income groups (Becher and Stegmueller, 2020). Overall, from the perspective of power resource theory (Esping-Andersen, 1985; Korpi, 1983, 1986), unions are key political actors driving redistribution and social policy.

Importantly, much of the evidence suggests that labor unions have a *contextual* impact that extends beyond their direct membership. The presence and strength of unions can generate spillovers that affect non-members. This is true both in industrial relations—where collective agreements often cover non-unionized workers—and in politics, where unions act as intermediaries for social constituencies that are broader than their membership (Freeman and Medoff, 1984). For instance, Ahlquist and Levi (2013) highlight how union leaders inspire collective actions that transcend narrow self-interest, underscoring the outward-facing nature of unions' political engagement. Union presence in workplaces and their integration into communities can shape political behavior even among non-members, and play a critical role in local political campaigns (Lyon and Schaffner, 2021; Lopez, 2004). Such contextual effects may be as impactful as the direct effects of union membership.

Overall, these considerations call for studying unions as contextually consequential actors. Empirically, this requires a shift away from focusing solely on individual union membership in surveys, moving to considering regional- and industry-level unionization measures. This type of research has thus far been hindered by data limitations. Yet, some indirect evidence on the contextual role of unions exists. For instance, the size of the workplace in which people are employed is one of the most commonly cited factors affecting union membership (e.g., Oesch, 2006). That is, unionization rates tend to be higher in larger plants, which provide a more fertile ground for unions' activities. In parallel, relating to the political repercussions of unionization, Arndt and Rennwald (2017) document how workers in larger plants are less likely to support the radical right, suggesting that this is related to higher unionization. Along the same lines, at the country level, the employment share of large plants is positively correlated with social democratic parties' vote shares (Pontusson, 1995). One of the main contributions of this study is to provide novel granular data enabling a more direct analysis of the contextual role of unions at both the sub-national and industry level.

Our novel data also allow to uncover the dynamics underlying the phenomenon of unionization decline observed in recent decades across most Western countries. In the United States, for instance, private sector unionization dropped dramatically between 1973 and 2007, falling from 34% to 8% among men, and from 16% to 6% among women (Western and Rosenfeld, 2011). Similar declines have been observed in Europe, even in Scandinavian countries that were traditionally characterized by relatively high unionization rates. Comparative analyses of unionization trends in OECD countries, based on country-level data, indicate that the 1980s marked a critical turning point (e.g., Western, 1997; Wallerstein and Western, 2000). Unionization started a decline that persisted in most cases into the 21st century. As a result, union density is now at historically low levels (Rosenfeld, 2014).

Not surprisingly, given the political significance of unions, the decline of unionization has been shown to be politically consequential. In the US, Feigenbaum et al. (2018) show that the weakening of unions reduces Democratic vote share and turnout. Moreover, it also affects who runs for office: as declining unionization forces unions to prioritize membership recruitment and retention, they allocate fewer resources to campaign contributions and political activity. Along similar lines, Carnes (2013) finds that diminishing unionization has contributed to reducing the number of legislators with working-class backgrounds, weakening the political representation of these groups. Weaker unions have also been found to be associated with increased legislative support for trade deregulation and reduced support for workers' compensatory measures (Becher and Stegmueller, 2024). According to Kitschelt (2012), the decline of unionization may have also contributed to the shift of social-democratic parties towards more centrist positions on redistribution, as they sought to attract middle-class constituencies drawn to their cosmopolitan stances on non-economic issues. The political significance of declining unionization makes it important to investigate the drivers of this phenomenon. In this article, we argue for the role of automation as a key structural factor. In the next section, we develop the conceptual framework linking automation and unionization.

2.2 The role of automation

Historically, technological progress has always generated aggregate welfare gains paired with substantial distributional consequences (Goldin and Katz, 1998). Technological innovations create new opportunities for workers whose skills are complementary to the new technologies, while posing challenges to workers that are more substitutable. This creates winners and losers of technological change, at least in relative terms.

The computer revolution that took place from the 1980s onwards, with the widespread adoption of IT and computer-based technologies, has been identified as a main driver of rising wage inequality and educational premia both in the US and in Europe (e.g., Acemoglu and Autor, 2011). Computerization has mainly substituted workers in jobs involving mostly routine tasks, both cognitive and manual, while it has complemented workers in jobs involving mostly non-routine tasks. Since routine jobs were predominantly middle-skill and middle-income occupations, this technological shift has led to a phenomenon known as "labor market polarization". In essence, employment has grown at both ends of the wage and skill spectrum, while the traditional middle class has contracted. This trend has been well-documented in both the United States and Europe (Autor and Dorn, 2013; Goos et al., 2014).

Computerization has eliminated many decently paid clerical and blue-collar jobs, with displaced workers largely absorbed into lower-wage, non-routine service roles (e.g., drivers and personal care workers). In terms of wage gains, the main winners of computerization have been high-skill, typically college-educated workers employed in non-routine cognitive jobs. They have been strongly complemented by the new technologies, and their incomes have diverged from those of the declining middle class. The latter falls into the group of losers, along with low-skill workers, who have benefited less from technological advances, and faced additional wage compression due to competition from displaced middle-skill workers (Autor, 2015).

Industrial robots represent the next spurt of automation, made possible by the expansion of the capabilities of computer based technologies. As highlighted by Frey and Osborne (2017), the novel aspect of robotization compared to earlier computerization is the extension of automation also to non-routine tasks, which were previously relatively unaffected. For instance, mobile robotics extends automation from routine assembly line operations to non-routine, more complex production activities, and to new domains such as maintenance of industrial plants, demolition and construction, and mining. According to data from the International Federation of Robotics (IFR), the stock of operational robots at the global level has grown exponentially from the mid-1990s onwards, with an acceleration from the mid-2000s. Importantly, the robotization shock comes after the peak of the so-called "China shock", i.e., the displacement driven by surging imports from China from the end of the 1980s until the Great Financial Crisis of 2007. At the same time, robotization precedes the widespread adoption of AI tools, which began gaining momentum in 2022. As a result, robotization stands out as the primary driver of structural change during the period of analysis (2002–2018), providing a key motivation for our focus.

A growing literature has provided evidence on the economic effects of automation through robots. Most studies adopt a regional level approach to focus on robot exposure, akin to the one adopted for evaluating the impact of the China shock (Autor et al., 2013). In this approach, stronger exposure to robotization is attributed to areas that were ex-ante specialized in industries witnessing higher robot adoption in subsequent years. In the US, Acemoglu and Restrepo (2020) find that stronger automation exposure reduces both employment and average wages at the commuting-zone level. The negative impact on employment is stronger in the manufacturing sector, particularly in industries such as automotive that are more exposed to robotization. Yet, it also extends to non-manufacturing sectors such as construction, personal services, and retail. The effect is more pronounced for workers with no college degree, for men in general, and for blue-collar workers employed in routine manual jobs, assembly, and related occupations. The negative effect of robotization on wages is mainly felt in the lower half of the wage distribution, thus contributing to the rise in wage inequality. Similar results for the US are also obtained by Borjas and Freeman (2019). Cross-country evidence of negative employment effects of robotization are found by Chiacchio et al. (2018) on 6 European countries, and by Carbonero et al. (2020) and Chen and Nabar (2018) on larger sets of countries including both advanced and emerging economies. Graetz and Michaels (2018), focusing on a sample of 17 industrialized countries, find that robot adoption increases productivity but reduces the share of hours worked by low-skill workers.

In Germany, Dauth et al. (2021) observe automation-induced job losses in manufacturing that are offset by employment gains elsewhere, especially in business services. Workers who are forced to switch plants, industries, or leave the manufacturing sector incur significant earnings losses. Overall, automation increases wage inequality by benefiting workers employed in occupations that are complementary to robots, such as managers and technical scientists, and penalizing substitutable workers like machine operators. Similar evidence is found by Bessen et al. (2023) for the Netherlands, and Dottori (2021) for Italy.

Overall, the surge in robotization has driven substantial distributional consequences, favoring mostly high-skill individuals vis-à-vis others. The losers of robotization tend to be concentrated in specific manufacturing industries, and in geographic areas where such industries were historically concentrated. In turn, these distributional effects have been found to be politically consequential. In the US, Frey et al. (2018) find that support for the Republican candidate Donald Trump in the 2016 presidential election was higher in local labor markets more exposed to robot adoption. In Europe, higher exposure to robotization at the regional level has been found to determine higher support for radical-right parties (Anelli et al., 2019; Dal Bó et al., 2023; Caselli et al., 2021; Milner, 2021). Similar results have been found at the individual level. For instance, Anelli et al. (2021) find that individuals more exposed to robot adoption, based on their positioning in the labor market, are more likely to support radical-right parties. Analogous evidence has been obtained measuring automation exposure through the automatability of one's occupation (e.g., Gingrich, 2019; Im et al., 2019; Milner, 2021).

While much is known about the economic and political effects of robotization, the impact on unionization has remained largely underexplored. Yet labor unions may play a crucial role in shaping the implications of automation, both in terms of distributional effects on the labor market, and in terms of the ensuing political consequences. From an economic perspective, Kristal (2013) and Kristal and Cohen (2015) find that computerization in the US increased wage inequality—and decreased the overall labor share in favor of the capital share of income—not only through direct labor market effects, but also through weakening unions. From a political perspective, Kitschelt (2012) suggests that the turn to the radical right of blue-collar constituencies experiencing economic distress may be partly explained by the decline of trade unions. Unionization decline may help make sense of a puzzle that has been often pointed out in the literature: why constituencies most penalized by automation (and globalization) are turning towards radical-right forces, rather than supporting parties with clearly redistributive stances (e.g., Kitschelt, 2012; Betz, 1993, 1994; Betz and Meret, 2012).

Radical-right parties are skeptical, or at a minimum ambiguous about redistribution (Kitschelt and McGann, 1997; Rovny, 2013); they are unsupportive of active labor market policies (Enggist and Pinggera, 2022), and of workers' rights more in general (Greilinger and Mudde, 2024; Rathgeb and Busemeyer, 2022). Unionization decline may explain why these parties are nevertheless successful at attracting the losers of structural economic changes such as automation. In fact, Rennwald and Mosimann (2023) have documented that non-unionized workers are less inclined to support parties that cater to their redistributive needs, and more likely to realign their vote based on cultural preferences. As observed by Colantone and Stanig (2018), the weakening of the link between blue-collar constituencies and left-wing parties opened the space for options that assign a central role to economic nationalism and nativism rather than the welfare state. Levi (2017) underscores that "the decline of labor unions has also facilitated the rise of populism by eliminating a source for a framework for understanding the situation of workers". Along these lines, Frymer and Grumbach (2021) discuss the feedback loop between union decline and racial prejudice. Overall, investigating the link between robotization and unionization is key to improve the understanding of both the economic and the political implications of automation.

From a theoretical standpoint, a main way in which robotization may reduce unionization is through a compositional effect in the labor market. The available evidence (e.g., Acemoglu and Restrepo, 2020; Dauth et al., 2021) suggests that robot adoption is higher in traditionally unionized manufacturing industries (e.g., automotive), and tends to shift employment towards less unionized industries (e.g., in logistics and personal care services), where it is also more difficult for unions to make inroads among workers. The result is a compositional change in employment that reduces overall unionization. This is likely to be visible not only at the country level but also at the sub-national level. The regional level is in fact most interesting for the purpose of our analysis, as it is where the role of unions as *contextual* factors may be particularly consequential, both economically and politically.

Besides this compositional effect, robotization may also reduce unionization within industries that are relatively more exposed to it. This could happen for at least three reasons. First, as suggested by Meyer (2019) in a study on computerization, the shift in the composition of the workforce induced by robots—i.e., relatively fewer blue-collar workers and more high-skill technical scientists and managers—may increase skill heterogeneity among workers, reducing their incentives for collective action. Relatedly, Checchi et al. (2010) suggest that rising earning inequality contributes to unionization decline by eroding solidarity among workers. Second, robotization may reduce the overall size of the workforce employed at each given plant; in turn, we know that unionization tends to be lower at smaller establishments (e.g., Oesch, 2006). Third, as suggested by Kristal (2013) for computerization, robotization may structurally diminish the bargaining power of workers, and thus inherently reduce their incentives to unionize. In our empirical analysis, we investigate the impact of robotization on unionization both at the regional level and at the industry level.

2.3 Automation and other factors of unionization decline

Our work is related to other streams of research that have investigated different factors of unionization decline. Brady (2007) categorizes these factors into three main families: institutional, solidaristic, and economic.

Institutional explanations argue that cross-national differences in unionization levels, as well as their trends, are largely shaped by institutional arrangements. These include union access to workplace representation, selective incentives such as union-administered unemployment schemes, recognition of employers through corporatist institutions, and closed-shop arrangements mandating union membership (Ebbinghaus and Visser, 1999). Government policies aimed at improving work security can also influence workers' incentives to join unions (Checchi and Lucifora, 2002). In the US, shifts in public policy seem to have played a key role in the decline of unionization rates (Feigenbaum et al., 2018; Farber, 2005; Hacker and Pierson, 2010; Lichtenstein, 2013).

Solidaristic explanations for unionization decline focus on the shift from class-based politics to identity- and status-oriented politics (Brady, 2007). For instance, Hechter (2004) argues that the expansion of welfare states has reduced the need for unions, as general welfare provisions now provide the social insurance that unions once offered exclusively to their members, leading to a decline in unionization and class consciousness.

Economic explanations consider both the short-term effects of business cycle fluctuations, and the long-term role of structural transformations such as de-industrialization (Brady, 2007). Studies that are most closely related to our work have investigated the link between globalization and unionization. Globalization entails easier international trade and capital mobility. Employers can exploit globalization to weaken workers' bargaining power, using the threat of production offshoring to countries with lower labor costs and lower unionization. According to Becher and Stegmueller (2020), economic shocks from global markets weaken labor unions, diminishing their influence on political representation and legislative support for compensatory policies, exacerbating inequality and heightening dissatisfaction with democratic processes. Slaughter (2007) finds a strong correlation between rising foreign direct investment and falling unionization in the US. Becher and Stegmueller (2024) find that import competition reduces district-level unionization, which in turn decreases legislative support for policies compensating economic losers and weakens opposition to trade deregulation. Ahlquist and Downey (2023) show that stronger import competition from China leads to a slight decline in unionization within manufacturing, but also to an increase in union membership in other sectors, such as healthcare and education, where unions are stronger. This suggests that structural forces may operate through compositional changes in the economy.

Technological progress, and in particular robotization, is the most important structural economic change taking place over the period we study. The consequences of robotization on labor unions are understudied. This study contributes to the literature on the economic drivers of unionization by focusing on this key dimension of structural change. In a study parallel to ours, Balcazar (2024) investigates the impact of robot adoption on unionization in the US. He finds that higher robotization at the level of congressional districts is related to lower unionization rates, and to lower responsiveness of elected representatives with respect to unions' interests, particularly on policies aimed at compensating the losers from international competition. Along similar lines, Agnolin (2025) finds that in areas of the US witnessing higher robot adoption candidates are less likely to come from a working-class background. This may contribute to weaken the representation of workers' interests in the democratic process.

3 Novel data on union density

Comparative research on labor unions has long been hampered by data limitations. As early as the beginning of the 1980s, Lipset (1983) observed that "a comparative analysis of working class movements in western society is limited by an obvious methodological problem: too many variables and too few cases". As Ahlquist (2017) aptly notes, "we have too many explanations chasing too few data points that are themselves interdependent in both time and space", and therefore recommends "research designs explicitly taking advantage of heterogeneity in context and population". To pursue such paths, better disaggregated data are required (Pontusson and Rueda, 2010).

At present, the only available source of comparable data on unionization across countries is the Visser (2019) dataset (ICTWSS). This is based on information collected and compiled by third parties within each country.[2] The main limitation of this dataset is that it only provides information on the overall unionization rate at the country-year level. Such aggregate data constitute an important constraint to unionization research. A sub-national

[2]Since 2021, the Visser (2019) database has been maintained as the OECD/AIAS ICTWSS database, reflecting a collaborative effort between the OECD and AIAS to ensure its continuation following Professor Visser's retirement. This version builds upon and consolidates earlier editions of the ICTWSS database and is publicly accessible at www.oecd.org/employment/ictwss-database.htm.

and sectoral analysis is in fact required not only to achieve clean identification of causal effects of unionization factors, but even just to address descriptive questions on unionization decline.

In the US, researchers have leveraged more fine-grained unionization data available via the Current Population Survey, where large samples of individual observations allow for meaningful aggregation at the sub-national and industry level (see, for instance, Ahlquist and Downey, 2023; Farber et al., 2021). Recently, a growing stream of research has also utilized the fine-grained data compiled by Becher et al. (2018) based on administrative records, with Balcazar (2024) and Becher and Stegmueller (2020, 2024) being notable examples.

For most European countries, instead, the national-level figures provided by the Visser (2019) database are the only available data on unionization. In fact, official national statistics only rarely collect or provide data on union density at any level, and administrative sources allow to retrieve such information only in a few contexts. While trade unions generally maintain records of their membership numbers, these records are often based on varying collection procedures, definitions of union membership, criteria for including students and retirees, and update frequencies. Moreover, statistics derived from the main trade unions' records do not account for the presence of union members in smaller trade unions.

Against this backdrop, the first contribution of this paper is to assemble a novel database on region- and industry-level unionization rates across 15 western European countries.[3] We obtain these unionization figures combining data from the European Social Survey (ESS) with country censuses over the period 2002-2018, using dynamic multilevel regressions with post-stratification (MRP, Park et al., 2004; Gelman et al., 2019). Our approach is very intuitive. First, we employ ESS data to estimate the predicted probability of unionization for

[3]These are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the UK.

different types of individuals, defined as combinations of age, gender, education, occupation, industry, and region of residence. Then, we compute the unionization rate at the regional level by taking a weighted average of unionization probabilities for different types of individuals. The weights are given by the share of each type of individuals out of the regional population, as obtained from census data. Similarly, we obtain industry-level figures by using as weights the share of each type of individuals out of all workers employed in a given industry, within each country.

This methodological approach allows us to obtain unionization estimates that are representative at the sub-national and industry level, and comparable across countries. It is an innovative application of the methodology initially proposed by Gelman and Little (1997) and Park et al. (2004), and popularized in applied political science research by Lax and Phillips (2009).[4] Thus far, this approach has mostly been used for predicting public opinion at the sub-national level in the US. We provide a novel cross-country application in the European context, focusing on unionization as an outcome.

3.1 Unionization estimates and validation

In the first step of the MRP approach, we employ individual-level data from nine rounds of the European Social Survey to estimate the probability of unionization for different types of individuals. As a starting point, we estimate models of the following general form:

$$Pr(Union_i = 1) = F(\text{gender}, \text{age}, \text{edu}, \text{occupation}, \text{industry}, \text{region}, \text{round}, \boldsymbol{X_r}) \quad (1)$$

where $Union_i$ is an indicator variable for whether the respondent is a union member, and $F(\cdot)$ is the probit link. The probability of unionization is modeled as a function of several individual characteristics: gender, age, education, occupation (ISCO 2-digit),

[4] See also Leemann and Wasserfallen (2020) for a textbook treatment.

industry (NACE 2-digit), and region of residence (NUTS 2-digit).[5] In addition, we include information on time—i.e., the ESS round in which the individual is observed—and a vector $\boldsymbol{X_r}$ of regional, pre-sample variables. This includes: the employment share of low- and medium-skill workers, the employment share of services, the employment share of low- and medium-tech industries, the employment share of primary sector, the employment share of finance and business services, and the share of foreign-born workers.[6] Importantly, we estimate the probit models separately for each sample country.

We explore a rich space of sixteen alternative specifications for the probit function $F(\cdot)$ in Equation 1, following the dynamic MRP approach developed by Gelman et al. (2019) to model time variation. These specifications feature different combinations of random effects and time trends based on the whole set of predictors. The full list of models can be found in Online Appendix A. The baseline model, selected via cross-validation, has the following form:

$$\Pr(Union_i = 1) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{occ} + \alpha^{ind} + \alpha^{region} + \alpha^{occ\_1d,\ edu} + \beta \cdot round + \xi \cdot \boldsymbol{X_r}) \quad (2)$$

where the α terms are random effects for gender, age category, education level, occupation, industry of employment, region of residence, and the combination of (ISCO 1-digit) occupation and education level. The specification includes also a time trend, captured by the ESS *round* variable, and the vector $\boldsymbol{X_r}$ of variables controlling for cross-regional differences in pre-sample conditions.

The baseline model outlined in Equation 2 is chosen through cross-validation in order to optimize the predictive performance at the region and industry level. In practice, in each country we: randomly split the sample into $K = 10$ folds; estimate each model on the training set (i.e., excluding fold k); form unionization predictions for fold k; and iterate over folds to obtain a vector of predictions for all observations. We then evaluate the accuracy of

[5] Industries are classified according to the Revision 1.1 of the NACE classification. Regions are at the NUTS-2 level for all countries except Germany and the UK, where data are only available at the NUTS-1 level due country-specific privacy limitations.

[6] Low-skill workers have up to lower secondary education. Medium-skill workers have upper secondary and post-secondary non-tertiary education.

predictions for each relevant group of observations, i.e., regions and industries within each country.

We denote as Ω_g the set of N_g observations in a given group g (e.g., region Île-de-France, or textile industry in France). The average predicted probability for group g is $\hat{P}_g = \frac{\sum_{i \in \Omega_g} \hat{P}_i}{N_g}$. Analogously, the observed empirical frequency in the ESS data is given by $F_g = \frac{\sum_{i \in \Omega_g} \mathbb{1}(Union_i = 1)}{N_g}$. The group-wise calibration RMSE based on the whole set of groups G within a given country is $\text{RMSE}_\text{G} = \left(\frac{\sum_g (\hat{P}_g - F_g)^2}{G} \right)^{\frac{1}{2}}$. This metric compares the (cross-validated) predicted probabilities for each region or industry g with the empirical frequency of unionization among survey respondents from that region or industry. We then rank specifications according to their RMSE performance within each country. The baseline specification in Equation 2 is the model that ranks on average best across all countries in the study.[7]

Our main results on the impact of robotization on unionization are robust to employing union density figures obtained from any of the sixteen different probit specifications. For ease of exposition, in the tables we only report two sets of robustness checks. First, we show that results are robust to using the highest ranked prediction model by country. In fact, the baseline model is not necessarily the best performing within each country. In this respect, Figure A1 of the Online Appendix displays the ranking of the different models in all sample countries. Second, we show results that rely on an alternative model we chose on conceptual grounds, to allow for sector-specific differential time trends. The specification is:

$$\Pr(\text{Union}_i = 1) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{occ} + \alpha^{ind} + \alpha^{region} + \beta \cdot round + \gamma^{ind\_sector} \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r}) \tag{3}$$

were $\gamma^{ind\_sector} \cdot round$ denotes time trends that are specific to NACE sub-sections, i.e., aggregations of 2-digit industries.

From the probit estimates, we compute predicted probabilities of unionization for each

[7]Full details on the cross-validation approach are provided in Online Appendix A.

socio-demographic type of individuals, in each country. Types, denoted by λ, are defined by combinations of gender, age group, education level, occupation, industry, and region of residence, at a given point in time (i.e., ESS round). The probability of unionization for each type is denoted by θ_λ. To illustrate, we can retrieve the probability of unionization for women in their 30s, with a MSc degree, working as doctors, in the healthcare industry, in the Paris region (Île-de-France), in each ESS round.

In the second step of the MRP approach, we obtain unionization rates at the region (or industry) level through post-stratification of the predicted probabilities of unionization for the different types of individuals. To do so, we employ as weights the frequency of each type within each region (or industry), obtained from census data.[8]

Specifically, the unionization rate of region r at time t is obtained as:

$$\text{Union Density}_{rt} = \sum_{\lambda \in \Lambda_{rt}} \left(\frac{N_\lambda}{N_{rt}} \right) \theta_\lambda$$

where Λ_{rt} is the set of types in region r at time t; N_λ is the number of λ-type individuals in the region, and N_{rt} is the total population in the region, based on census data. The unionization rate represents the share of workers in the region that are union members.[9]

The unionization rate of industry j, in country c, at time t is:

$$\text{Union Density}_{jct} = \sum_{\lambda \in \Lambda_{jct}} \left(\frac{N_\lambda}{N_{jct}} \right) \theta_\lambda$$

where Λ_{jct} is the set of types in industry j of country c at time t; N_λ is the number of λ-type individuals working in the industry, and N_{jct} is the total number of individuals working in the industry, based on census data. The unionization rate represents the share of workers in the industry that are union members.

[8] Full details about the sources of census data, and their harmonization for the construction of the weights, are available in Online Appendix B.

[9] For the computation, we consider all workers in the labor force of the region—i.e., employed plus unemployed—excluding self-employed workers.

3.2 External validation and granular evidence on unionization decline

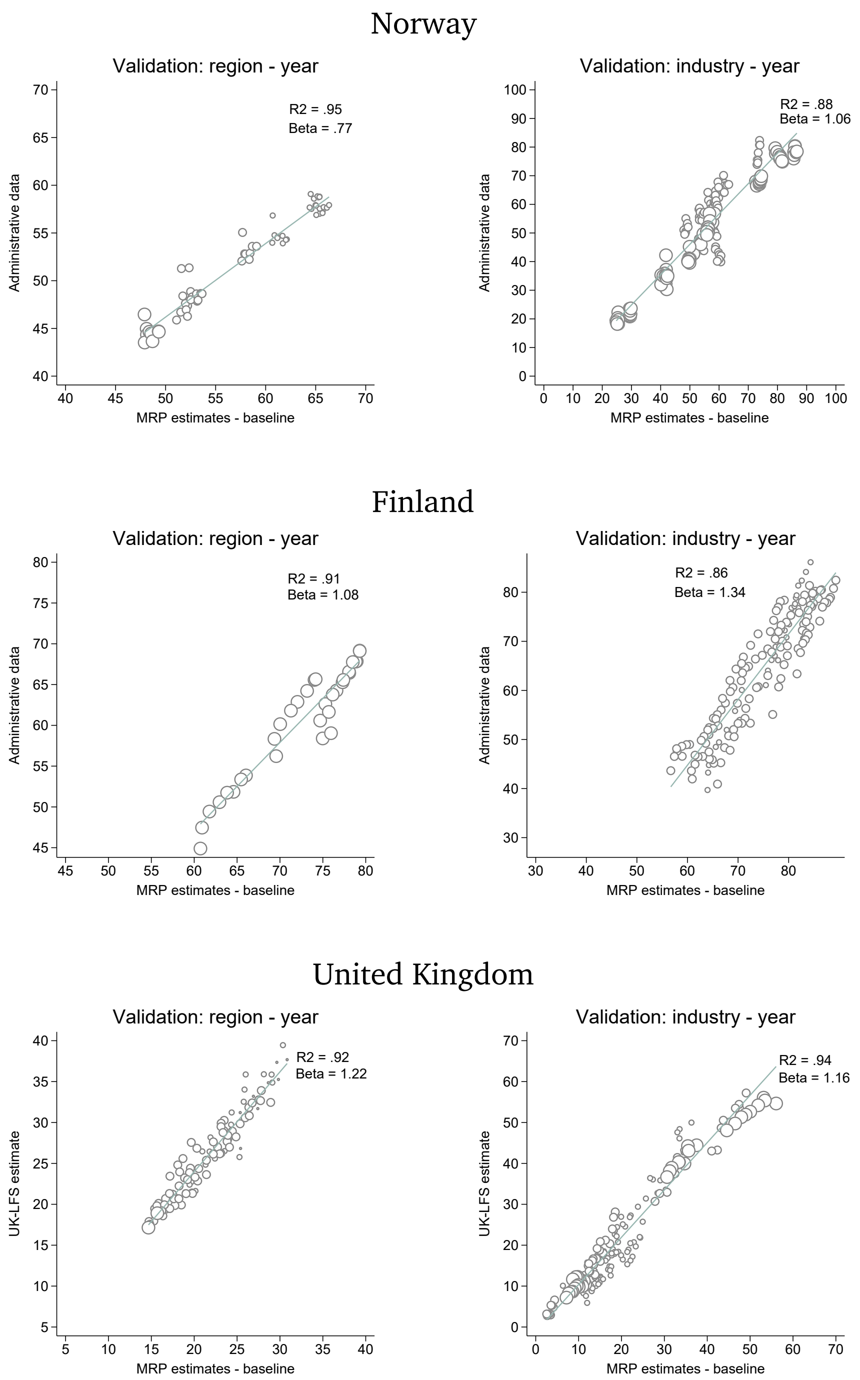
We validate the unionization estimates against official figures for Norway and Finland, two countries that provide administrative records of unionization at the sub-national and industry level. We perform a similar comparison for the UK, relying on data from the UK Labor Force Survey, that is representative at the granular level. This external validation exercise is presented in Figure 1. The left panels compare unionization rates at the region-year level, while the right panels focus on the industry-year level, with industries aggregated at the NACE sub-section level. In all panels, the horizontal axis reports the MRP unionization estimates, as obtained through the baseline model, while the vertical axis reports the external data. It is important to remark that neither the administrative records for Norway and Finland, nor the information about union membership in the UK-LFS survey data, were employed as inputs in the MRP estimation. Hence, the MRP estimates and these external benchmarks are entirely independent.

The correlation between our unionization estimates and the external data is very high. In particular, focusing on the regional figures, a regression of external data on MRP estimates yields R-squared values of 0.95 in Norway, 0.91 in Finland, and 0.92 in the UK. For industry-level variation, the R-squared values are slightly smaller, yet the overall convergence between the MRP estimates and the external data is still tight.

As a further validation exercise, we aggregate our sub-national estimates at the country-year level, and compare the resulting unionization figures with data from the Visser (2019) dataset. Reassuringly, the correlation between our aggregated data and Visser's data is very high: 0.96. Based on Visser's data, the average annual decline in unionization across the 15 western European countries in our sample was 0.384 percentage points between 2002 and 2018. Our data replicate closely this trend, showing an yearly average decline of 0.378 percentage points.

The advantage of our data is that they allow us to provide evidence of unionization

Figure 1: External validation



Note: Each observation in the chart corresponds to either a region (left panels) or an industry (right panels) in a specific survey year. The size of the circles is proportional to group size (regional population or industry employment). External validation data are from administrative records provided by Statistics Norway and Statistics Finland, and from the UK Labor Force Survey.

decline also at the region and industry level. This is done in Table 1. Specifically, in column 1 we regress the unionization rate in a given region and year on a time trend, controlling for the unionization rate of the region at the beginning of the sample (i.e., 2002), and absorbing average differences in unionization across countries through country fixed effects. The estimated coefficient on the time trend, -0.384, points to a strong decline in unionization over time, by around 3.84 percentage points on average every ten years. This is in line with the macro-level figures in Visser (2019). In column 2, we show that the declining trend is steeper in those regions that were initially more unionized. For instance, in a region with 80% initial unionization (e.g., East Middle Sweden or North and East Finland), the trend is expected to be around 5.4 percentage points every ten years, while it is around 3.3 p.p. in a region starting at 10% unionization rate (e.g., Lorraine, or Franche-Comté). In columns 3-4, we replicate the same analysis on unionization by country-industry (NACE subsections). Also in this case, we find evidence of a significant declining trend, by around 3.81 percentage points on average every ten years. Unlike the regional evidence, the decline is not more pronounced in country-industries that were initially more unionized.

3.3 Union density in Western Europe

Our novel dataset on union density spans over around two decades (2002-2018) and includes 15 western European countries. Figure 2 illustrates union density across Western European regions in 2002 and 2018, representing the first and last years currently covered by our data. The results reveal a rich heterogeneity in unionization levels, with countries such as Finland, Sweden, and Norway exhibiting union densities exceeding 50%, while countries like France, Portugal, and Greece display significantly lower levels of unionization. Beyond the well-documented variation between countries—rooted in deep-seated historical and institutional factors—our data also uncovers substantial intra-country variation. For example, union density is notably higher in northern Spain compared to the south, and in the industrial regions of West Germany compared to other parts of the country. In France,

Table 1: Unionization decline

Dep. Var.:	Union Density			
	(1)	(2)	(3)	(4)
Analytical Unit:	Region		Industry	
Year	-0.384***	-0.300***	-0.381***	-0.350***
	[0.035]	[0.048]	[0.025]	[0.037]
Initial Union Density	0.911***	6.835**	0.928***	2.712
	[0.021]	[2.880]	[0.012]	[1.779]
Initial Union Density X Year		-0.003**		-0.001
		[0.001]		[0.001]
Country FE	X	X	X	X
Observations	1,467	1,467	2,700	2,700
R-squared	0.984	0.984	0.987	0.987

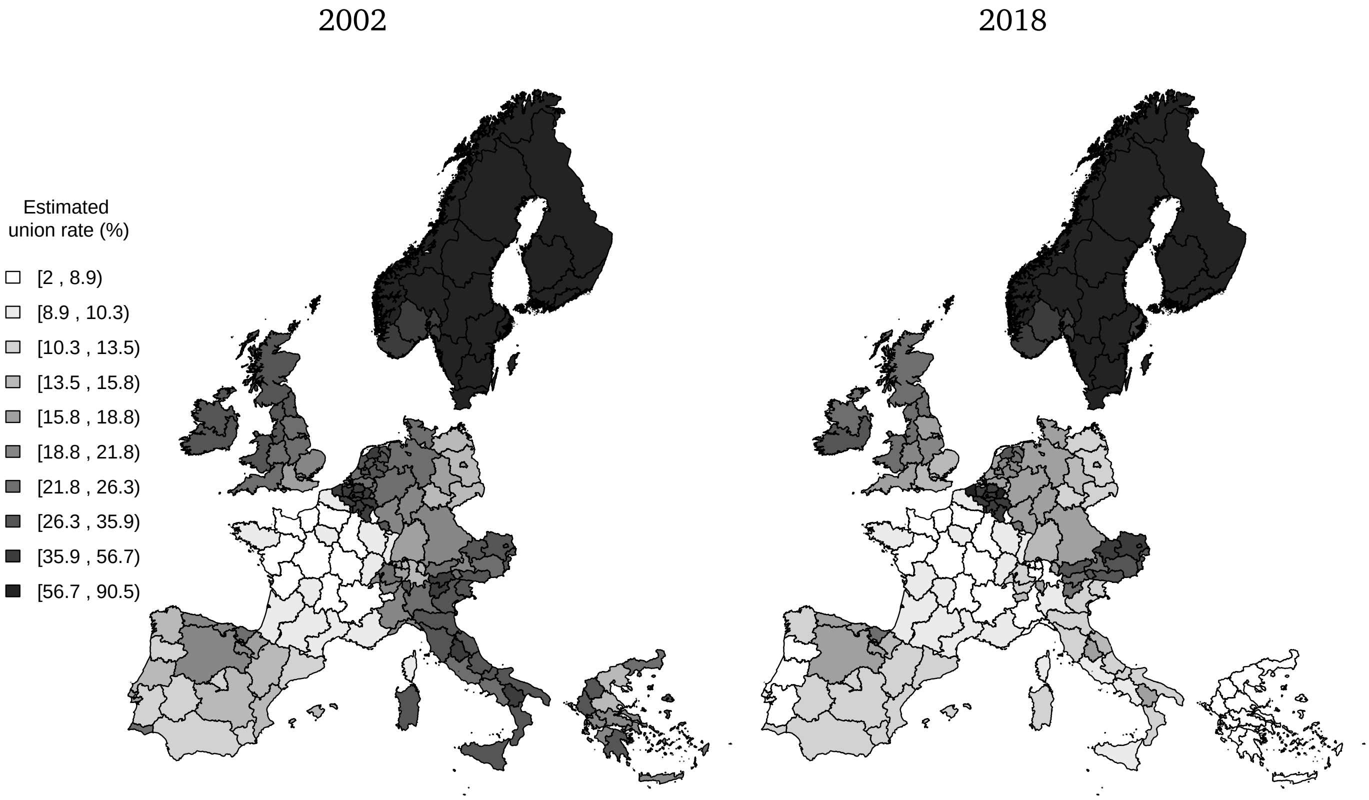
Note: In columns 1-2, the dependent variable is the unionization rate at the region-year level; in columns 3-4, the dependent variable is the unionization rate at the country-industry-year level. Unionization estimates are obtained through MRP employing the baseline model. Initial union density refers to 2002. The models are estimated by OLS and include country fixed effects. Standard errors are clustered by region in columns 1-2, and by country-industry in columns 3-4.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

union density remains uniformly low, ranging from 8% to 10% in most regions.

Figure 3 unveils a pronounced trend of deunionization in several European regions, especially in Greece, Ireland and Italy. Importantly, the implications of a higher or lower union density levels are shaped by complex, context-specific factors that can vary significantly across nations. As Ahlquist (2017) insightfully notes, trade unions are inherently diverse organizations, differing across multiple dimensions. Therefore, while cross-national comparisons of unionization trends should be approached with nuanced understanding, subnational variation provides a valuable analytical lens. To rigorously investigate trade unions, it is crucial to employ research designs that capitalize on this subnational diversity.

In the empirical analysis, we will connect within-country variation in unionization dynamics with regional variation in automation exposure, absorbing between-country variation through country-year fixed effects.

Figure 2: Estimated union density by region



3.4 Regional variation in union density and electoral outcomes

Besides describing unionization decline at the granular level, our sub-national unionization data also allow us to document, for the first time, the association between union density and electoral outcomes across European regions. For instance, we can provide empirical insights into the long-assumed positive correlation between unionization and support for mainstream left parties at the electoral-district level.

Specifically, we relate regional unionization figures to support for different party families at the district level, estimating specifications of the following form:

$$\text{Share}_{\ell dt} = \alpha + \beta \text{ Union Density}_{r(d)t} + \eta_{ct} + \epsilon_{\ell dt}, \qquad (4)$$

where $\text{Share}_{\ell dt}$ denotes the cumulative vote share for parties belonging to family ℓ, in

Figure 3: Estimated union density variation (2002-2018)

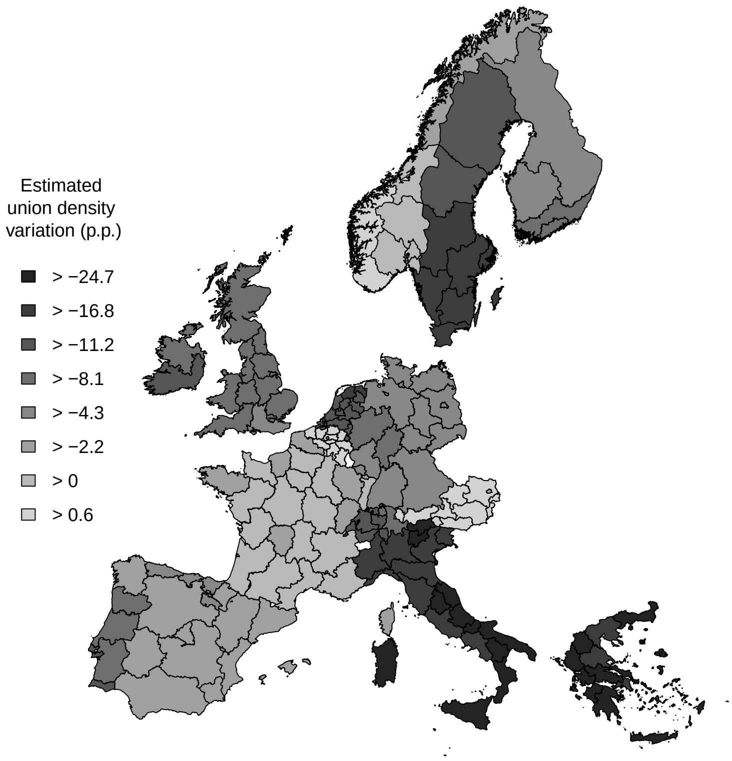


Table 2: Unionization and voting

	(1) Radical Left	(2) Mainstr. Left	(3) Mainstr. Right	(4) Radical Right	(5) Ethno-Region.	(6) Other Single
Union Density	-0.211***	0.861***	-1.034***	-0.104**	0.494***	-0.023
	[0.055]	[0.176]	[0.189]	[0.049]	[0.171]	[0.027]
Country-year FE	X	X	X	X	X	X
Observations	7,157	7,157	7,157	7,157	7,157	7,157
R-squared	0.530	0.324	0.362	0.731	0.169	0.889
Std dev. Y	4.939	11.94	13.52	4.102	6.501	2.009
Std dev. X	9.617	9.617	9.617	9.617	9.617	9.617
Magnitude	-0.412	0.693	-0.735	-0.244	0.731	-0.110

Note: The dependent variable is the cumulative vote share for parties belonging to a given party family at the electoral-district level. Unionization estimates are obtained through MRP employing the baseline model. The model is estimated by OLS and includes country-year fixed effects. The table reports the standard deviation of the dependent variable and of the unionization rate, along with the magnitude of their relationship, after residualizing with respect to country-year fixed effects. Standard errors are clustered by region-year.
*** p<0.01, ** p<0.05, * p<0.1

electoral district d, in country c, in the election taking place in year t. The sample includes elections held between 2002 and 2018. Union $\text{Density}_{r(d)t}$ is the unionization rate in the (NUTS-2) region r where district d is located, as measured in year t.[10] η_{ct} are country-year fixed effects, which are equivalent to election fixed effects. Standard errors are clustered at the region-year level.

We focus on six party families: radical left, mainstream left, mainstream right, radical right, ethno-regionalist, and other, single-issue parties. Parties are assigned to a given family based on the Manifesto Project classification (Volkens et al., 2016), except for the radical left and right families, which are identified based on consensus in the literature.[11]

Results are presented in Table 2. Consistent with the expectations, we detect a positive association between unionization and support for parties of the mainstream left. There

[10] In many cases, a district is itself a NUTS-2 region. In other cases, a given region may contain two or more districts. Importantly, a district is always fully contained within the boundaries of one single NUTS-2 region, with no overlaps.

[11] See Online Appendix C for full details on party classification. Some residual parties are not included in the analysis, as they are too small to be coded by the Manifesto Project. The same applies to three "agrarian" parties from Norway, Sweden, and Finland.

is also a positive correlation between unionization and support for ethno-regionalist parties. Conversely, union density is uncorrelated with support for single-issue parties, and negatively associated with support for the other party families.

4 Automation and unionization

We investigate the impact of automation exposure on unionization at the regional level estimating regressions of the following form:

$$\text{Union Density}_{\text{rt}} = \alpha + \beta \text{ Regional Robot Exposure}_{\text{rt}} + \eta_{ct} + \eta_r + \epsilon_{rt}, \tag{5}$$

where r indexes (NUTS-2) regions, and t years. Union $\text{Density}_{\text{rt}}$ is the unionization rate in region r and year t. The terms η_{ct} and η_r are country-year and region fixed effects, respectively. Regional Robot $\text{Exposure}_{\text{rt}}$ is the exposure to industrial robot adoption in region r, evaluated in year t. Following Acemoglu and Restrepo (2020), this is measured as:

$$\text{Regional Robot Exposure}_{\text{rt}} = \sum_j \frac{L_{rj}^{\text{pre-sample}}}{L_r^{\text{pre-sample}}} * \frac{R_{cj}^{t-1} - R_{cj}^{t-k}}{L_{cj}^{\text{pre-sample}}}, \tag{6}$$

where r indexes regions, j manufacturing industries, c countries, and t years. $R_{cj}^{t-1} - R_{cj}^{t-k}$ is the change in the operational stock of industrial robots over the past k years, in country c and industry j. In the baseline analysis, $k = 3$. This change is normalized by the pre-sample number of workers employed in the same country and industry, $L_{cj}^{\text{pre-sample}}$, yielding a ratio that measures the intensity of robot adoption at the industry level.

The regional-level exposure is a weighted summation of the industry-level changes, where the weights capture the historical importance of each industry in the region. Specifically, each weight is the ratio between the number of workers employed in a given region and industry, $L_{rj}^{\text{pre-sample}}$, and the total number of workers employed in the same region, $L_r^{\text{pre-sample}}$. Weights are based on pre-sample figures, dating before the surge in the adoption of industrial robots observed from the mid-1990s onwards. Intuitively, regions that were initially specialized in industries in which the adoption of robots has later been faster are assigned stronger exposure to automation.

Data on robot adoption by industry are sourced from the International Federation of Robotics (IFR). They refer to eleven industries encompassing the whole manufacturing sector. These correspond mostly to NACE Rev. 1.1 sub-sections (details in Table D.1 of the Online Appendix). Employment data are from national sources or Eurostat. Detailed information on all data sources employed to measure automation exposure can be found in Online Appendix D.

To address potential endogeneity concerns, we employ an overidentified model with three instrumental variables capturing complementary aspects of technological progress that are relevant for robots. These instrumental variables exploit: the producer price index of computers, sourced from Federal Reserve Economic Data (FRED, US Bureau of Labor Statistics, 2023); and two indexes of advances in computing, specifically single-thread performance and number of transistors per microprocessor, both sourced from Rupp (2021). For each of these variables, we compute the instrument as follows:

$$\text{IV Regional Robot Exposure}_{\text{rt}} = \sum_j \frac{L_{rj}^{\text{pre-sample}}}{L_r^{\text{pre-sample}}} * Rep_j * \Delta Index_{t-1,t-k}, \tag{7}$$

where $\Delta Index_{t-1,t-k}$ is the change in the relevant variable between $t-1$ and $t-k$, and Rep_j is an industry-level replaceability index—i.e., the share of hours worked within industry j in occupations replaceable by robots—as computed by Graetz and Michaels (2018) based on US Census data of 1980. These instruments are designed to capture the role of plausibly exogenous global technological shifts in robotics and computing (i.e., $\Delta Index_{t-1,t-k}$), which vary over time and are common across countries. These shifts have differential effects across industries based on their ex-ante predisposition to robotization (i.e., Rep_j). In turn, the impact on regions depends on the pre-sample composition of employment across industries (i.e., the terms $\frac{L_{rj}^{\text{pre-sample}}}{L_r^{\text{pre-sample}}}$).

Columns 1-2 of Table 3 display the baseline estimates of Equation 5. The first column reports ordinary least squares (OLS) estimates, while the second shows instrumental variables (IV) results, where regional exposure to robot adoption is instrumented as described above. In these regressions, we employ the baseline unionization estimates, obtained through MRP according to Equation 2. The coefficient on regional robot exposure is negative and precisely estimated in both columns, pointing to a negative effect of automation on unionization. The IV estimate is larger than the OLS one in absolute value. This is consistent with there being unobserved factors related at the same time with both higher automation and higher unionization. For instance, the presence of stronger unions may raise company incentives to accelerate automation to replace workers. In terms of magnitudes, according to the IV estimate in column 2, a one standard deviation increase in robot exposure (i.e., 17 robots per 100,000 workers) leads to a reduction in regional union density by 34.4% of a standard deviation. This figure is obtained net of country-year and region fixed effects, as per the Mummolo and Peterson (2018) approach.

These results are robust to employing union density figures obtained from any of the other 15 probit specifications used in the MRP approach. For ease of exposition, in Table 3 we only report two robustness checks. Specifically, in columns 3-4 we use unionization estimates obtained from the highest ranked prediction model by country. In columns 4-5, we employ unionization estimates based on the model outlined in Equation 3, which allows for sector-specific differential time trends. The results are very similar to the baseline evidence of columns 1-2, in terms of both statistical and substantive significance. Hence, for the next battery of robustness checks we only focus on the baseline unionization estimates.

Table 4 reports twelve robustness checks on the baseline IV result of column 2 in Table 3, which is replicated in the first row for convenience. Each row refers to a different estimation, and reports the coefficient and standard error on regional robot exposure. In rows 2-6,

Table 3: Regional robot exposure and unionization

Dep. Var.:	Regional Union Density					
	(1)	(2)	(3)	(4)	(5)	(6)
MRP Model:	Baseline		Best by country		Sector trends	
Regional Robot Exposure	-0.177***	-0.430**	-0.184***	-0.466***	-0.176***	-0.347**
	[0.051]	[0.171]	[0.052]	[0.177]	[0.048]	[0.161]
Country-year FE	X	X	X	X	X	X
Region FE	X	X	X	X	X	X
Estimator	OLS	2SLS	OLS	2SLS	OLS	2SLS
Observations	1,413	1,413	1,413	1,413	1,413	1,413
Std dev. Y	0.526	0.526	0.533	0.533	0.521	0.521
Std dev. X	0.421	0.421	0.421	0.421	0.421	0.421
Magnitude	-0.141	-0.344	-0.145	-0.368	-0.143	-0.281
First stage F-stat		15.45		15.45		15.45

Note: The dependent variable is the union density at the regional level, estimated through MRP according to the model indicated on top. The table reports the standard deviation of the dependent variable and of robot exposure, along with the magnitude of their relationship, after residualizing with respect to country-year and region fixed effects. Robust standard errors in brackets.
*** p<0.01, ** p<0.05, * p<0.1.

we augment the baseline specification with interactions between the year dummies and a number of initial regional conditions, measured pre-sample and specified in each row. By doing this, we control for differential regional trajectories—as related to the initial characteristics of the labor force—which might be confounded with exposure to automation. The results are not significantly affected.

In rows 7-10, we replicate the baseline regressions excluding the largest regions, which might arguably provide a sub-optimal approximation of the local labor market concept behind the theoretical and measurement approach by Acemoglu and Restrepo (2020). We drop, alternatively, regions in the top 10% or 25% of the distribution in terms of geographical area, either within each country (rows 7-8) or overall (rows 9-10). The results remain close to the baseline. In rows 11-13, we exclude from the computation of the instruments three industries that are particularly relevant for robotization. These are the automotive industry, which is the most robot-intensive, and the electronics and metal industries, which provide key robot inputs.[12] Results do not hinge on the inclusion of any

[12] In particular, we exclude the NACE sub-sections DM (manufacture of transport equipment), DL (manufacture of electrical and optical equipment) and DI-DJ (manufacture of other non-metallic mineral products and

Table 4: Additional robustness checks

Dep. Var.:	Regional Union Density
1) Baseline	-0.430**
	[0.171]
2) Year dummies * Initial share low-skill workers	-0.427**
	[0.176]
3) Year dummies * Initial share med-skill workers	-0.406**
	[0.174]
4) Year dummies * Initial share high-skill workers	-0.401**
	[0.174]
5) Year dummies * Initial share foreign workers	-0.431**
	[0.178]
6) Year dummies * Initial stock foreign workers	-0.417**
	[0.178]
7) Excluding largest 10% regions by country	-0.424**
	[0.193]
8) Excluding largest 25% regions by country	-0.414*
	[0.232]
9) Excluding largest 10% regions overall	-0.521***
	[0.178]
10) Excluding largest 25% regions overall	-0.708***
	[0.210]
11) Excluding automotive industry	-0.419*
	[0.228]
12) Excluding electronics industry	-0.397**
	[0.177]
13) Excluding metals and minerals industries	-0.434***
	[0.163]

Note: All reported coefficients refer to regional robot exposure. All models include country-year and region fixed effects. Robust standard errors in brackets.
*** p<0.01, ** p<0.05, * p<0.1.

one of these industries, thus corroborating the robustness of our findings.

Finally, we assess whether the estimated effect of automation is sensitive to the choice of different lags for the computation of regional robot exposure. In the baseline specification we consider the change in the operational stock of robots over the previous three years, i.e., between $t-1$ and $t-3$. In Figure E1 of the Online Appendix, we show that results are robust to considering alternative time periods, ranging from the previous two to six years.

5 Potential channels underlying the main effect

The main finding of the analysis, thus far, is that higher exposure to robotization reduces unionization at the regional level. This result could be driven by two non-mutually exclusive channels, as outlined in Section 2.2. First, there could be a systematic reduction of unionization *within* industries where robot adoption is higher. Second, automation may induce a reallocation of workers *between* industries, shifting employment from highly unionized industries towards less unionized ones. In this section, we explore both channels.

We begin by investigating the within-industry impact of automation on unionization, estimating the following specification:

$$\text{Union Density}_{\text{cjt}} = \alpha + \beta \, \frac{R_{cj}^{t-1} - R_{cj}^{t-k}}{L_{cj}^{\text{pre-sample}}} + \eta_{ct} + \eta_j + \epsilon_{cjt}, \tag{8}$$

where c indexes countries, j industries, and t years. η_{ct} and η_j are country-year and industry fixed effects, respectively. $\text{Union Density}_{\text{cjt}}$ is the unionization rate in industry j, country c, and year t. This is regressed on the industry-level exposure to automation, $\frac{R_{cj}^{t-1} - R_{cj}^{t-k}}{L_{cj}^{\text{pre-sample}}}$, with $k = 3$. Note that this is the same measure of industry exposure that is used as an input in Equation 6 to compute regional robot exposure.[13]

We instrument automation exposure using the same three indexes of technological progress employed in the regional analysis. Specifically, the three instrumental variables are defined as follows:

$$\text{IV Industry Robot Exposure}_{jt} = Rep_j * \Delta Index_{t-1,t-k}, \tag{9}$$

manufacture of basic metals and fabricated metal).

[13]Industrial robot adoption is recorded for 11 manufacturing industries, mostly at the NACE sub-section level, as presented in Table D.1 of the Online Appendix. Non-manufacturing industries are included in the analysis at the NACE subsection level of disaggregation, and have zero exposure to robot adoption, with industries in the primary sector (section A-B-C) being aggregated in a single industry. Results are substantially unchanged when excluding these industries from the analysis.

where Rep_j is the labor replaceability of industry j, and $\Delta Index_{t-1,t-k}$ is the change in one of the three indexes employed for the computation of regional instruments in Equation 7. In fact, the industry-level instruments are the inputs used for the computation of the instrumental variables at the regional level.

Table 5 shows the estimation results of Equation 8. The estimated coefficients on robot exposure are positive, although not statistically different from zero in the more robust specification shown in column (2). Coefficient estimates are also small in magnitude. In particular, according to the IV specification in column 2, a one standard deviation increase in robot exposure would yield an effect as small as 2.8% of a standard deviation in union density. These results are based on the baseline estimates of unionization. Findings are very similar when employing alternative unionization estimates, as reported in Table E.1 of the Online Appendix. Overall, the analysis does not support the conclusion that automation significantly alters unionization rates within industries. While the first-stage F-statistic is lower for industry variation than for regional variation, the comparison with the clearly detected and negative regional-level relationship between automation and unionization suggests that the null results at the industry level are likely to reflect a genuinely weak or ambiguous effect. In sum, the industry-level results do not support the idea that the decline in unionization driven by automation operates primarily through within-industry channels.

Lack of evidence in favor of the within-industry channel is not surprising in light of the empirical studies on the employment effects of automation in Europe (e.g., Dauth et al., 2021; Dottori, 2021). In fact, these studies suggest that employment declines in automated manufacturing industries occur primarily through reduced hiring rather than layoffs. Since incumbent workers are more likely to be unionized than new hires, halting turnover may not necessarily decrease union density, even as automation may lower incentives to unionize, as discussed in Section 2.2.

Next, we examine the second potential channel underlying the main region-level results, i.e., an automation-induced compositional change in the labor market, with a relative shift

Table 5: Industry-level analysis

Dep. Var.:	Industry Union Density	
	(1)	(2)
Industry Robot Exposure	0.264**	0.196
	[0.127]	[1.016]
Country-year FE	X	X
Industry FE	X	X
Estimator	OLS	2SLS
Observations	2,540	2,540
Std dev. Y	6.112	6.112
Std dev. X	0.869	0.869
Magnitude	0.0375	0.0279
First stage F-stat		4.507

Note: The dependent variable is the unionization rate at the country-industry-year level. Unionization estimates are obtained through MRP employing the baseline model. The table reports the standard deviation of the dependent variable and of robot exposure, along with the magnitude of their relationship, after residualizing with respect to country-year and industry fixed effects. Robust standard errors in brackets.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

in employment from more to less unionized industries. For this channel to be relevant, two conditions must hold: (1) industries that are relatively more exposed to automation witness a relative decline in employment; and (2) automation exposure is higher in industries that were more unionized at the beginning of the sample period. In Table 6, we provide evidence consistent with these two conditions.

In columns 1-2, we regress the national share of workers employed in a given industry on exposure to robotization. The instrumental variable results indicate that industries with higher exposure to robot adoption experience a decline in their employment share. Specifically, a one standard deviation increase in automation exposure determines a decrease of the industry employment share by about 8% of a standard deviation.

Columns 3-4 of Table 6 display linear regressions where the dependent variable is the average automation exposure by country-industry over the whole sample period (2002-2018), and the predictor is the initial unionization rate at the country-industry level, measured in 2002. Column 3 includes all industries, both manufacturing and non-manufacturing, while column 4 considers only manufacturing industries. These purely descriptive regressions indicate that industries that were initially more unionized experienced greater automation over time. Taken together, the results presented in Table 6 point to automation-induced reallocation of employment as the main channel underlying the negative impact of robotization on unionization identified at the regional level.

Figure 4 provides further descriptive evidence consistent with the overall argument of a broader labor market reconfiguration that penalizes traditional union strongholds. The figure displays changes in employment shares across broad sectors from 2002 to 2018, expressed relative to their levels in 2002. Employment shares are calculated by aggregating Eurostat employment data across the fifteen countries included in the study, grouped into broad sectors. Thus, the measure captures the average relative importance of each sector within the total employed population. The figure also displays the average initial union density levels in 2002, with darker and wider lines indicating sectors with higher initial union density. The data highlights a steady decline in employment shares for traditionally unionized sectors, such as Manufacturing, Transport and Communication, Utilities and Construction, where union density initially ranged between 30% and 40%. For instance, in Manufacturing—–where union density averaged around 35%—–employment share fell from 18% in 2002 to 13% in 2018. The only exception among highly unionized sectors is the public sector, which has expanded in terms of employment. Conversely, sectors with lower initial union density, such as Finance and Business Services, Hotels and Restaurants, and Personal and Community Services, have increased their share of total employment over time. Figure E2 shows that a similar overall pattern holds across different unionization regimes in Europe. Broadly speaking, the public sector stands out among the most highly unionized sectors in all regimes, suggesting it may have acted as a buffer on the overall decline in unionization. In contrast, the share of employment in manufacturing has declined sharply across all European unionization regimes.

This pattern plausibly reflects systematic processes with an economic and social logic. Robots have primarily automated activities at industrial plants, i.e., contained environments where most manufacturing workers were concentrated, facilitating union organization and collective action. In contrast, automation has spurred employment growth in sectors that are more challenging to unionize, where collective action is harder to coordinate. Logistics, for example, has expanded significantly due to increased productivity and lower production costs driven by automation. However, logistics workers often operate in isolation, and services are frequently outsourced to companies that are detached from production and easily replaceable. Such factors create substantial barriers to unionization and hinder effective collective action in the expanding industries.

5.1 Complementary individual level analysis

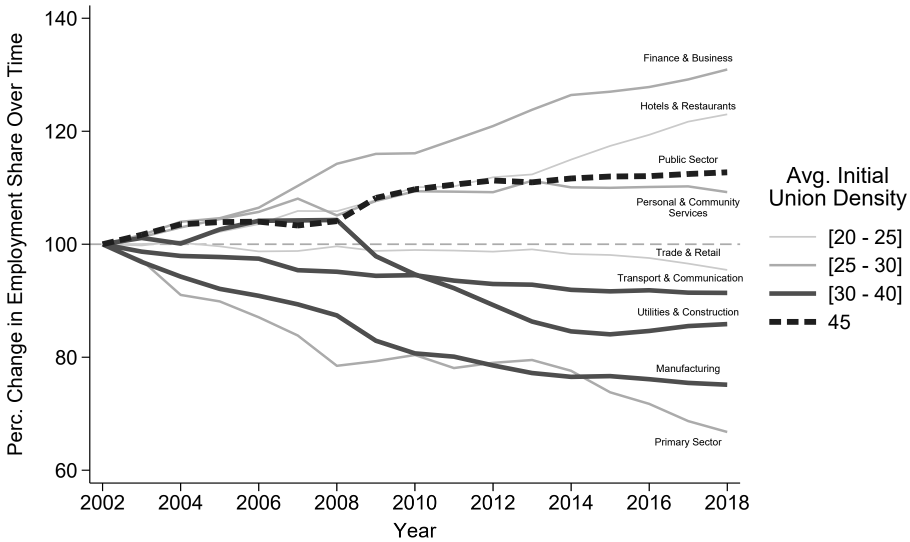
To further substantiate our findings and provide additional insight into the mechanism through which automation affect unionization, we now conduct an individual-level analysis

Table 6: Automation, employment share, and initial unionization

Panel A Automation and Empl. Share			**Panel B** Initial Unionization and Automation		
Dep. Var.	Employment share		Dep. Var.	Avg. robot exposure	
	(1)	(2)		(3)	(4)
Industry robot exposure	0.042	-0.145**	Initial union density	0.016**	0.132***
	[0.030]	[0.061]		[0.007]	[0.044]
Country FE	X	X	Country FE	X	X
Industry FE	X	X			
Estimator	OLS	2SLS	Estimator	OLS	OLS
Industries	All	All	Industries	All	Manuf.
Observations	2,540	2,540	Observations	300	120
Std dev. Y	1.610	1.610	Std dev. Y	1.113	1.568
Std dev. X	0.891	0.891	Std dev. X	9.769	5.023
Magnitude	0.0233	-0.0801	Magnitude	0.141	0.424
First stage F-stat		5.890			

Note: In Panel A: the dependent variable is the share of workers employed in a given country-industry-year relative to the total number of employed workers in the corresponding country-year; the reported coefficients refer to industry robot exposure. In Panel B: the dependent variable is the average industry-level robot exposure over 2002-2018, measured at the country-industry level; the reported coefficients refer to initial union density in 2002, based on baseline MRP unionization estimates. Standard errors are clustered at the country-year level in Panel A, while robust standard errors are reported in Panel B.
*** p<0.01, ** p<0.05, * p<0.1.

Figure 4: Employment share by sector and initial union density



Note: The figure displays changes in employment shares across broad sectors from 2002 to 2018, expressed relative to their levels in 2002. Employment shares are calculated by aggregating Eurostat employment data across the fifteen countries included in the study, grouped into broad sectors: Primary Sector (NACE Rev 1.1: sections A, B, C); Manufacturing (D); Utilities and Construction (E, F); Trade and Retail (G); Hotels and Restaurants (H); Transport and Communication (I); Finance and Business (J, K); Public Sector (L, M, N, Q); and Personal and Community Services (O, P). Each country's contribution is weighted equally, so the series reflects the unweighted average sectoral composition across countries. Darker and wider lines indicate sectors with higher initial union density in 2002.

to examine the effect of personal exposure to automation on the likelihood of being a union member and on a set of associated individual outcomes.

In conducting individual-level analyses, it is crucial to use measures that extend beyond the automation exposure associated with a person's current occupation. This is because an individual's current occupation may already be influenced by prior automation-driven changes, either through direct or indirect displacement. For example, some workers may have transitioned to low-automation service occupations after being displaced by robots from previous manufacturing jobs. Similarly, new labor market entrants may find themselves in less secure and lower-paying occupations due to the shrinking availability of "good jobs" in manufacturing, a consequence of automation. These individuals, based on their current occupations, might appear to have low automation exposure, yet they represent classic cases of automation-induced displacement. Moreover, unemployed individuals, who may have been displaced due to automation, lack current occupational data altogether, further complicating the assessment.

To address these issues, we employ a counterfactual measure of individual vulnerability to automation that accounts for potential displacement. This approach is particularly relevant for new labor market entrants who may face unemployment or find themselves in suboptimal, low-automation jobs. We adopt the individual automatability measure developed by Anelli et al. (2021).

This methodology utilizes historical data on labor market conditions from the early 1990s, prior to the most recent wave of automation. Instead of relying on observed current occupations, we estimate predicted probabilities of employment in various occupations based on factors such as education, age, gender, and region of residence. These estimates are derived from the early 1990s European Labor Force Survey (EU-LFS) data, calculated separately for each country. We then apply these estimates to respondents from the European Social Survey (ESS), who were interviewed between 2000 and 2018. By combining these predicted probabilities with the automatability measure for each occupation, we generate a metric of "individual vulnerability to automation." This metric is then multiplied by the pace of robot adoption in each country and year, yielding a measure of "individual exposure to automation."

Our approach effectively assigns higher automation exposure to individuals who, based on pre-automation labor market conditions, would have been more likely to be employed in automatable occupations. This method captures the reality that, due to automation, certain job opportunities have diminished, potentially resulting in unemployment or employment in less desirable occupations.

Following Anelli et al. (2021), we define individual exposure to automation as:

$$\text{Individual Exposure}_{icrt} = \Delta R_{ct} * \underbrace{\sum_j \widehat{Pr}(o_i = j|\text{age}, \text{gender}, \text{edu}, r) * \theta_j}_{\text{Individual Vulnerability}}$$

where: $\Delta R_{crt} = \sum_j \frac{L_{crj}^{\text{pre-sample}}}{L_{cr}^{\text{pre-sample}}} * \frac{R_{cj}^{t-1} - R_{cj}^{t-n}}{L_{cj}^{\text{pre-sample}}}$ represents region-level exposure to robot adoption over the past n years. The term $\widehat{Pr}(o_i = j|\text{age}, \text{gender}, \text{edu}, r)$ denotes individual i's predicted probability of working in occupation j, estimated using pre-shock data from the European Labor Force Survey (EU-LFS). The parameter θ_j captures the automation threat associated to occupation j, based on the scores provided by Frey and Osborne (2017). We calculate these individual exposure scores for respondents of the European Social Survey (ESS) interviewed between 2000 and 2018 across the 15 Western European countries included in our study.

We now proceed to estimate the effect of individual exposure to automation on likelihood of being a union member. Specifically, we estimate the following equation:

$$\text{Union}_{it} = \alpha + \beta_1 \text{Individual Exposure}_{icrt} + \beta_2 X_{it} + \lambda_{ct} + \lambda_r + \varepsilon_{icrt}$$

Table 7: Individual level analysis

VARIABLES	(1) Union (best)	(2) Union (best)
Individual Exposure	-0.011***	-0.074***
	[0.004]	[0.013]
Estimator	OLS	2SLS
Instrument		Technological
Observations	224,922	224,922
Age, gender, edu	X	X
Country-year FE	X	X
Region FE	X	X
Std dev. Y	0.375	0.375
Std dev. X	0.421	0.421
Magnitude	-0.0121	-0.0831
R-squared	0.166	
Kleibergen-Paap F-stat		54.04

Standard errors clustered by region-year
*** p<0.01, ** p<0.05, * p<0.1

where i indexes individuals, c countries, r regions, and t years. Individual Exposure$_{icrt}$ represents our measure of individual exposure to automation. The percentage change in total operational robots is between year t – 1 and t – 3, in the country c where individual i resides. X_{it} is a vector of individual control variables, including age, gender, and years of education. Standard errors are clustered by region-year. As in our previous regression analysis, the change in the stock of robots is instrumented using our three indexes of technological progress.

Table 7 presents the results of the regression analysis. The coefficient on individual exposure to automation is both negative and statistically significant in the OLS and 2SLS estimations. In the IV model, a one standard deviation increase in individual exposure to automation is associated with a 7.4 percentage point reduction in the probability of being a union member. After accounting for fixed effects and control variables, this reduction corresponds to approximately 8 percentage points.[14] These results suggest that robotization significantly contributes to lower unionization rates among individuals more vulnerable to automation.

It is important to emphasize that our measure of exposure is counterfactual, indicating a decreased propensity to unionize among those who would have been more likely to work in automatable occupations. This finding reinforces the expectation that, due to automation and the resulting structural changes in the labor market, many workers have been displaced into employment contexts with lower union density.

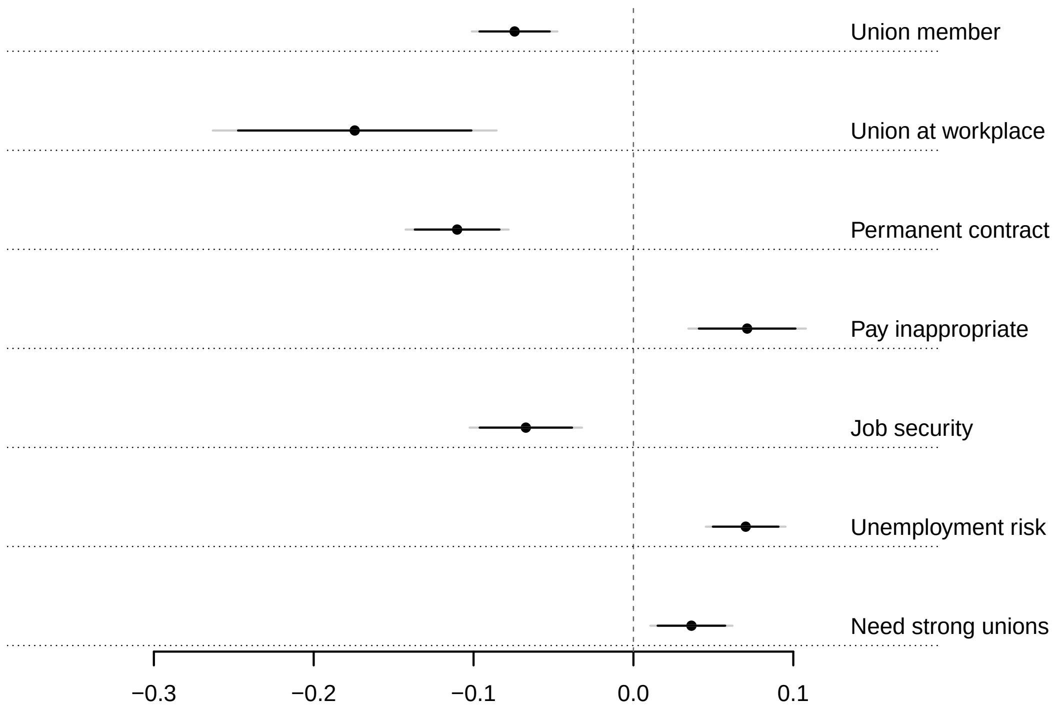
To gain a deeper understanding of our findings, we re-estimate the model using a different set of dependent variables designed to capture economic conditions and perceptions. These variables are constructed based on specific items from the European Social Survey (ESS) and include whether a trade union is present at the respondent's workplace, whether the respondent has a permanent contract, their perception of being fairly compensated relative to their efforts and achievements, their sense of job security, their likelihood of becoming unemployed within the next 12 months, and their agreement with the statement that strong trade unions are necessary to protect working conditions and wages.

The results, presented in Figure 5, reveal several key findings. In addition to the previously discussed negative effect on individual union membership, the analysis suggests that individuals with higher exposure to automation are less likely to work in a unionized workplace, less likely to hold a permanent contract, more likely to perceive their pay as inadequate, and more likely to be in insecure jobs with a higher risk of unemployment. Additionally, these individuals are more inclined to believe in the importance of strong unions to safeguard working conditions and wages.

In summary, our analysis shows that industrial automation has led to a decline in

[14]The magnitude of the effect of individual exposure would be similar, though slightly smaller—around 7 percentage points in a simple OLS model—if, instead of our counterfactual measure of individual vulnerability interacted by the regional robot shock, we used the exposure threat index θ_j alone, based on the respondent's observed occupation.

Figure 5: Effect of individual exposure to robots on economic conditions and perceptions



unionization, largely by shifting employment from highly unionized sectors to those with lower union presence. The individual-level analysis further supports this, indicating that workers more vulnerable to automation are less likely to be union members and experience less secure working conditions. These findings suggest that robotization is reshaping labor markets, contributing to a reduced overall union presence.

6 Conclusion

Technological change generates economic distributional consequences that have been shown to be politically consequential. Labor unions are potentially key actors shaping both the economic and the political implications of technological change. Yet the impact of technological change on unionization has remained largely underexplored. This study offers new insight on this issue.

We assemble a novel dataset with unionization data at the region and industry level, encompassing fifteen western European countries over two decades (2002-2018). These data allow us to document for the first time the decline of unionization in Europe at the granular level. We also shed empirical light on the long-assumed positive association between unionization and support for mainstream left parties at the local level.

We provide comprehensive evidence that technological change, in the form of robotization of manufacturing activities, has been a significant driver of unionization decline. Specifically, we find that an increase in exposure to robotization decreases the unionization rate at the regional level. The decline in unionization is driven by a compositional effect in the labor market: automation induces a reallocation of employment from traditionally unionized industries, where robotization is more intensive, towards less unionized ones. Conversely, there is no evidence of a systematic reduction of unionization within industries that are relatively more exposed to automation.

These findings speak to the role of unionization as a contextual factor that may shape electoral dynamics at the local level, influencing the political repercussions of automation. For instance, as proposed by Kitschelt (2012), the decline of labor unions may be a key factor behind the rise of radical right support among the losers of structural change in advanced democracies. In fact, labor unions are intermediary organizations that historically connected blue-collar constituencies to mainstream parties, in particular of the left. These organizations not only linked workers to parties but also framed their interests in terms of class antagonism rather than ethno-national identities. As automation displaces workers and weakens trade unions, the capacity of unions to mobilize working-class voters and serve as political intermediaries for social-democratic parties diminishes. Economic grievances among the losers of structural change are then increasingly intercepted by nationalist and radical-right forces.

Our results also have important implications for trade unions and the future of labor politics. We have documented how the ongoing broad reconfiguration of the labor market can lower the overall share of unionized workers. One fundamental driver of union decline—though not exclusive—is the reduction in manufacturing employment caused by automation, which reduces the number of workers in industries historically characterized by strong union presence. Given the crucial role of unions in shaping distributive outcomes and supporting democratic functions such as voter turnout and civic engagement, their decline raises important questions about their future adaptability. A key challenge lies in organizing workers in sectors outside traditional union strongholds while countering ongoing efforts to limit unionization. So far, unions have largely endured by maintaining a strong foothold in the public sector. However, their ability to mobilize collective action in nontraditional sectors will be crucial to their future relevance.

The novel dataset developed in this study will enable further research on these topics, and more broadly on the social, and political implications of labor unions, whose significance is difficult to overstate.

References

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: implications for employment and earnings. In *Handbook of Labor Economics*, Volume 4, pp. 1043–1171. Elsevier.

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: evidence from US labor markets. *Journal of Political Economy 128*(6), 2188–2244.

Agnolin, P. (2025). The candidate factory: technological change and political supply. *Unpublished manuscript*.

Ahlquist, J. S. (2017). Labor unions, political representation, and economic inequality. *Annual Review of Political Science 20*, 409–432.

Ahlquist, J. S. and M. Downey (2023, November). The effects of import competition on unionization. *American Economic Journal: Economic Policy 15*(4), 359–89.

Ahlquist, J. S. and M. Levi (2013). *In the interest of others: organizations and social activism*. Princeton University Press.

Anelli, M., I. Colantone, and P. Stanig (2019). We were the robots: automation and voting behavior in western Europe. *BAFFI CAREFIN Centre Research Paper* (2019-115).

Anelli, M., I. Colantone, and P. Stanig (2021). Individual vulnerability to industrial robot adoption increases support for the radical right. *Proceedings of the National Academy of Sciences 118*(47).

Arndt, C. and L. Rennwald (2017). Workplace characteristics and working class vote for the old and new right. *British Journal of Industrial Relations 55*(1), 137–164.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives 29*(3), 3–30.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review 103*(5), 1553–1597.

Autor, D. H., D. Dorn, and G. H. Hanson (2013, October). The China syndrome: local labor market effects of import competition in the United States. *American Economic Review 103*(6), 2121–68.

Balcazar, C. F. (2024). Globalization, unions and robots: the effects of automation on the power of labor and policymaking. *Available at SSRN: https://ssrn.com/abstract=4574527*.

Becher, M. and D. Stegmueller (2019). Cognitive ability, union membership, and voter turnout. *IAST Working Paper* (19-97).

Becher, M. and D. Stegmueller (2020). Reducing unequal representation: the impact of labor unions on legislative responsiveness in the US Congress. *Perspectives on Politics 19*(1), 92–109.

Becher, M. and D. Stegmueller (2024). Global competition, local unions, and political representation: disentangling mechanisms. *American Journal of Political Science*.

Becher, M., D. Stegmueller, and K. Käppner (2018). Local union organization and law making in the US Congress. *The Journal of Politics 80*(2), 539–554.

Bessen, J., M. Goos, A. Salomons, and W. Van den Berge (2023). Automatic reaction - what happens to workers at firms that automate? *The Review of Economics and Statistics* (Feb. 6, 2023).

Betz, H.-G. (1993). The new politics of resentment: radical right-wing populist parties in western Europe. *Comparative Politics*, 413–427.

Betz, H.-G. (1994). *Radical right-wing populism in western Europe*. Palgrave Macmillan.

Betz, H.-G. and S. Meret (2012). Right-wing populist parties and the working-class vote: what have you done for us lately? In *Class Politics and the Radical Right*, pp. 107–121. Routledge.

Boix, C. (2019). Democratic capitalism at the crossroads: technological change and the future of politics. *Princeton University Press*.

Borjas, G. J. and R. B. Freeman (2019). From immigrants to robots: the changing locus of substitutes for workers. *RSF: The Russell Sage Foundation Journal of the Social Sciences 5*(5), 22–42.

Brady, D. (2007). Institutional, economic, or solidaristic? Assessing explanations for unionization across affluent democracies. *Work and Occupations 34*(1), 67–101.

Carbonero, F., E. Ernst, and E. Weber (2020). Robots worldwide: the impact of automation on employment and trade.

Carnes, N. (2013). *White-collar government: the hidden role of class in economic policy making*. Chicago, IL: University of Chicago Press.

Caselli, M., A. Fracasso, and S. Traverso (2021). Globalization, robotization, and electoral outcomes: evidence from spatial regressions for Italy. *Journal of Regional Science 61*(1), 86–111.

Chang, T. F. (2001). The labour vote in US national elections, 1948–2000. *Political Quarterly 72*(3).

Checchi, D. and C. Lucifora (2002). Unions and labour market institutions in Europe. *Economic Policy 17*(35), 361–408.

Checchi, D., J. Visser, and H. G. Van De Werfhorst (2010). Inequality and union membership: the influence of relative earnings and inequality attitudes. *British Journal of Industrial Relations 48*(1), 84–108.

Chen, W. and M. Nabar (2018). The employment impact of automation following the global financial crisis: the case of industrial robots. In IMF (Ed.), *World Economic Outlook: Challenges to steady growth*, Chapter ch. 2, pp. 90–92. Washington DC: International Monetary Fund.

Chiacchio, F., G. Petropoulos, and D. Pichler (2018). The impact of industrial robots on EU employment and wages: a local labour market approach. *Bruegel Working Paper 2018*(04).

Colantone, I. and P. Stanig (2018). The trade origins of economic nationalism: import competition and voting behavior in western Europe. *American Journal of Political Science 62*(4), 936–953.

Dal Bó, E., F. Finan, O. Folke, T. Persson, and J. Rickne (2023). Economic and social outsiders but political insiders: Sweden's populist radical right. *The Review of Economic Studies 90*(2), 675–706.

Dark, T. E. (2018). *The unions and the Democrats: an enduring alliance*. Cornell University Press.

Dauth, W., S. Findeisen, J. Suedekum, and N. Woessner (2021). The adjustment of labor markets to robots. *Journal of the European Economic Association 19*(6), 3104–3153.

Dottori, D. (2021). Robots and employment: evidence from Italy. *Economia Politica 38*(2), 739–795.

Ebbinghaus, B. and J. Visser (1999). When institutions matter: union growth and decline in western Europe, 1950–1995. *European Sociological Review 15*(2), 135–158.

Enggist, M. and M. Pinggera (2022). Radical right parties and their welfare state stances–not so blurry after all? *West European Politics 45*(1), 102–128.

Esping-Andersen, G. (1985). Power and distributional regimes. *Politics & Society 14*(2), 223–256.

Eurostat (2008). *NACE Rev. 2: Statistical Classification of Economic Activities*. Methodologies and Working Papers. Luxembourg: Eurostat, European Commission. Product code: KS-RA-07-015.

Farber, H. S. (2005). Union membership in the United States: the divergence between the public and private sectors. *Princeton University Industrial Relations Section Working Paper*.

Farber, H. S., D. Herbst, I. Kuziemko, and S. Naidu (2021). Unions and inequality over the twentieth century: new evidence from survey data. *The Quarterly Journal of Economics 136*(3), 1325–1385.

Feigenbaum, J., A. Hertel-Fernandez, and V. Williamson (2018). From the bargaining table to the ballot box: political effects of right to work laws. Technical report, National Bureau of Economic Research.

Freeman, R. B. and J. L. Medoff (1984). What do unions do? *Basic Book*.

Frey, C. B., T. Berger, and C. Chen (2018). Political machinery: did robots swing the 2016 US presidential election? *Oxford Review of Economic Policy 34*(3), 418–442.

Frey, C. B. and M. A. Osborne (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change 114*, 254–280.

Frymer, P. and J. M. Grumbach (2021). Labor unions and white racial politics. *American Journal of Political Science 65*(1), 225–240.

Gallego, A. and T. Kurer (2022). Automation, digitalization, and artificial intelligence in the workplace: implications for political behavior. *Annual Review of Political Science 25*(1), 463–484.

Gelman, A., J. Lax, J. Phillips, J. Gabry, and R. Trangucci (2019). Using multilevel regression and poststratification to estimate dynamic public opinion. *Unpublished manuscript*.

Gelman, A. and T. C. Little (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology 23*(2), 127–135.

Gingrich, J. (2019). Did state responses to automation matter for voters? *Research & Politics 6*(1), 2053168019832745.

Goldin, C. and L. F. Katz (1998). The origins of technology-skill complementarity. *The Quarterly Journal of Economics 113*(3), 693–732.

Goos, M., A. Manning, and A. Salomons (2014). Explaining job polarization: routine-biased technological change and offshoring. *American Economic Review 104*(8), 2509–2526.

Graetz, G. and G. Michaels (2018). Robots at work. *The Review of Economics and Statistics 100*(5), 753–768.

Greilinger, G. and C. Mudde (2024). Talking left, voting right: an assessment of far-right voting on socio-economic issues in the european parliament. Policy Brief 2024.05, ETUI.

Hacker, J. S. and P. Pierson (2010). *Winner-take-all politics: how Washington made the rich richer–and turned its back on the middle class*. Simon and Schuster.

Hechter, M. (2004). From class to culture. *American Journal of Sociology 110*(2), 400–445.

Häusermann, S. and H. Kitschelt (Eds.) (2023). *Beyond social democracy: the transformation of the left in emerging knowledge societies*. Cambridge, United Kingdom: Cambridge University Press Assessment.

Im, Z. J., N. Mayer, B. Palier, and J. Rovny (2019). The losers of automation: a reservoir of votes for the radical right? *Research & Politics 6*(1), 2053168018822395.

Iversen, T. and D. Soskice (2015). Information, inequality, and mass polarization: ideology in advanced democracies. *Comparative Political Studies 48*(13), 1781–1813.

Kim, S. E. and Y. Margalit (2017). Informed preferences? The impact of unions on workers' policy views. *American Journal of Political Science 61*(3), 728–743.

Kitschelt, H. (2012). Social class and the radical right: conceptualizing political preference formation and partisan choice. *J. Rydgren (ed.), Class Politics and the Radical Right*, 224–251.

Kitschelt, H. and A. J. McGann (1997). *The radical right in western Europe: a comparative analysis*. University of Michigan Press.

Korpi, W. (1983). *The democratic class struggle*. Boston: Routledge.

Korpi, W. (1986). *Power resources approach vs. action and conflict: on causal and intentional explanation in the study of power*. Swedish Institute for Social Research.

Kristal, T. (2013). The capitalist machine: computerization, workers' power, and the decline in labor's share within U.S. industries. *American Sociological Review 78*(3), 361–389.

Kristal, T. and Y. Cohen (2015). What do computers really do? Computerization, fading pay-setting institutions and rising wage inequality. *Research in Social Stratification and Mobility 42*, 33–47.

Lax, J. R. and J. H. Phillips (2009). Gay rights in the states: public opinion and policy responsiveness. *American Political Science Review 103*(3), 367–386.

Leemann, L. and F. Wasserfallen (2017). Extending the use and prediction precision of subnational public opinion estimation. *American Journal of Political Science 61*(4), 1003–1022.

Leemann, L. and F. Wasserfallen (2020). Measuring attitudes–multilevel modeling with post-stratification (mrp). In *The Sage Handbook of Research Methods in Political Science and International Relations*, pp. 371–384. SAGE Publications Ltd.

Leighley, J. E. and J. Nagler (2007). Unions, voter turnout, and class bias in the U.S. electorate, 1964–2004. *The Journal of Politics 69*(2), 430–441.

Levi, M. (2017). Labor unions as critical intermediate associations. *FSI Memo 2017*(3).

Lichtenstein, N. (2013). State of the union: a century of American labor-revised and expanded edition. In *State of the Union*. Princeton University Press.

Lipset, S. M. (1983). Radicalism or reformism: the sources of working-class politics. *American Political Science Review 77*(1), 1–18.

Lopez, S. H. (2004). *Reorganizing the rust belt: an inside study of the American labor movement*. Univ of California Press.

Lyon, G. and B. F. Schaffner (2021). Labor unions and non-member political protest mobilization in the United States. *Political Research Quarterly 74*(4), 998–1008.

Mach, A., L. Rennwald, and A. Zimmermann (2024, July). The role of trade union officials in the political representation of the working class in Switzerland (1880-2020). In *2024 European Political Science Association Conference*, Cologne, Germany. 4-6 July 2024.

Marks, G. (1982). *Trade unions in politics*. Ph. D. thesis, Ph. D. dissertation, Department of Political Science, Stanford University.

Meyer, B. (2019). Financialization, technological change, and trade union decline. *Socio-Economic Review 17*(3), 477–502.

Milner, H. V. (2021). Voting for populism in Europe: globalization, technological change, and the extreme right. *Comparative Political Studies 54*(13), 2286–2320.

Mummolo, J. and E. Peterson (2018). Improving the interpretation of fixed effects regression results. *Political Science Research and Methods 6*(4), 829–835.

Müller, C. (2022). *World Robotics 2022 – Industrial Robots*. Frankfurt am Main, Germany: IFR Statistical Department, VDMA Services GmbH.

Oesch, D. (2006). *Redrawing the class map: stratification and institutions in Britain, Germany, Sweden and Switzerland*. Palgrave Macmillan.

Park, D. K., A. Gelman, and J. Bafumi (2004). Bayesian multilevel estimation with post-stratification: state-level estimates from national polls. *Political Analysis 12*(4), 375–385.

Perani, G., V. Cirillo, et al. (2015). Matching industry classifications. A method for converting Nace rev. 2 to Nace rev. 1. *WP-EMS Working Papers Series in Economics, Mathematics and Statistics*.

Pontusson, J. (1995). Explaining the decline of European social democracy: the role of structural economic change. *World Politics 47*(4), 495–533.

Pontusson, J. and D. Rueda (2010). The politics of inequality: voter mobilization and left parties in advanced industrial states. *Comparative Political Studies 43*(6), 675–705.

Przeworski, A. and J. Sprague (1986). *Paper Stones: a History of Electoral Socialism*. Chicago: University of Chicago Press.

Radcliff, B. and P. Davis (2000). Labor organization and electoral participation in industrial democracies. *American Journal of Political Science 44*(1), 132–141.

Rathgeb, P. and M. R. Busemeyer (2022). How to study the populist radical right and the welfare state? *West European Politics 45*(1), 1–23.

Rennwald, L. and N. Mosimann (2023, February). Union membership and electoral demand for redistribution among left-authoritarians. Working Paper 37, FORS, Lausanne, Switzerland.

Rennwald, L. and J. Pontusson (2021). Paper stones revisited: class voting, unionization and the electoral decline of the mainstream left. *Perspectives on Politics 19*(1), 36–54.

Rosenfeld, J. (2014). *What unions no longer do*. Harvard University Press.

Rovny, J. (2013). Where do radical right parties stand? Position blurring in multidimensional competition. *European Political Science Review 5*(1), 1–26.

Rupp, K. (2021). Data from "42 years of microprocessor trend data". GitHub repository. Accessed: 1 December 2023.

Schlozman, D. (2015). *When movements anchor parties: electoral alignments in American history*. Princeton University Press.

Slaughter, M. J. (2007). Globalization and declining unionization in the United States. *Industrial Relations: A Journal of Economy and Society 46*(2), 329–346.

US Bureau of Labor Statistics (2023). Producer price index by industry: computer and electronic product manufacturing. Retrieved from FRED, Federal Reserve Bank of St. Louis. Accessed 1 December 2023.

Visser, J. (2019). Ictwss: Database on institutional characteristics of trade unions, wage setting, state intervention and social pacts. Dataset. Available at: https://b2share.eudat.eu.

Volkens, A., P. Lehmann, T. Matthieß, N. Merz, S. Regel, and A. Werner (2016). The manifesto data collection. Manifesto project (mrg/cmp/marpor). Version 2016a. *Wissenschaftszentrum Berlin für Sozialforschung (WZB)*.

Wallerstein, M. and B. Western (2000). Unions in decline? What has changed and why. *Annual Review of Political Science 3*(1), 355–377.

Western, B. (1997). *Between class and market: postwar unionization in the capitalist democracies*. Princeton University Press.

Western, B. and J. Rosenfeld (2011). Unions, norms, and the rise in US wage inequality. *American Sociological Review 76*(4), 513–537.

Robots Replacing Trade Unions: Novel Data and Evidence from Western Europe

Online Appendix

Contents

A MRP prediction models

At the core of estimations based on Multilevel Regression with Post-stratification (MRP) is a multilevel (or hierarchical) response model in which the outcome is the variable of interest (Leemann and Wasserfallen, 2020). In our case, the outcome variable indicates whether the respondent is currently a member of a trade union and is a function of individual-level characteristics such as gender, age, educational attainment, occupation, region of residence and industry, modeled as random intercepts, so that the intercept for each category of a given variable is modeled as a draw from a common normal distribution. A full list of the categories used for each variable is provided in Section B. The model yields estimates for the realization of the random effects, which allows to predict the probability of trade union membership for every possible combination of the included covariate categories.

In addition to these individual-level effects, we account for time trends, following the dynamic MRP approach developed by Gelman et al. (2019). Specifically, we include a term for the survey round as a linear trend ($\beta \cdot round$). To capture heterogeneous time trends, we also allow the time effect to vary across groups defined by a given individual-level characteristic. For example, the effect of time may differ by age group ($\gamma^{age} \cdot round$): these age-specific time trends are modeled as random slopes.

Industries (*ind*) and occupation (*occ*) variables are defined at the 2-digit level of the NACE Rev. 1.1 and ISCO-88 classifications, respectively. When allowing for specific time trends, we employ the more aggregated NACE sub-section (*ind_sector*), as well as ISCO-88 1-digit (occ_1d) categories. The same is done in models that include interaction effects for combinations of variables. For instance, we include random intercepts for specific occupation–sector combinations ($\alpha^{occ\_1d,\ ind\_sector}$), allowing for more flexible patterns of variation across individual profiles.

Finally, we model contextual (regional-level) effects by including both random intercepts for regions and linear effects for regional characteristics. That is, alongside the random intercept for each region, α^{region}, we include fixed coefficients $\boldsymbol{\xi}$ for region-level predictors $\boldsymbol{X_r}$. These pre-sample predictors include: the employment share of low- and medium-skill workers, the employment share of services, the employment share of low- and medium-tech industries, the employment share of primary sector, the employment share of finance and business services, and the share of foreign-born workers.

This structure allows us to account for both individual-level and contextual heterogeneity in trade union membership, while modeling its evolution over time. We employ the following list of sixteen alternative specifications:

1. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \gamma^{ind\_sector} \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$
2. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$
3. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \gamma^{occ\_1d} \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

4. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \gamma^{age} \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

5. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \gamma^{edu} \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

6. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{edu} + \alpha^{age} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \gamma^{gndr} \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

7. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{occ\_1d,\ ind\_sector} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

8. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{occ\_1d,\ reg} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

9. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{occ\_1d,\ age} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

10. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{occ\_1d,\ gndr} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

11. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{occ\_1d,\ edu} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

12. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{ind\_sector,\ reg} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

13. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{ind\_sector,\ age} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

14. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{ind\_sector,\ gndr} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

15. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \beta \cdot round + \alpha^{ind\_sector,\ edu} + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

16. $Pr(\text{Union}_i) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{ind} + \alpha^{occ} + \alpha^{region} + \alpha^{round} + \beta \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$

Specification 1 reflects our preferred approach on conceptual grounds, as it allows for sector-specific differential time trends. Specification 2 provides a more parsimonious alternative, relying only on a common time trend. Specifications 3 through 6 replicate the structure of specification 1 incorporating differential time trends for individual characteristics (occupation, age, education, gender). Specifications 7 through 15 include interaction effects for pairwise combinations of occupation (1-digit) or sector and each of the other individual-level variables. These models allow the joint effect of two variables to vary flexibly across grouping levels by estimating separate random intercepts for each combination. For instance, specification 15 includes random intercepts for sector-education, capturing variation across the 27 industry sub-sections (plus a category for the unemployed) crossed with three levels of educational attainment. Specification 16 adds a random intercept for survey round (α^{round}) alongside a linear time trend ($\beta \cdot round$). While we generally avoid modeling round-specific effects due to the risk of overfitting wave-specific idiosyncrasies, this specification is included to assess MRP performance under an alternative strategy for modeling temporal heterogeneity.

A.1 RMSE and assessment of prediction models

To evaluate the relative predictive performance of each model specification, we compute, for each model, a group-wise calibration Root Mean Square Error (RMSE). In this section, we first outline how this metric is constructed, then describe its distribution across models, and finally explain the procedure used to rank models based on their RMSE scores.

First, we implement a 10-fold cross-validation procedure (within each sample country). We randomly split the sample into $K = 10$ folds, then iteratively estimate the model on the training set (excluding fold k), and generate predicted probabilities of union membership for the held-out fold k. This process is repeated across all folds, yielding a complete vector of out-of-sample predictions.

We assess calibration at relevant levels of aggregation, i.e., regions and industries within each country, summarized by computing a group-level root mean squared (RMSE) metric. This metric summarizes departures of the observed frequencies by region from the average of the predicted probabilities by region; and the same for industries. A well-calibrated model will display empirical frequencies by group in line with the predicted probabilities. The metric we adopt takes the squared discrepancies between frequencies and predicted probabilities in each group, averages them, and then moves back to the original scale by taking the square root.

Formally, let Ω_g denote the set of N_g observations in a given group $g = 1, 2, \ldots, G$. A group is defined either as a region or an industry within each country. We first calculate the average predicted probability of union membership $\hat{P}_g$ for group g as:

$$\hat{P}_g = \frac{\sum_{i \in \Omega_g} \hat{P}_i}{N_g} \tag{10}$$

The observed empirical frequency F_g in the ESS data is given by:

$$F_g = \frac{\sum_{i \in \Omega_g} \mathbb{1}(Union_i = 1)}{N_g} \tag{11}$$

The RMSE for a given level of aggregation (region or industry) within a given country is then defined as:

$$\text{RMSE}_\text{G} = \left(\frac{\sum_g (\hat{P}_g - F_g)^2}{G} \right)^{\frac{1}{2}} \tag{12}$$

This metric compares the (cross-validated) predicted probabilities for each region or industry g with the empirical frequency of unionization among survey respondents from region or industry g.

To facilitate meaningful comparisons across countries with different levels of unionization, we rescale the RMSE by the observed range of unionization rates across groups. That is, we divide the raw RMSE by the range of the observed group-level unionization rates:

$$\text{Rescaled RMSE}_G = \frac{\text{RMSE}_G}{\max_g F_g - \min_g F_g} \tag{13}$$

Finally, because our interest lies in assessing model calibration across both geographic and industry dimensions, we create an overall metric as the sum of the rescaled RMSEs for regions and industries:

$$\text{RMSE} = \text{Rescaled RMSE}_{region} + \text{Rescaled RMSE}_{industry} \tag{14}$$

This measure captures how well each model reproduces the observed spatial and industry variation in unionization in the survey data, and allows us to compare the predictive accuracy of alternative specifications.

Figure A1 displays the relative performance of each model specification by country, based on RMSE. To facilitate comparison, RMSE values are expressed as a percentage of the highest (i.e., worst) RMSE within each country. Since RMSE values are close in absolute terms, this approach highlights relative differences in model performance. For example, in Austria, specification 8 yields the lowest RMSE, which is approximately 12% lower than the highest RMSE for the country, observed in specification 16.

We then rank the different model specifications by their RMSE performance within each country. Specifically, we assign a rank to each model in every country based on its RMSE value. For example, model 1 is the top-performing model in France, the Netherlands, Sweden, and the UK, while it ranks 15th in Portugal and Spain. To compare overall performance across models, we consider the average ranking by country. That is, we prioritize models that, even if not consistently the best in every country, tend to perform well relative to others on average.

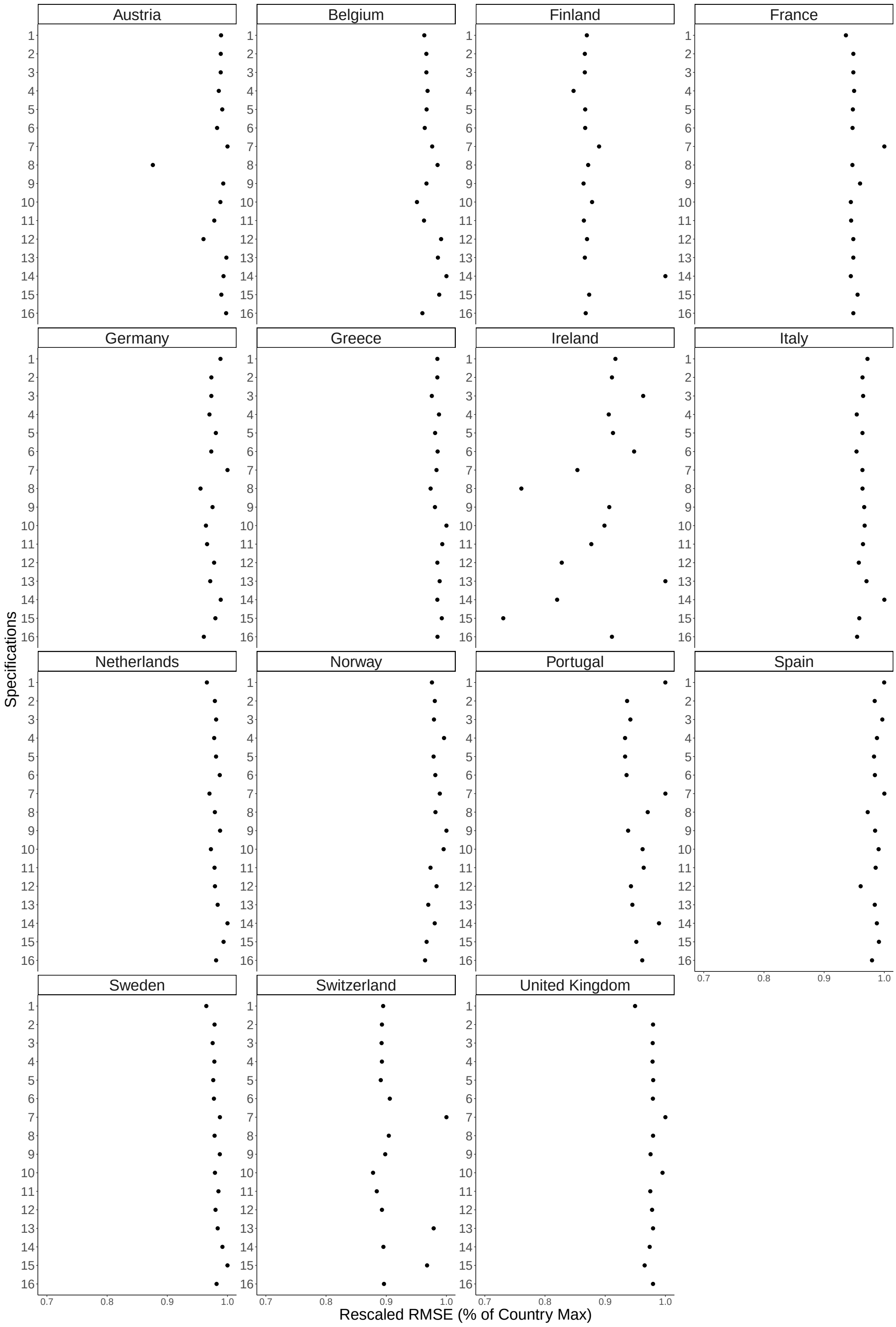
According to this criterion, model 11 emerges as the best-performing model overall. While this is not the top-ranked model in any single country, it consistently achieves high rankings across most countries. We therefore adopt specification 11 as the baseline model. Its functional form is:

$$\Pr(Union_i = 1) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{occ} + \alpha^{ind} + \alpha^{region} + \alpha^{occ\_1d,\ edu} + \beta \cdot round + \xi \cdot \boldsymbol{X_r}) \tag{15}$$

where the α terms are random effects for gender, age category, education level, occupation, industry of employment, region of residence, and the combination of (1-digit) occupation and education level. The specification includes also a linear time trend, captured by the ESS *round* variable, and the vector $\boldsymbol{X_r}$ of variables controlling for cross-regional differences in pre-sample conditions.

Our main results on the impact of robotization on unionization are robust to employing union density figures obtained from any of the sixteen different probit specifications. Throughout the paper, for ease of exposition we only show results based on two alternative approaches to model selection. The first is to use the best-performing model within each country. This leads us to select different models for different countries. Under this approach,

Figure A1: Prediction models performance by country



Note: The figure plots RMSE values rescaled within each country, expressed as a percentage of the highest RMSE observed in that country. For each country, the model with the highest RMSE is set to 1, and all other values are expressed relative to it.

we would choose specification 1 for France, the Netherlands, Sweden, and the UK; specification 4 for Finland and Portugal; specification 6 for Italy; specification 8 for Austria, Germany, and Greece; specification 10 for Belgium and Switzerland; specification 12 for Spain; specification 15 for Ireland; and specification 16 for Norway.

The second alternative approach relies on conceptual considerations and adopts specification 1, which incorporates sector-specific differential time trends. The model is defined as follows.

$$\Pr(\text{Union}_i = 1) = \text{Probit}(\alpha^{gndr} + \alpha^{age} + \alpha^{edu} + \alpha^{occ} + \alpha^{ind} + \alpha^{region} + \beta \cdot round + \gamma^{ind\_sector} \cdot round + \boldsymbol{\xi} \cdot \boldsymbol{X_r})$$

were $\gamma^{ind\_sector} \cdot round$ denotes time trends that are specific to NACE sub-sections, i.e., aggregations of 2-digit industries.

B Data sources and harmonization of data across countries

In Table B.1, we report the data sources used for each country in both the prediction and post-stratification stages of the MRP estimation procedure. While the prediction stage relies exclusively on the European Social Survey (ESS), post-stratification draws on national census or registry data wherever available. In countries where recent census data are unavailable or access is highly restricted, we instead rely on Labor Force Surveys. The last column shows the external validation data, available only for Finland, Norway, and the UK.

Table B.1: Data sources

Country	Prediction data	Post-stratification data	External validation data available
Austria	ESS	Mikrozensus (Statistics Austria)	-
Belgium	ESS	National LFS (Statbel)	-
Finland	ESS	Registry Data (Statistics Finland)	Registry Data (Statistics Finland)
France	ESS	Census (IPUMS)	-
Germany	ESS	Mikrozensus (Statistisches Bundesamt)	-
Greece	ESS	Census (IPUMS)	-
Ireland	ESS	Census (IPUMS)	-
Italy	ESS	National LFS (ISTAT)	-
The Netherlands	ESS	Survey Workforce New Series EEBnw (CBS)	-
Norway	ESS	Registry Data (Statistics Norway)	Registry Data (Statistics Norway)
Spain	ESS	Census (IPUMS)	-
Sweden	ESS	Registry Data (Statistics Sweden)	-
Switzerland	ESS	EU-LFS (FSO - Eurostat) + Admnistrative data	-
Portugal	ESS	Census (IPUMS)	-
United Kingdom	ESS	Office for national statistics (National LFS)	National LFS

To ensure consistent application of the MRP estimation across space and time, for each country we harmonized six key variables present in both the ESS and the post-stratification datasets. These harmonized variables are listed, along with their categories, in Table B.2 and include age, education, region, industry, occupation, and gender.

For the regional variable, we adopt the NUTS 2-digit classification for all countries except the UK and Germany, for which the data we use are available only at the NUTS 1-digit level. The dataset excludes Northern Ireland, Ceuta and Melilla, and the French overseas territories.

The education variable is categorized into three levels: “less than secondary education”, “secondary education”, and “tertiary education or more”. Occupation is classified using the ISCO-88 scheme at the 2-digit level. When original data use different classification systems (e.g. ISCO-08, SOC, or national classifications), we apply crosswalks to convert them to ISCO-88.

Table B.2: Summary of harmonized variables and categories

Variable	Categories
Gender	Male, Female
Education	Less than secondary education, Secondary education, Tertiary education or more
Age	Below 25, 25-34, 35-44, 45-54, 55-64, Over 65
Region	NUTS 2-digit (NUTS 1-digit UK, DE)
Occupation	ISCO-88 2-digit
Industry	NACE Rev. 1.1. 2-digit

The industry variable follows the NACE Rev. 1.1 classification, at the 2-digit level. When original industry data are based on other systems (mainly NACE Rev. 2, and in some cases SIC or national classifications), we convert them using crosswalks. The mapping from NACE Rev. 2 to NACE Rev. 1.1 is detailed in the next section. Unemployed individuals are classified under a residual, fictitious industry category to reflect their temporary lack of affiliation with any specific industry.

Throughout the data harmonization process, for both the ESS and post-stratification datasets, we identify and exclude individuals who are out of the labor force, such as retirees and those not actively seeking employment. Additionally, we exclude self-employed individuals.

When post-stratification data for a given country-year are unavailable—due to census data being collected only every 5 or 10 years, or limited temporal coverage of registry data—we adopt two strategies based on informed assumptions. We either rely on the weights calculated with data from the following available year (e.g., we use 2004 weights for the year 2002) or estimate the missing data through interpolation or conditional extrapolation from surrounding years. In the case of conditional extrapolation, we project the region-industry employment frequencies while preserving the relative demographic composition within each industry-region cell, as observed in the closest available year (e.g., we extrapolate the number of workers in a given industry-region for years after the latest census, preserving the gender, age, education, occupation composition characterizing that industry-region in the latest census). The complete list of imputations and interpolations/extrapolations, by country, is provided in Section B.2.

In other instances, one or more variables used to post-stratify are only available at a higher level of aggregation than our standard. For instance, industry data may be available only at the sub-section level of the NACE classification, rather than the preferred 2-digit level. In such cases, we adopt an approach inspired by the multilevel regression with synthetic post-stratification (MrsP) method introduced by Leemann and Wasserfallen (2017). This method allows us to reconstruct post-stratification frequency weights at the target granular level by combining coarse joint-distribution data with more granular marginal distributions for one variable. A concrete example of this approach in our study concerns Switzerland. For this country, we rely on detailed post-stratification data from the national version of the EU Labor Force Survey, which provides cell frequencies defined by gender, education, age, region, occupation, and industry. However, industry information is available only at the sub-section level of NACE Rev. 1.1. To refine this to the 2-digit level, we draw Eurostat data that detail employment shares across NACE 2-digit categories within each region-year and NACE sub-section group. Specifically, for each region and year, we compute the share of employees in each 2-digit NACE category relative to the total employment in the corresponding sub-section-level category. These shares are then used to proportionally allocate individuals from the Swiss EU-LFS data—classified by demographic characteristics, region and sub-section—into the more detailed 2-digit NACE categories. A detailed list of instances where this method is used is also included in Section B.2.

The next section describes the mapping procedure from NACE Rev. 2 to NACE Rev. 1.1. The following section provides a comprehensive list of country-specific adjustments to the general harmonization approach, due to data limitations such as those described above.

B.1 Industry classification and harmonization to NACE Rev 1.1

The data in our study span the period 2002-2018, during which time the European classification system for economic activities underwent a substantial revision. In particular, Eurostat and other statistical agencies coordinated the transition from NACE Revision 1.1, introduced in 1993, to NACE Revision 2, adopted in December 2006 and implemented from January 1, 2008, onwards. This transition involved significant changes to the coding system, including at the 2-digit level. To ensure temporal consistency in our analysis, it was essential to harmonize all industry data to a common standard. For this purpose, we constructed a deterministic crosswalk from NACE Revision 2 to NACE Revision 1.1, at the 2-digit level. This means that each Rev. 2 industry is uniquely associated with a single Rev. 1.1 industry category.

The mapping was based on a close reading of official Eurostat documentation (Eurostat, 2008). In cases where a Rev. 2 industry could reasonably be linked to multiple Rev. 1.1 industries, we adopted a majority principle: the Rev. 2 industry was assigned to the Rev. 1.1 industry employing the largest share of its workforce (see Perani et al., 2015).

To further improve temporal consistency and account for structural changes in the classification system, we combined a few NACE Rev. 1.1 2-digit industries. Specifically:

- NACE Rev. 1.1 – 12 (Mining of uranium and thorium ores) was incorporated into NACE Rev. 1.1 – 13 (Mining of metal ores)
- NACE Rev. 1.1 – 30 (Manufacture of office machinery and computers) was incorporated into NACE Rev. 1.1 – 32 (Manufacture of radio, television and communication equipment and apparatus)
- NACE Rev. 1.1 – 33 (Repair, maintenance and installation of machinery and equipment) was incorporated into NACE Rev. 1.1 – 36 (Manufacture of furniture; manufacturing n.e.c.)
- NACE Rev. 1.1 – 37 (Recycling) was incorporated into NACE Rev. 1.1 – 90 (Sewage and refuse disposal, sanitation and similar activities)

The complete crosswalk from NACE Rev. 2 to NACE Rev. 1.1 is presented in Table B.3.

Table B.3: Correspondence table mapping NACE Rev. 2 to NACE Rev. 1.1

NACE Rev. 2	Industry Name	NACE Rev. 1.1	Industry Name
01	Agriculture, farming of animals, hunting and related service activities	01	Agriculture, hunting and related service activities
02	Forestry and logging	02	Forestry, logging and related service activities
03	Fishing and aquaculture	05	Fishing, fish farming and related service activities
05	Mining of coal and lignite	10	Mining of coal and lignite; extraction of peat
06	Extraction of crude petroleum and natural gas	11	Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying
07	Mining and preparation of metal ores	13 (+12)	Mining of metal ores (+12: Mining of uranium and thorium ores)
08	Other mining and quarrying	14	Other mining and quarrying
09	Mining and quarrying related service activities	11	Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying
10	Manufacture of food products	15	Manufacture of food products and beverages
11	Manufacture of beverages	15	Manufacture of food products and beverages
12	Manufacture of tobacco products	16	Manufacture of tobacco products
13	Manufacture of textiles	17	Manufacture of textiles
14	Manufacture of wearing apparel	18	Manufacture of wearing apparel; dressing and dyeing of fur
15	Manufacture of leather and related products	19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
17	Manufacture of paper and paper products	21	Manufacture of pulp, paper and paper products
18	Printing and reproduction of recorded media	22	Publishing, printing and reproduction of recorded media

NACE Rev. 2	Industry Name	NACE Rev. 1.1	Industry Name
19	Manufacture of coke, refined petroleum products and fuels briquettes	23	Manufacture of coke, refined petroleum products and nuclear fuel
20	Manufacture of chemicals, chemical products and man-made fibres, except pharmaceutical products	24	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	24	Manufacture of chemicals and chemical products
22	Manufacture of rubber and plastic products	25	Manufacture of rubber and plastic products
23	Manufacture of other non-metallic mineral products	26	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals	27	Manufacture of basic metals
25	Manufacture of fabricated metal products, except machinery and equipment	28	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, communication equipment, electronic and optical products	32 (+30)	Manufacture of radio, television and communication equipment and apparatus (+30: Manufacture of office machinery and computers)
27	Manufacture of electrical equipment	31	Manufacture of electrical machinery and apparatus n.e.c.
28	Manufacture of machinery and equipment n.e.c.	29	Manufacture of machinery and equipment n.e.c.
29	Manufacture of motor vehicles, trailers, semi-trailers and parts and accessories for motor vehicles	34	Manufacture of transport equipment
30	Manufacture of other transport equipment	35	Manufacture of other transport equipment
31	Manufacture of furniture	36 (+33)	Manufacture of furniture; manufacturing n.e.c. (+33: Repair, maintenance and installation of machinery and equipment)
32	Other manufacturing activities	36 (+33)	Manufacture of furniture; manufacturing n.e.c. (+33: Repair, maintenance and installation of machinery and equipment)
33	Repair, maintenance and installation of machinery and equipment	29	Manufacture of machinery and equipment n.e.c.
35	Electricity, gas, steam, cold and hot water and cold air	40	Electricity, gas, steam and hot water supply
36	Water collection, treatment and distribution	41	Collection, purification and distribution of water
37	Collection, drainage and treatment of sewage	90 (+37)	Sewage and refuse disposal, sanitation and similar activities (+37: Recycling)
38	Waste collection, treatment and disposal activities; materials recovery	90 (+37)	Sewage and refuse disposal, sanitation and similar activities (+37: Recycling)
39	Remediation and similar activities	90 (+37)	Sewage and refuse disposal, sanitation and similar activities (+37: Recycling)
41	Development of building projects; Construction of buildings	45	Construction
42	Civil engineering	45	Construction
43	Specialised construction activities	45	Construction
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	50	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods
46	Wholesale trade (include commission trade), except of motor vehicles and motorcycles	51	Wholesale trade and commission trade, except of motor vehicles and motorcycles
47	Retail trade, except of motor vehicles and motorcycles	52	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods
49	Land transport and transport via pipelines	60	Land transport, transport via pipelines
50	Water transport	61	Water transport
51	Air transport	62	Air transport
52	Warehousing and support activities for transportation (include cargo handling)	63	Supporting and auxiliary transport activities; activities of travel agencies
53	Postal and courier activities	64	Post and telecommunications
55	Accommodation	55	Hotels and restaurants
56	Food and beverage service activities	55	Hotels and restaurants
58	Publishing activities	22	Publishing, printing and reproduction of recorded media

NACE Rev. 2	Industry Name	NACE Rev. 1.1	Industry Name
59	Motion picture, video and television programme production, sound recording and music publishing activities	92	Recreational, cultural and sporting activities
60	Radio and television activities	92	Recreational, cultural and sporting activities
61	Telecommunications	64	Post and telecommunications
62	Computer programming, consultancy and related activities	72	Computer and related activities
63	Information service activities	72	Computer and related activities
64	Financial service activities, except insurance and pension funding	65	Financial intermediation, except insurance and pension funding
65	Insurance, reinsurance and pension funding, except compulsory social security	66	Insurance and pension funding, except compulsory social security
66	Activities auxiliary to financial services and insurance activities	67	Activities auxiliary to financial intermediation
68	Real estate activities	70	Real estate activities
69	Legal and accounting activities	74	Other business activities
70	Activities of head offices; management consultancy activities	74	Other business activities
71	Architectural, engineering and related technical activities; technical testing and analysis	74	Other business activities
72	Scientific research and development	73	Research and development
73	Advertising, market research and public opinion polling	74	Other business activities
74	Other consultancy, scientific and technical activities	74	Other business activities
75	Veterinary activities	85	Health and social work
77	Renting activities	71	Renting of machinery and equipment without operator and of personal and household goods
78	Employment activities	74	Other business activities
79	Travel agency, tour operator, reservation service and related activities	63	Supporting and auxiliary transport activities; activities of travel agencies
80	Security and investigation activities	74	Other business activities
81	Services to buildings and landscape activities	74	Other business activities
82	Office administrative, office support and other business support activities	74	Other business activities
84	Public administration and defence; compulsory social security	75	Public administration and defence; compulsory social security
85	Education	80	Education
86	Human health activities	85	Health and social work
87	Residential care activities	85	Health and social work
88	Social work activities without accommodation	85	Health and social work
90	Creative, arts and entertainment activities	92	Recreational, cultural and sporting activities
91	Libraries, archives, museums and other cultural activities	92	Recreational, cultural and sporting activities
92	Gambling and betting activities	92	Recreational, cultural and sporting activities
93	Sports activities and amusement and recreation activities	92	Recreational, cultural and sporting activities
94	Activities of membership organizations	91	Activities of membership organizations n.e.c.
95	Repair of computers and personal and household goods	52	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods
96	Other personal service activities	93	Other service activities
97	Activities of households as employers of domestic personnel	95	Activities of households as employers of domestic staff
98	Undifferentiated goods- and services-producing activities of private households for own use	97	Undifferentiated services producing activities of private households for own use
99	Activities of extraterritorial organizations and bodies	99	Extra-territorial organizations and bodies

B.2 Country-level adjustments to the harmonization procedure

Table B.4 provides details on country-specific adjustments to the harmonization of post-stratification data due to data availability constraints.

Table B.4: Country-specific adjustments

Country	Notes
Austria	Micro-census data for 2002 are incomplete; hence, we replicate the values from 2004 for that year.
Belgium	While the joint distributions are estimated using national LFS data from Statbel, we further disaggregate them by self-employed versus the rest of the labor force. This breakdown relies on additional EU-LFS data on the distribution of self-employed and non-self-employed individuals by year, sector (NACE sub-section), region, and occupation (ISCO-88, 2-digit), following the MRSP approach (Leemann and Wasserfallen, 2017) described in the first section of Appendix B. This allows us to identify and exclude the self-employed from the dataset, consistent with the approach used for the other countries. Data on context-level variables are estimated at the NUTS 1-digit level; post-stratification data are at the NUTS 2-digit level.
Finland	Registry data for 2002 are incomplete; hence, we replicate the values from 2004 for that year.
France	Data between 2006 and 2011 are linearly interpolated using the full joint distribution available for those two years. For years prior to 2006 and after 2011, we extrapolate region-industry frequencies, while holding constant the relative demographic composition within each industry–region cell at the 2006 and 2011 levels, respectively, as described in the first section of Appendix B.
Germany	Data are available only at the NUTS 1-digit level.
Greece	Data between 2001 and 2011 are linearly interpolated using the full joint distribution available for those two years. For years after 2011, we extrapolate region-industry frequencies, while holding constant the relative demographic composition within each industry–region cell at the 2011 levels.
Ireland	Occupation data are classified at the ISCO-88 1-digit level instead of the 2-digit level. For the year 2016, the NACE 2-digit level breakdown is unavailable. Hence, we compute the NACE 2-digit industry distribution within each NACE sub-section, and use it to proportionally allocate individuals with each given demographic combination into the more-detailed 2-digit NACE categories. The shares of 2-digit NACE employment among industry subsections in 2016 are derived from the Labor Force Survey (LFS) of the United Kingdom for the same year. The choice of using UK data is made because a similar survey is not available for Ireland in 2016. Data for the years between 2006 and 2011, and between 2011 and 2016 are linearly interpolated using the full joint distribution available for each respective pair of years. Data for 2018 are unavailable; hence, we replicate the values from 2016 for that year.
Italy	Data for the years between 2006 and 2012, 2012 and 2016, and 2016 and 2018 are linearly interpolated using the full joint distribution available for each respective pair of years. For the years prior to 2004, we extrapolate region-industry frequencies, while holding constant the relative demographic composition within each industry–region cell at the 2004 level.
Netherlands	Data for 2002 are unavailable; hence, we impute the values of 2003 for that year.
Norway	Data for 2002 are incomplete; hence, we impute the values of 2003 for that year.
Portugal	Data between 2001 and 2011 are linearly interpolated using the full joint distribution available for those two years. For years after 2011, we extrapolate by adjusting weights at the region–industry level, while holding constant the relative distribution within each industry–region cell at the 2011 levels.
Spain	Data between 2001 and 2011 are linearly interpolated using the full joint distribution available for those two years. For years after 2011, we extrapolate region-industry frequencies, while holding constant the relative demographic composition within each industry–region cell at the 2011 levels.
Sweden	No specific adjustments.
Switzerland	While frequency weights are estimated using national EU-LFS data, the industry variable is initially at the NACE sub-section level. We refine the industry variable from NACE Rev. 1.1 sub-section to the 2-digit level using Eurostat data, following an approach inspired by the MRSP method of Leemann and Wasserfallen (2017). For each region and year, we compute the share of employees in each 2-digit NACE category relative to the total employment in the corresponding sub-section category. These shares are then used to proportionally allocate individuals from the Swiss EU-LFS data—classified by demographic characteristics, region and sub-section-level—into the more detailed 2-digit NACE categories.
UK	Data are available only at the NUTS 1-digit level. For observations with unspecified qualifications, we impute educational attainment based on the individual's age at the time of education completion.

C Details on party classification

The list of radical-right parties includes: the Austrian Freedom Party (FPÖ), the Alliance for the Future of Austria (BZÖ) and the Team Stronach for Austria in Austria; the Flemish Bloc (VB) and the Flemish Interest (VB) in Belgium; the True Finns (PS) in Finland; the National Front (FN) in France; the Alternative for Germany (AfD) in Germany; the Popular Orthodox Rally (LAOS) and Golden Dawn (XA) in Greece; the Brothers of Italy - National Centre-right (FDI-CDN), the National Alliance (AN), the Northern League (LN), and the League (L) in Italy; the Forum for Democracy (FvD), the List Pim Fortuyn (LPF) and the Party of Freedom (PVV) in the Netherlands; the Progress Party (FrP) in Norway; VOX in Spain; the Sweden Democrats (SD) in Sweden; the Swiss Democrats (SD/DS), the Federal Democratic Union (EDU/UDF), and the Swiss People's Party (SVP/UDC) in Switzerland; and the United Kingdom Independence Party (UKIP) in the United Kingdom.

The list of radical-left parties includes: the Workers' Party of Belgium (PTB/PVDA) in Belgium; the Left Front (FDG) and the French Communist Party (PCF) in France; the Left Wing Alliance (VAS) in Finland; the Party of Democratic Socialism (PDS), The Left. Party of Democratic Socialism (L-PDS), and The Left (LINKE) in Germany; the Coalition of the Radical Left (SYRIZA), the Communist Party of Greece (KKE), and the Progressive Left Coalition (SYN) in Greece; Civil Revolution (RC), the Communist Refoundation Party (PRC), the Party of Italian Communists (PdCI), and Left Ecology Freedom (SEL) in Italy; the Socialist Party (SP) in the Netherlands; the Socialist Left Party (SV) in Norway; the Left Bloc (BE) and the Unified Democratic Coalition (CDU) in Portugal; United We Can, We Can, United Left (IU), Compromís–Podemos–EUPV, the Galician Nationalist Bloc (BNG), and the Aragonist Council (CHA) in Spain; the Left Party (V) in Sweden; the Swiss Labour Party (PdAS/PdTS) in Switzerland; and We Ourselves (SF) in both the United Kingdom and Ireland.

We classify as mainstream left all the parties that, according to the Manifesto Project data (Volkens et al., 2016), belong to the Ecological, Socialist, and Social-Democratic party families, and are not classified as radical-left parties according to the list above. We classify as mainstream right all the parties that, according to (Volkens et al., 2016), belong to the Liberal, Christian-Democratic, and Conservative party families, and are not classified as radical-right parties according to the list above.

The list of ethnic and regionalist parties (following Volkens et al., 2016) includes: the New Flemish Alliance (N-VA) in Belgium; the Swedish People's Party (RKP/SFP) in Finland; the Basque Nationalist Party (PNV/EAJ), Catalan Republican Left (ERC), Canarian Coalition (CC), Convergence and Union (CiU), Democratic Convergence of Catalonia (CDC), Together for Catalonia (JxCat), Popular Unity Candidacy (CUP), and Amaiur in Spain; and the Scottish National Party (SNP) in the United Kingdom.

The list of single-issue or otherwise hard-to-classify parties includes: the Pirates in Germany; the European

Democracy (DE), List Di Pietro – Italy of Values (IdV), Popular Democratic Union for Europe (P-UDEUR), and Five Star Movement (M5S) in Italy; the 50Plus (50PLUS), Party for the Animals (PvdD), and Reformed Political Party (SGP) in the Netherlands; and the National Solidarity Party (PSN) in Portugal.

D Measure of exposure to robot adoption: data sources

To construct the measure of exposure to robot adoption, we use data on the adoption of industrial robots for the 15 European countries of interest between 1993 and 2018. For Greece and Ireland, data are only available from 1999 and 2002, respectively. In these two cases, we begin measuring robot exposure in 2006, considering that earlier data classify all robots as “unspecified” and lack variation across industries. Robot data are classified into eleven industries that cover the entire manufacturing sector, mainly corresponding to the sub-sections of Section D (i.e. Manufacturing) of the NACE Rev. 1.1 classification. These are presented in Table D.1.

Table D.1: Description of industries

Industry description	NACE Rev. 1.1 code
Food, beverages, tobacco	DA
Textiles and leather	DB-DC
Wood and wood products	DD
Pulp, paper, publishing and printing	DE
Coke, refined petroleum, chemicals, rubber and plastic	DF-DG-DH
Other non-metallic mineral products	DI
Basic metals and fabricated metal products	DJ
Machinery and equipment n.e.c.	DK
Electrical and optical equipment	DL
Transport equipment	DM
Manufacturing n.e.c. (furniture, toys, sports goods, etc.)	DN

Annual data on the stock of operational industrial robots by country and industry are sourced from the International Federation of Robotics (IRF), which compiles data reported by robot suppliers, with support from national robotics associations. The operational stock of robots measures the number of robots in use each year in each given country and industry, based on an assumed average service life of 12 years and immediate withdrawal from service thereafter. The IRF adopts the ISO 8373:2012 definition of an industrial robot as an “automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which can be either fixed in place or mobile for use in industrial automation applications” (Müller, 2022).

The pre-sample employment data used for the computation of robotization exposure are obtained from national sources and Eurostat. They refer to an year between 1988 to 1995, depending on the country. Table D.2 provides an overview of the data sources used for each country. The data are obtained at the NUTS 2-digit level, except for Germany and the United Kingdom, where they are obtained at the NUTS 1-digit level.

For the instruments, we use annual data from 1993 to 2018 on three variables designed to capture technological shifts. The first instrument relies on the Producer Price Index for the industry “Electronic Computer Manufacturing”, sourced from the Federal Reserve Bank of St. Louis’s FRED Economic Data (US Bureau of Labor Statistics, 2023). The electronic computer manufacturing category includes both primary and secondary products. The former

Table D.2: Employment data sources

Country	Data Source
Austria	Eurostat Regional Employment Statistics; Austria Statistics
Belgium	National Bank of Belgium
Finland	Statistics Finland
France	National Institute of Statistics and Economic Studies (INSEE)
Germany	Federal Employment Agency
Greece	Hellenic Statistical Authority – Enterprise Census of 1988
Ireland	Eurostat Regional Employment Statistics; Central Statistics Office
Italy	National Institute for Statistics (ISTAT)
Netherlands	Statistics Netherlands (CBS)
Norway	Statistics Norway (SSB)
Portugal	Statistics Portugal (INE)
Spain	National Statistics Institute (INE)
Sweden	Statistics Sweden (SCB)
Switzerland	Swiss Statistics (SFSO)
United Kingdom	Office for National Statistics (ONS)

encompasses both single-user computers and other types of computers, including host and multiuser computers. This index is set to 100 in 2004 and reported annually without seasonal adjustments. The second instrument relies on single-thread performance (measured in SpecINT x 10^3), which reflects the speed at which a single thread (i.e., a sequence of instructions within a program) can be executed by a processor core. The third instrument relies on the number of transistors (in thousands) per microprocessor, which influences the complexity and performance of calculations. Data for the second and third instruments are sourced from Rupp (2021).

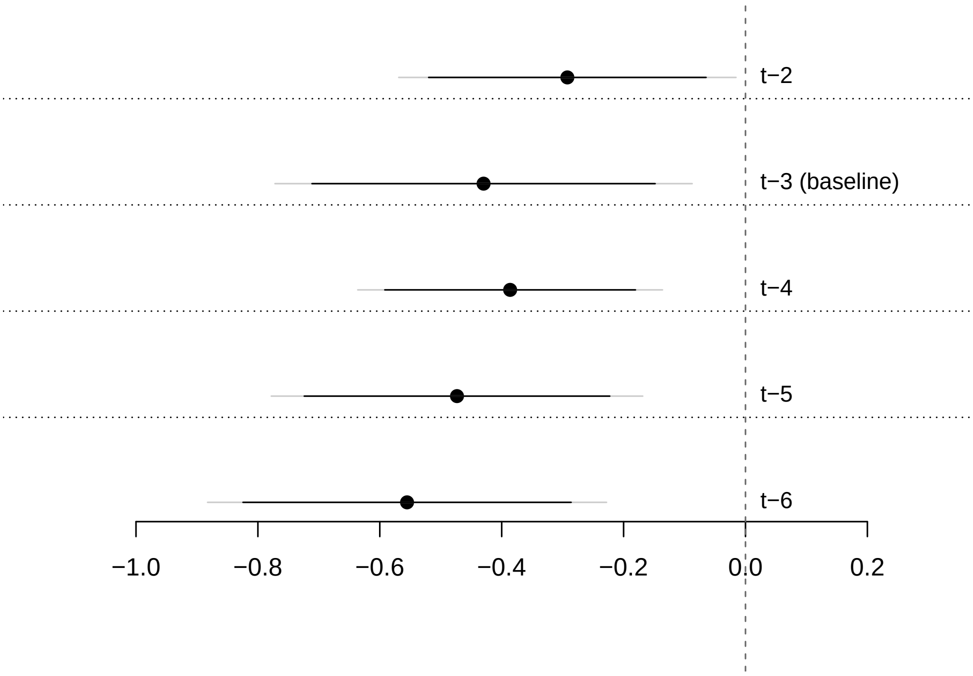
E Additional results

Table E.1: Industry-level analysis - robustness

Dep. Var.:	Industry Union Density			
	(1)	(2)	(3)	(4)
MRP Pred. Model	Best by country		Sector trends	
Estimator	OLS	2SLS	OLS	2SLS
Industry Robot	0.266**	0.142	0.233*	0.066
Exposure	[0.126]	[1.028]	[0.125]	[1.031]
Country-year FE	X	X	X	X
Industry FE	X	X	X	X
Observations	2,540	2,540	2,540	2,540
Std dev. Y	6.153	6.153	6.153	6.153
Std dev. X	0.869	0.869	0.869	0.869
Magnitude	0.0376	0.0200	0.0329	0.00938
First stage F-stat		4.507		4.507

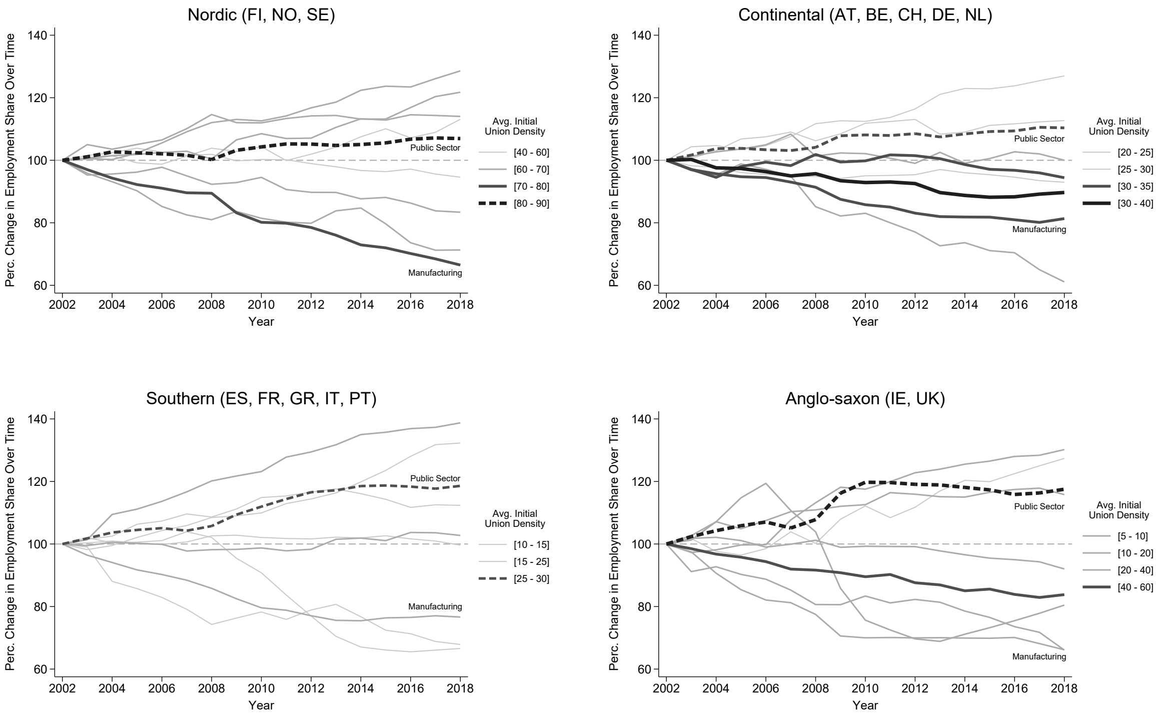
Note: The dependent variable is the unionization rate at the country-industry-year level. Unionization estimates are obtained through MRP employing the highest ranked prediction model by country in columns 1-2, while model 1 of Section A is employed in columns 3-4. The table reports the standard deviation of the dependent variable and of robot exposure, along with the magnitude of their relationship, after residualizing with respect to country-year and industry fixed effects. Robust standard errors in brackets.
*** p<0.01, ** p<0.05, * p<0.1.

Figure E1: Robustness: different periods



Note: The dependent variable is the unionization rate at the region-year level. Unionization estimates are obtained through MRP employing the baseline model. The coefficients reported in the plot correspond to the effect of regional robot exposure, measured over alternative time periods ranging from the previous two to six years. All models include country-year and region fixed effects. The figure displays point estimates along with 95% (dark bars) and 90% (light bars) confidence intervals, computed using robust standard errors.

Figure E2: Employment trend by unionization regimes



Note: The figure displays changes in employment shares across broad sectors from 2002 to 2018, expressed relative to their levels in 2002. Employment shares are calculated by aggregating Eurostat employment data across the fifteen countries included in the study, grouped into broad sectors: Primary Sector (NACE Rev 1.1: sections A, B, C); Manufacturing (D); Utilities and Construction (E, F); Trade and Retail (G); Hotels and Restaurants (H); Transport and Communication (I); Finance and Business (J, K); Public Sector (L, M, N, Q); and Personal and Community Services (O, P). Each country's contribution is weighted equally, so the series reflects the unweighted average sectoral composition across countries. The aggregation is performed by different regimes of unionization: nordic (Finland, Norway and Sweden), continental (Austria, Belgium, Switzerland, Germany and the Netherlands), southern (France, Greece, Italy, Portugal and Spain) and anglo-saxon (Ireland and the UK) regime. Darker and wider lines indicate sectors with higher initial union density in 2002.

Chapter 3

The Candidate Factory: Technological Change and Political Supply

The Candidate Factory: Technological Change and Political Supply[*]

Paolo Agnolin[†]

Abstract

The transition from industrial to post-industrial societies is reshaping political landscapes, with studies documenting a decline in working-class representation. This paper provides a structural explanation for the changing composition of political elites, focusing on the role of technological advancements such as automation and robotization. I argue that the distributional effects of automation play a key role in shaping political representation. Specifically, I theorize that automation alters the distribution of economic resources and opportunities across occupations, influencing who runs for office. I propose that those most negatively affected by automation, particularly individuals from working-class or automatable backgrounds, are increasingly underrepresented among political candidates. This research explores the impact of material changes on political candidacy, drawing on data from U.S. state legislatures, local offices in California, and representative surveys. The findings reveal that in areas with higher exposure to automation, there is a notable decline in candidates from occupations most affected by robotization. This research contributes to a deeper understanding of the structural factors shaping the supply of political candidates and their implications for representation and policy outcomes.

[*]I thank Catherine de Vries, Massimo Anelli, Pablo Beramendi, Italo Colantone, Silja Häusermann, Todd Makse, Michele Pellizzari, Piero Stanig, Georg Vanberg, and seminar participants at the 2025 APSA Labor Politics Workshop, APSA 2024, MPSA 2024, Bern University, Bocconi University and Duke University for insightful comments. The usual disclaimer applies.

[†]Bocconi University and Dondena Centre for Research on Social Dynamics and Public Policy, Via Roentgen 1, 20136 Milan, Italy. Contact: paolo.agnolin@unibocconi.it.

1 Introduction

Several studies have documented a relevant transformation in the composition of political elites in recent decades. The share of female candidates has increased significantly in most democracies around the world, along with an increasing percentage of candidates and MPs with a university education (Lamprinakou *et al.*, 2017). At the same time, many parties have seen declines in working-class legislators and increases in professional career politicians (Borchert, 2003; Evans and Tilley, 2017; MacKenzie, 2015). O'Grady (2019) argues that many parties have become dominated by middle-class professional politicians with little experience outside of politics itself, while working-class people find it increasingly difficult to enter politics. Examining these changes in the supply side of politics is essential, as shifts in political recruitment directly influence the diversity of interests represented in policy debates and lead to real political consequences (Burden, 2015; O'Grady, 2019).

Post-industrial democracies are characterized by a general transformation of the power relationships between traditional parties and new incumbents. However, the implications of the significant changes in the occupational structure and the emergence of a new knowledge economy are not only restricted to competition between parties. They also deeply shape intra-party dynamics and the succession of political candidates. A prominent example is provided by mainstream social democratic parties in western countries, whose political leaders have generally moved away from their traditional working-class constituencies and have embraced new social and civic libertarian issues that appeal to urban professionals and cosmopolitan voters. The transition from an industrial to a post-industrial society is accompanied by a transformation in the composition of political elites and by changes in the socio-economic background of party members.

Despite the existing evidence on the compositional change in the population of politicians and the importance of political elites for the democratic process, we still know relatively little about the structural economic determinants of this phenomenon. As noted by Carnes and Lupu (2023), there is less research on the causes of inequalities in politicians' backgrounds than there is on the consequences. Research in political science and political economy has focused on the effects of economic shocks (e.g. Colantone and Stanig, 2018) and occupational realignment (Kitschelt and Rehm, 2023) to explain changes in voting behavior and party systems. However, our understanding of the causes of this occupational realignment in the pool of political candidates remains limited.

In this paper, I provide a structural explanation for the changes on the supply side of politics, and in particular on the economic background of political candidates. I theorize that the changes from a traditional occupational structural to a post-industrial society, as driven by technological change, determine a reconfiguration of the political elite, characterized by a new composition of occupational backgrounds and distinct policy platforms. In particular, I focus on the occupational background of politicians, given how individual's work experiences on the job play an important part in shaping political attitudes (Kitschelt and Rehm, 2014).

I theorize that changes in the composition of political elites and candidates are affected by shifts in the distribution of resources across the occupational structure of electoral districts. In particular, I argue that technological changes such as computerization and robotization have accelerated the reallocation of workers in different industries and occupations and have affected the material well-being and earning trajectories of such workers. This occupational realignment has certainly reshaped the political demands of different social groups and their relative political weight in the electorate, as extensively documented in the literature (e.g. Beramendi, 2015). At the same time, however, changes in individuals' material conditions have altered their economic incentives and access to resources, impacting their ability to run for office and, consequently, the composition of the candidate pool. I theorize a model of behavior in which the probability to enter the political arena and to run for office is a function of economic material condition. By altering the distribution of resources across different social groups and by transforming the power economic relationship in society, technological change thus affects the propensity to run for office of citizens from different socio-economic backgrounds.

For instance, in an electoral district in which the economic transformation has moved resources to high-skilled workers in less automatable occupations, I expect an increase in the propensity of such workers to enter the political arena and to run for office. By the same token, the loss of employment and career prospects for workers in industries and occupations more exposed to technological change may lead to exacerbate the under-representation of these workers in the population of politicians.

I argue against economic models that suggest that free-riding incentives and lower opportunity costs give citizens with lower-market wages have a comparative advantage in entering political life (Caselli and Morelli, 2004). On the contrary, I expect increasing over-representation of candidates representing the "winners" of technological change, and that are thus characterized by better pre-office earning prospects.

At the same time, I challenge the notion of an "inclusive meritocracy," by which politics selects competent individuals broadly representative of all social backgrounds (Dal Bó *et al.*, 2017). On the contrary, I argue that individuals from socio-economic backgrounds facing declining economic prospects (i.e. the "losers" of technological progress) become disproportionately less likely to pursue political candidacy. This dynamic exacerbates political inequality by deepening the under-representation of less affluent backgrounds among political candidates, highlighting a significant link between economic and political inequality.

In a nutshell, I argue that the transition from a traditional occupational structure to a post-industrial society, driven by technological change, reshapes the occupational composition of political candidates. Technological change alters the distribution of resources across occupations, redefining who benefits and who loses economically. These shifts in resource allocation and economic inequality shape the pool of individuals entering politics, influencing the characteristics of new candidates.

This study focuses on the United States, a nation characterized by a strong relationship between electoral districts and the representatives elected from those districts. It begins by documenting broad changes in elite composition, focusing on state legislative representatives across the U.S. from 1997 to 2012. To explore structural shifts in the broader pool of political candidates, the empirical analysis centers on the full population of candidates for lower-level offices in California—such as city councils, school boards, and city supervisors—between 1998 and 2022. Candidates are classified by their occupational backgrounds using a text-as-data approach, and this data is linked to district-level indicators, including exposure to automation and robotization. To address potential endogeneity, district-level exposure to automation is instrumented using industry-level spread of robots in other advanced economies. The findings reveal a systematic decline in the share of candidates from working-class occupations and those most adversely affected by automation in counties with greater exposure to robotization.

To further support my findings, I use data from the nationally representative Cooperative Election Studies (CES) to examine the impact of exposure to robotization on political candidacy across social groups. The analysis reveals that automation exacerbates inequality in political participation along skill lines and individual exposure to automation, with evidence pointing to the distributional effects of automation on income as a potential driver.

By exploring the impact of technological and occupational shifts on participation and candidate selection, this research sheds new light to key political developments, including trends in political representation, inequality, and the reconfiguration of political elites. As political elites become more socio-economically homogenous, the disconnect between the political class and economically insecure groups, such as the working class and routine workers, risks deepening, raising critical implications for democratic representation.

2 Background and theory

2.1 The distributional effects of technological change

Automation and technological change have emerged as significant determinants of labor market outcomes and earnings distribution in the economic literature. Recent advances in technology and the automation of work have driven significant productivity gains but also profound distributional effects. Notably, these changes have exacerbated wage inequality both across geographic areas and, more critically, between social groups and occupations. Early research highlighted widening wage disparities between broad occupational categories, particularly along the skill spectrum. More recent studies, however, have revealed even sharper divergences between occupations, as earnings trajectories increasingly depend on the level of automatability and the intensity of routine tasks associated with each role. These findings underscore the nuanced and uneven economic impacts of automation and technological change.

Many observers have used the concept of skill-biased technological change (SBTC) to explain diverging earnings patterns, particularly the rising wage premium for workers with tertiary education (Katz and Murphy, 1992). This explanation is supported by evidence of the sustained growth in demand for highly skilled, educated workers and women, resulting in significant implications for wage structures and labor market dynamics.

The wage premium for education has increased across several industrialized countries, especially in the late 20th century. In the United States, for example, Goldin and Katz (2007) estimate that the increase in the education premium, particularly with post-secondary education, explains about two-thirds of the overall rise of earnings polarization between 1980 and 2005. The dramatic rise in wage premium is empirically linked to higher education and cognitive ability and is consistent with the SBTC hypothesis, with both supply and demand for skills that matter to shape inequality (Autor, 2014). Based on this framework, it can be argued that technological change has led to distributional effects across broad occupational categories, depending on the skill level. This is particularly relevant in political science research on working-class representation, given that most working-class occupations tend to require lower skill levels than non-working-class occupations.

The wage premium for education has risen significantly across industrialized countries, particularly during the late 20th century. In the United States, Goldin and Katz (2007) estimate that the increasing returns to post-secondary education account for approximately two-thirds of the rise in earnings polarization between 1980 and 2005. This sharp increase in the education premium has been empirically linked to higher levels of education and cognitive ability, aligning with the skill-biased technological change (SBTC) hypothesis. According to this framework, both the supply and demand for skills play a central role in shaping inequality (Autor, 2014). Technological change has thus produced substantial distributional effects across broad occupational categories, with skill levels serving as a critical determinant. From the perspective of this paper, it is important to note that most working-class occupations are typically characterized by lower skill levels compared to nonworking-class occupations.

Despite its success in explaining some tendency in the earning distribution, however, SBTC cannot fully account for the recent phenomenon of job polarization in the U.S. and other Western countries. This phenomenon, marked by income gains for both high-skilled, educated workers and low-wage service-sector employees, contrasts with stagnating or declining earnings for the middle class (Autor *et al.*, 2006; Autor and Dorn, 2013; Goos and Manning, 2007). To address this, the routine-biased technological change (RBTC) hypothesis has emerged, offering a more comprehensive explanation for job polarization. Technological change is biased toward replacing labour in routine tasks and often leads to task offshoring, resulting in a decrease in the demand for middling relative to high-skilled and low-skilled occupations, with subsequent polarization in the earnings trajectories of high relative to low routine task intensity (Acemoglu and Autor, 2011; Autor *et al.*, 2003;

Goos *et al.*, 2014).

As a consequence, the structural economic relationships between different social groups in a given area is affected by the introduction of technological innovation, such as for instance computerization and robotization. Workers in regions reliant on industries that heavily adopted robotics, as well as those whose skills were easily replaced by new technologies, have been particularly impacted. Stable, relatively well-paying jobs in manufacturing have largely disappeared, replaced by lower-paying roles in service industries and emerging sectors like the gig economy. However, these new roles often come with lower wages, precarious contracts, and limited labor protections.

The consequence has been a sharp rise in income inequality and widening wage disparities. Occupations more susceptible to automation have seen declining wages and fewer opportunities, while those less automatable have benefited from improving earnings trajectories. This shift has fundamentally altered the economic power dynamics within society, deepening divides between social groups and exacerbating disparities in economic opportunities and outcomes.

It is crucial to recognize that the acceleration of automation produces distributional effects that cannot be captured through a broad distinction between occupations. Instead, a more nuanced differentiation based on the automability of occupations—both within and outside of the working class—is necessary. While working-class occupations tend to be more automatable on average than white-collar jobs, the literature on routine-biased technological change suggests that understanding shifts in economic resources and power requires a detailed investigation of the varying levels of exposure to automation. This more precise analysis helps illuminate how technological change affects different social groups and who bears the economic costs of progress.

2.2 A structural theory of changes in the supply side of politics

Several recent studies have documented a significant transformation in the composition of political elites across the globe. Notably, these studies highlight an increasing presence of female candidates and parliamentarians with university degrees (Lamprinakou *et al.*, 2017). This trend, however, appears to be accompanied by a concurrent decline in the representation of working-class individuals within political institutions, with a corresponding rise in professional career politicians (Borchert, 2003; Evans and Tilley, 2017; MacKenzie, 2015). O'Grady (2019) further suggests that this shift has resulted in the dominance of middle-class professionals within many political parties, individuals who often lack substantial experience outside the realm of politics itself. This trend, coupled with the growing challenges faced by working-class individuals in entering the political sphere, necessitates a closer examination of the supply side of politics. As noted by Burden (2015) and O'Grady (2019), these shifts in political recruitment have significant implications for the representation of diverse interests within policy debates and ultimately lead to tangible political consequences.

While much of the existing literature has focused on the consequences of political elites' economic backgrounds, there has been limited research into the causes behind the underrepresentation of working-class individuals in politics. Carnes and Lupu (2023) note that, while earlier scholars proposed various hypotheses about the origins of this inequality, there has been little consensus or conclusive progress in understanding why certain economic groups are more likely to enter politics. Modern scholarship has made some advances, but the underlying mechanisms remain unclear. This gap in research may be attributed to the complexity of the issue, as studying the causes of economic inequality among politicians presents greater challenges than examining its consequences. In their recent work, Folke and Rickne (2024) explore potential explanations for the "class ceiling" in political representation, rejecting the notions that working-class individuals lack political ambition, public service motivation, honesty, or voter support. They argue that political parties are significantly less likely to promote working-class candidates, even when they possess the necessary qualifications. However, the question of when and how less affluent socio-economic groups are excluded from political careers still requires further investigation.

While factors such as gatekeeping, party systems, and voter biases undoubtedly play a role, focusing on structural explanations—particularly the economic conditions that shape different groups' ability and propensity to engage in politics—can offer valuable insights into the broader dynamics influencing political representation. This perspective aligns with arguments in the literature on representation, which suggests that a more comprehensive understanding of the underrepresentation of certain social groups must consider the possibility that the gap in representation originates at the supply stage (Dancygier *et al.*, 2021).

The broader literature on the descriptive representation of social groups often conceptualizes the path to becoming an elected politician as a sequence of distinct stages, often likened to a pipeline that can be examined for potential leaks (Carnes and Lupu, 2023; Gulzar, 2021; Fox and Lawless, 2005; Norris and Lovenduski, 1993). To attain elected office, individuals must first be able to run, have the desire to hold office and then formally enter the race. In many countries, they must then secure selection by their party before ultimately winning enough votes to be elected. If members of a particular social group are disproportionately filtered out at any stage of this process, they will be numerically underrepresented in elected office. Theories explaining these patterns are often categorized into supply and demand perspectives (see Carnes and Lupu, 2023). Supply-based explanations focus on differences in the availability and predisposition to run of candidates from a given social group, while demand-based explanations examine voters and other political actor facilitate or hinder their political participation

This paper advances a supply-side, structural economic explanation for working-class representation and the representation of social groups affected by technological change. I theorize a reciprocal relationship between the transformation of political elites, as evidenced by changes in candidate composition, and distributional dynamics across the occupational structure within electoral districts. Technological advancements, such as computerization and robotization, have intensified the reallocation of workers across industries and occupations, thereby affecting both their material well-being and earning potential (Autor *et al.*, 2003). Prior research (e.g., Beramendi, 2015) has explored how such occupational realignment can reshape the political demands of different social groups, altering their relative political influence. However, this paper offers a novel perspective by arguing that such transformations also shape the pool of political candidates. Specifically, the impact of automation on the composition and economic resources of various groups seeking office is central to understanding how technological change affects political participation.

I propose a model of behaviour where the propensity to enter the political arena is determined by the economic well-being of individuals and job conditions, which can arguably also affect factors such as social status. Technological change redistributes resources across different social groups, creating disparities in the material and social incentives for political engagement. As a result, automation is theorized to differentially affect the likelihood that individuals from various socio-economic backgrounds will pursue political careers. In essence, automation acts as a shock to the distribution of economic resources, with individuals in more automatable occupations experiencing a decline in economic prospects. Consequently, individuals with stronger economic and labor market prospects are more likely to run for office. Political parties, aiming to maximize electoral success and minimize campaign spending, are expected to prioritize candidates with greater economic and social advantages. By shifting the distribution of economic resources across occupations, technological change thus has the potential to significantly alter the composition of political candidates and, by extension, the political elite.

My argument opposes the prediction of adverse selection, by which able people with higher private incomes and more promising careers face a higher cost to enter public life. Standard models of occupational choice suggest that free-riding incentives and lower opportunity costs give the less competent and poorer a comparative advantage at entering political life (Caselli and Morelli, 2004). While pecuniary considerations matter for self-selection decisions[1], some have shown that a combination of strong intrinsic motivation and high wages for full-time positions preserves the incentives of successful individuals to enter politics (Dal Bó *et al.*, 2017).

From an empirical perspective, previous research has shown that politicians report higher level of achievement, at least for mayors and national legislators, easily above the general population (Dal Bó *et al.*, 2017; Ferraz and Finan, 2009; Merlo *et al.*, 2010; Galasso and Nannicini, 2011; Gagliarducci and Paserman, 2012; Dal Bó and Rossi, 2011; Dal Bó *et al.*, 2009). Also, parties appear to screen positively, promoting richer and more competent individuals to higher ranks (Carnes, 2020; Dal Bó *et al.*, 2017).

In line with empirical findings, I expect individuals with more economic endowments to be more likely to enter the political arena. The relative economic position of the individuals is a function of their exposure to technological change and automation. Exposure to technological change does not only affect the material condition of workers, but also their social status, which is similarly a determinant of political success. To provide an analogy with the marriage "market", Anelli *et al.* (2024) have shown that in areas more exposed to robots, gender gaps in income and labor force participation declined, reducing the relative economic stature of men. Robot penetration also triggered an increase in both divorce and cohabitation and a decline in the number of marriages, which highlights the loss of social status of the economic "losers". Just as the loss of economic and social status has led the economic "losers" to be less successful in the marriage "market", exposure to automation here is expected to make them less successful in the political arena as well.

[1] Dal Bó *et al.* (2017) shows how municipalities with higher relative remuneration for mayors and vice mayors, compared to average earnings, tend to attract individuals with greater competence. Similarly, parties with a higher probability of filling these positions full-time exhibit display higher competence among their top politicians. Additionally, those from steeper age-earnings professions may sometimes exhibit a lower likelihood of entering politics

Consistent with my claim, Dal Bó *et al.* (2017) document strong positive selection into the pool of candidates for more able and successful politicians even though politicians have substantially higher pre-office incomes—hence, higher opportunity costs–than the general population. However, they also show that representation of social background, whether measured by parental earnings or social class, is very even. While they interpret this as "inclusive meritocracy" in Sweden, by which politics selects competent politicians broadly representative of all social backgrounds, I contend that this dynamic might shift in contexts with lower social mobility and weaker safety nets[2]. In such contexts, I propose the concept of "inclusive plutocracy," where individuals from diverse backgrounds may enter politics, but their likelihood of candidacy is still significantly influenced by economic resources, job prospects, and social status. All these factors, are a function of structural economic transformations, including technological change and the automation of work.

In the United States, the likelihood to become a politician is disproportionally higher for richer citizens. Carnes (2020) has documented how individuals with working-class backgrounds make up around half of the population but a mere 2% of Congress and 10% of city councilors. Workers are less likely to hold office because they are less likely to run in the first place, not because they are unqualified or because voters prefer more affluent candidates. Carnes (2020) argues that workers are just as likely to have the qualifications that both party leaders and voters deem important, but they do not eventually run because of the many burdens associated with office seeking and the extensive resources that campaigning requires. In addition, workers are less likely to be recruited and encouraged to run by political elites, party leaders, and interest groups.

Furthermore, the decision to pursue a political career inherently involves financial risk. While the potential for a high salary in a full-time political position might hold less appeal

[2]While a regime on inclusive meritocracy may be observed in a country like Sweden, in which intergenerational mobility is high and parties represent all segments of the income distribution, I argue that the same logic does not apply to countries in which such conditions does not hold. In particular, a lower degree of intergenerational mobility and the lower presence of active labor market policies and safety nets jeopardize a broad representation across social backgrounds.

for individuals with strong existing labor market prospects, they may be more willing to navigate the risks associated with a political campaign. Conversely, for individuals facing precarious employment due to factors like automation, the risks associated with running for office may appear particularly daunting.

A more general interpretation of the relationship between structural shocks with distributional effects and political participation can also be framed within the well-documented link between economic and political inequality. Rising economic inequality constrains the political participation and representational opportunities of lower-income citizens while simultaneously enhancing the ability of more affluent social groups to secure representation and exert influence (e.g. Ansell and Gingrich, 2024; Beramendi *et al.*, 2024).

In conclusion, changes in the likelihood to run for office for citizens from different socioeconomic background is relevant because it predicts changes in policy platforms. In fact, politicians' actions as a representative can be influenced by their background (Burden, 2015). For instance, occupations help shape legislators' ideologies by affecting their material interest and help determine which groups of people legislators instinctively empathize with (O'Grady, 2019). Different trends in political selection affect the range of political interests represented in the policy debate and lead to tangible political consequences.

3 Data and methods

This article examines the United States, a political context where there is a strong connection between electoral districts and the representatives elected to those districts. Candidates typically run on platforms that address the specific needs and concerns of their constituents.

This study utilizes multiple datasets to examine the relationship between technological change and the socio-economic composition of political candidates. The data spans different levels of government and geographic regions in the US. To begin, I use data on U.S. state legislators from 1997 to 2012 to document descriptive trends in political representation by occupational background. Data on the occupational background of state legislators are sourced from Makse (2019).

I then turn to an analysis of representation at lower-level offices in California from 1998 to 2022. Here, I combine data from the California Elections Data Archives (CEDA) and assign occupational background using a text-as-data approach. Studying lower-level candidates offers the advantage of examining selection patterns at a stage where economic barriers to political participation are lower. However, it also highlights the ongoing exclusion of significant portions of society from the political process, with underrepresentation of less affluent groups that arguably tends to become more pronounced at higher levels of political office. Lastly, I draw on nationally representative survey data from the Cooperative Election Studies (CES) to further explore the impact of automation on political candidacy across skill levels and individual exposure to automation, shedding light on the underlying mechanisms driving these dynamics.

3.1 U.S. State Legislatures Data

The first major source of data is the dataset on the occupational background of state legislators, which is drawn from the work of Makse (2019). This dataset includes detailed information on 14,826 legislators serving in state legislatures across the United States between 1997 and 2012. The dataset includes legislators from both the upper and lower houses of state legislatures, with occupational data classified into 44 distinct categories. These categories span a broad range of occupations, from those in managerial and professional fields to manual labor and service occupations. This allows for an analysis of the occupational background of legislators and changes in these patterns over time.

The state legislatures data are crucial for understanding long-term trends in the composition of political elites at the state level. By examining these trends, this study is able to descriptively explore trends in representation by occupational background.

While the dataset on state legislators is comprehensive, it has several limitations. One challenge is the broad occupational categories, which may obscure important distinctions between occupations that are differently impacted by automation. Another limitation is the blending of demand-side and supply-side dynamics. The U.S. State Legislatures dataset captures only the final pool of representatives who have successfully navigated the selection process and won their elections, meaning it does not account for potential candidates who were screened out earlier or those who lost in the race. To address this, the study places greater emphasis on the CEDA dataset, which includes a broader population of candidates, including those running for lower-level offices, to better capture the broader dynamics of political candidate selection.

3.2 California Elections Data Archives (CEDA)

The main data source employed in this study is the California Elections Data Archives (CEDA), which includes the entire population of candidates running for local office. The CEDA dataset includes information on 95,721 candidates who ran for various lower-level political offices in California between 1998 and 2022. These offices include positions on city councils, county supervisors, school boards, and other local government bodies. This dataset provides both individual-level data on candidates and contextual information about the offices they sought.

The CEDA dataset is particularly valuable for this research because it captures a wide range of candidates, including those who may not be part of the political elite but still play an essential role in local governance. By focusing on candidates running for lower-level offices, this study is able to explore the effect of technological change on the broader pool of potential political candidates, beyond the elite circles that dominate state legislatures. Additionally, the CEDA data include detailed ballot designation worksheets, which provide information about candidates' occupations, making it possible to classify them into specific occupational categories.

The detailed occupational information in the CEDA dataset is essential for identifying trends in the types of individuals who are running for local office. By linking these data with measures of technological exposure, it is possible to assess how automation impacts the occupational backgrounds of candidates at the local level.

For the CEDA data, the occupational background of candidates is determined by text strings from the ballot designation worksheets. These text strings describe candidates' occupations in varying levels of specificity (e.g., "electrical foreman," "technology instructor/parent"). To classify these occupations into broader categories, I use a matching technique that combines the ballot designation data with the ONET-SOC (Occupational Information Network - Standard Occupational Classification) system. The ONET-SOC system provides a standardized classification of occupations and is used to assign each candidate to an appropriate occupational category based on their description.

To ensure accuracy in matching occupations, I employ a term frequency-inverse document frequency (TF-IDF) model, a method commonly used in text analysis. The TF-IDF model assigns a weight to each term in the occupation description based on its frequency in the dataset, and the model then measures the similarity between ballot strings and a set of alternative occupational titles from the O*NET-SOC database. By calculating the cosine similarity between these text strings, I assign the closest matching occupational category to each candidate. This process allows for a consistent classification of candidates.

The classification of candidates into occupational groups is essential for the analysis of how automation affects political supply. It enables the study to compare changes in the share of candidates from different occupational backgrounds and assess how these changes are linked to contextual exposure to automation.

3.3 Classification of occupation and analytical categories

Based on the text from the candidate's worksheet, each candidate is assigned a detailed 6-digit SOC (Standard Occupational Classification) category. This granular occupational classification is then used to categorize candidates into broader groups relevant to the analysis. These categories are based on a combination of standard occupational classifications and specific measures of automability, which help to distinguish between occupations that are more or less vulnerable to automation. The construction of these categories is informed by both established classification systems and measures of routine versus non-routine tasks, which are essential for understanding the distributional impacts of technological change.

One of the primary classification systems used is the International Standard Classification of Occupations (ISCO), which groups occupations into broad categories based on the tasks they involve. To begin, occupations are categorized into working-class and non-working-class groups. The working-class category (ISCO 4 through 8) includes clerical support workers, service and sales workers, skilled agricultural workers, craft-related trades, plant and machine operators, and elementary occupations, while non-working-class occupations encompass managers, professional technicians and associate professionals.

Alternatively, I employ the occupational classification by Oesch (2013), which offers a more refined distinction between different types of professionals and workers. Oesch's framework classifies occupations into categories such as self-employed professionals, small business owners, technical professionals, production workers, and service workers. This classification allows for a nuanced understanding of the socio-economic structure of the labor force and its representation in politics. Oesch's distinction between routine and non-routine occupations is particularly valuable, as it enables an analysis of how technological change impacts different groups. Routine tasks, which involve repetitive activities, are more likely to be automated, while non-routine tasks, which require problem-solving and interpersonal skills, tend to be more resistant to automation.

To further refine the analysis, the study incorporates measures of automability based on the content of tasks within each occupation. These measures are drawn from existing literature on automation and labor markets, such as the Routine Task Intensity (RTI) index developed by Goos *et al.* (2014), which quantifies the extent to which an occupation involves routine tasks. The RTI index, along with measures of manual, abstract, and cognitive tasks (as used by Autor and Dorn, 2013), helps to assess the degree to which each occupation is exposed to automation. These indices are particularly useful for understanding how automation is reshaping the political candidate pool, as they allow the study to distinguish between occupations that are more susceptible to automation and those that are less so.

3.4 County-level exposure to automation

To quantify automation exposure, this study draws on the methodology used by Acemoglu and Restrepo (2020) and combines data on industrial robot adoption with employment data from the U.S. Census Bureau. Automation exposure is calculated by determining the extent to which different counties are affected by the increase in the use of industrial robots, particularly in manufacturing industries. This measure captures the exposure to automation at the county level over time, allowing the study to examine how technological disruptions influence the local supply of political candidates.

In particular, county-level exposure to automation is estimated as follows

$$\text{County Exposure}_{ct} = \sum_j \frac{L_{cj}^{\text{pre-sample}}}{L_c^{\text{pre-sample}}} * \frac{R_j^{t-1} - R_j^{t-n}}{L_j^{\text{pre-sample}}},$$

where c indexes county, j represents manufacturing industries and t is years. $R_j^{t-1} - R_j^{t-n}$ indicates the change in the operational stock of industrial robots over the past 4 years in industry j. Data on employment by county and industry are sourced from the US Census Bureau, whereas data on the number of operation industrial robots are sourced from the International Federation of Robotics.

The ratio $\frac{R_j^{t-1} - R_j^{t-n}}{L_j^{\text{pre-sample}}}$ provides a measure of the intensity of robot adoption at the industry level. To retrieve the county-level exposure, we take a weighted summation of the industry-level changes, where the weights capture the relative importance of each industry in each county. Specifically, each weight is the ratio between the number of workers employed in a given county and industry ($L_{crj}^{presample}$), and the total number of workers employed in the same county ($L_{cr}^{presample}$). Importantly, weights are based on pre-sample figures, dating before the surge in the adoption of industrial robots observed from the 1992. Intuitively, districts that were initially specialized in industries in which the adoption of robots has later been faster are assigned stronger exposure to automation.

In order to address endogeneity concerns, I then instrument the robot shock with change in the stock of operational robots in industry j in other countries, specifically Japan and South Korea.

$$\text{IV County Exposure}_{ct} = \sum_j \frac{L_{cj}^{\text{pre-shock}}}{L_c^{\text{pre-shock}}} * \frac{\bar{R}_{-C,j}^{t-1} - \bar{R}_{-C,j}^{t-n}}{\bar{L}_{-Cj}^{\text{pre-shock}}},$$

where:
$\frac{\bar{R}_{-C,j}^{t-1} - \bar{R}_{-C,j}^{t-n}}{\bar{L}_{-Cj}^{\text{pre-shock}}}$ is the change in the average stock of operational robots per worker in industry j in other countries (Japan and South Korea), between year $t-1$ and year $t-4$. The intuition is to exploit industry-specific trajectories in automation that are driven by technological innovations shared across countries.

3.5 Empirical analysis

The empirical analysis in this study seeks to quantify the impact of technological change, specifically automation, on the occupational composition of political candidates. To do so, I estimate a model that links county-level exposure to automation with the likelihood that candidates from various occupational backgrounds run for office.

The model used to estimate the relationship between automation exposure and the occupational background of political candidates is as follows:

$$Candidate\_background_{cit} = \alpha + \beta Robots_{ct} + \gamma \mathbf{X}_{cit} + \lambda_t + \lambda_c + \epsilon_{cit} \quad (1)$$

Where: i, indexes candidate, c indexes county and t years.

$Candidate\_background_{cit}$ represents the occupational background of candidate i in county c during year t. This is the dependent variable, capturing the occupational categories of political candidates (e.g., working-class, routinary occupation, etc.). $Robots_{ct}$ represents Robot adoption at the district level between year t-1 and t-4. The model includes controls for incumbancy status and type of lower-level office. The model doesn't include individual controls characterizing the candidate background such as gender or race because these would introduce post-treatment bias. λ_t are year fixed effects and λ_c are county-fixed effects. The year fixed effects control for any unobserved time-varying factors that might affect candidate selection, such as national political trends or major economic shocks. The county fixed effects account for any time-invariant differences across counties, such as geographical or historical factors that might influence the political landscape in a given county.

The model is estimated using Ordinary Least Squares (OLS) for baseline specifications, where I examine the relationship between automation exposure and the occupational background of the candidate. To account for potential endogeneity, I also employ a two-stage least squares (2SLS) estimator. The analysis includes robust standard errors, clustered at the county-year level, to account for any correlation in the error terms within county-year and given that exposure to automation varies at such level.

The dependent variable in this model is $Candidate\_background_{cit}$, which captures the occupational background of the candidate. To classify candidates, I use several approaches. First, I define a binary variable that takes the value of 1 if the candidate is from a working-class occupation and 0 otherwise. Next, I explore the automability of occupations by incorporating an index based on the work of Autor and Dorn (2013), which measures the intensity of routine and manual tasks in each occupation. Additionally, I include a Routine Task Intensity (RTI) index, based on Goos *et al.* (2014), to further assess the extent to which occupations are susceptible to automation due to their routine task content. I also categorize occupations according to the social class framework proposed by Oesch (2013), which distinguishes between different groups based on occupational characteristics. Lastly, I use a binary variable that takes the value of 1 if the occupation is considered routine (as per Oesch's granular classification) and 0 if it is non-routine.

3.6 Complementary evidence from survey data

To deepen the analysis and explore the factors influencing the decision to run for office, this study also leverages a representative survey of the general U.S. population. I use survey data to investigate the relationship between local-level exposure to automation and the likelihood of individuals from various groups running for political office at any level. I use data from the 2010–2022 Cooperative Election Studies (CES), a large, nationally representative internet survey of the American public, featuring thousands of respondents in each wave. The dataset includes 524,435 observations spanning the years 2010 to 2022.[3] The survey includes a question about respondents' individual political experience, specifically asking whether they have ever run for political office. Each year, respondents are posed the following question: "Have you ever run for elective office at any level of government (local, state, or federal)?" Responses to this question are coded as binary variables (0 = no, 1 = yes). Although only a small proportion of respondents (3.92%) report having run for office, the large sample size ensures sufficient variation for robust estimation.

I estimate exposure to automation for U.S. commuting zones (CZs) using the same methodology applied to California counties. The only difference is that, for the U.S. as a whole, variation in the robot shock is analyzed at the commuting zone level rather than the county level to ensure consistency across the entire country. Respondents' counties are thus mapped to their respective commuting zones, resulting 719 commuting zones

[3] In particular, the respondents are from the 2010, 2011, 2012, 2014, 2016, 2018, 2020, and 2022 waves, which include the item on political experience utilized in this analysis.

with an associated robot shock measure. To characterize the heterogeneous effects of the automation shock across different groups, I use educational attainment as a proxy for individuals' skills in the labor market. I classify respondents into four levels of educational attainment: no high school, high school, college, and higher graduate education. While this may be a relatively coarse measure of individual exposure to automation, this approach is justified by two key considerations. First, the survey lacks detailed occupational information, limiting more granular assessments of exposure. Second, previous educational attainment, unlike current occupation, is less likely to be influenced by automation and therefore serves as a more reliable moderator in survey data, minimizing concerns about post-treatment bias or on the fact that results may be driven by compositional effects across occupations. Although the heterogeneous effects of automation across broad skill levels may not fully align with the routine-biased technological change framework, this approach still provides valuable insights into skill-biased technological change and its implications for different segments of the labor market.

To estimate the impact of technological change on the propensity to run for office across different skill groups, I estimate the following equation:

$$Run\_office_{ict} = \alpha + \beta Robots_{ct} + \delta Robots_{ct} * Educ_{ict} + \phi Educ_{ict} + \gamma \mathbf{X}_{ict} + \lambda_t + \lambda_c + \epsilon_{ict} \quad (2)$$

Where: i indexes individual respondent, c indexes commuting zone and t years. The dependent variable, Run_office_{cit}, is the binary indicator of whether an individual has ever run for political office. $Robots_{ct}$ measures robot exposure at the commuting zone level, which is instrumented following the same approach used in the California analysis. $Educ_{cit}$ is a vector of dummy variables capturing educational attainment, while $\mathbf{X}_{cit}$ includes individual-level controls for gender, age, and race (classified into white, black, hispanic and other). The year and commuting zone fixed effects, λ_t and λ_c, play a critical role. They do not only account for time-invariant unobservables specific to the local area and time-varying general trends. Given that the dependent variable captures whether someone has *ever* run for office, these fixed effects are also essential to ensure that the estimated impact of the robot shock reflects changes in political experience of a certain type of individual *during* the treatment period. Without these fixed effects, such an interpretation would not be possible. Standard errors are clustered by commuting zone-year. Finally, I estimate a more saturated and arguably robust version of the model that includes interactions between the robot shock and other covariates, in addition to education, to mitigate potential bias arising from the omission of these interactions (Yzerbyt *et al.*, 2004).

Building on the idea that technological change influences the propensity to run for office by shifting resources across occupations - thereby altering individuals' socio-economic status- I conduct an analysis where I replace the political experience variable with a measure of the respondent's family financial situation. Specifically, this variable categorizes income into several brackets, ranging from less than $10,000 to over $150,000, capturing income levels from the poorest to the wealthiest. Given the anticipated distributional effects of technological change, the robot shock is expected to exacerbate income inequality, shifting individuals' socio-economic positions and, in turn, influencing their likelihood of becoming a political candidate.

A key limitation of the analysis estimating heterogeneous effects by skill level is that it does not fully capture the logic of routine-biased technological change. In particular, it fails to identify the effects on those most adversely affected by automation, such as individuals who often hold intermediate levels of education, like much of the population, but are employed in routine occupations. To address this, I conclude the empirical analysis by adopting a measure of individual exposure to automation based on demographic characteristics and commuting zone of residence, following the approach of Anelli *et al.* (2021).

This methodology uses historical labor market data from 1990, prior to the onset of the most recent wave of automation. Rather than relying on observed current occupations, I estimate the predicted probability of employment in different occupations using factors such as education, age, gender, race, and commuting zone of residence. These probabilities are derived from the 1990 U.S. Census data provided by IPUMS and then applied to CES respondents. By combining these predicted occupational probabilities with occupational-level measures of automatability, I construct a metric of "individual vulnerability to automation." This vulnerability metric is then multiplied by the level of robot adoption in each commuting zone and year, yielding a measure of "individual exposure to automation."

This approach improves on the original measure by Anelli *et al.* (2021) in three key ways. First, it draws on a larger estimation sample, based on the U.S. Census rather than labor force surveys, for predicting occupational probabilities. Second, it expands the set of individual-level demographic variables including race, a particularly relevant factor in the U.S. context, and uses more detailed information on local labor markets by relying on U.S. commuting zones. Third, it interacts individual-level automatability with local (CZ-level) exposure to robot adoption, rather than national-level trends, allowing for greater geographic precision.

This approach assigns higher automation exposure to individuals who, based on pre-automation labor market conditions, were more likely to be employed in automatable occupations. It captures the fact that automation has reduced certain job opportunities, potentially leading to unemployment or occupational shifts. Crucially, it allows me to estimate the effect of automation on political candidacy beyond what can be inferred from candidates' current occupations. In doing so, it helps rule out the possibility that automation has no impact on political representation aside from the broader labor market reconfiguration that may have already reshaped occupational structures.

Following Anelli *et al.* (2021), I define individual exposure to automation as:

$$\text{Individual Exposure}_{ict} = \Delta R_{ct} * \underbrace{\sum_j \widehat{Pr}(o_i = j | \text{age, gender, edu, race}, c) * \theta_j}_{\text{Individual Vulnerability}}$$

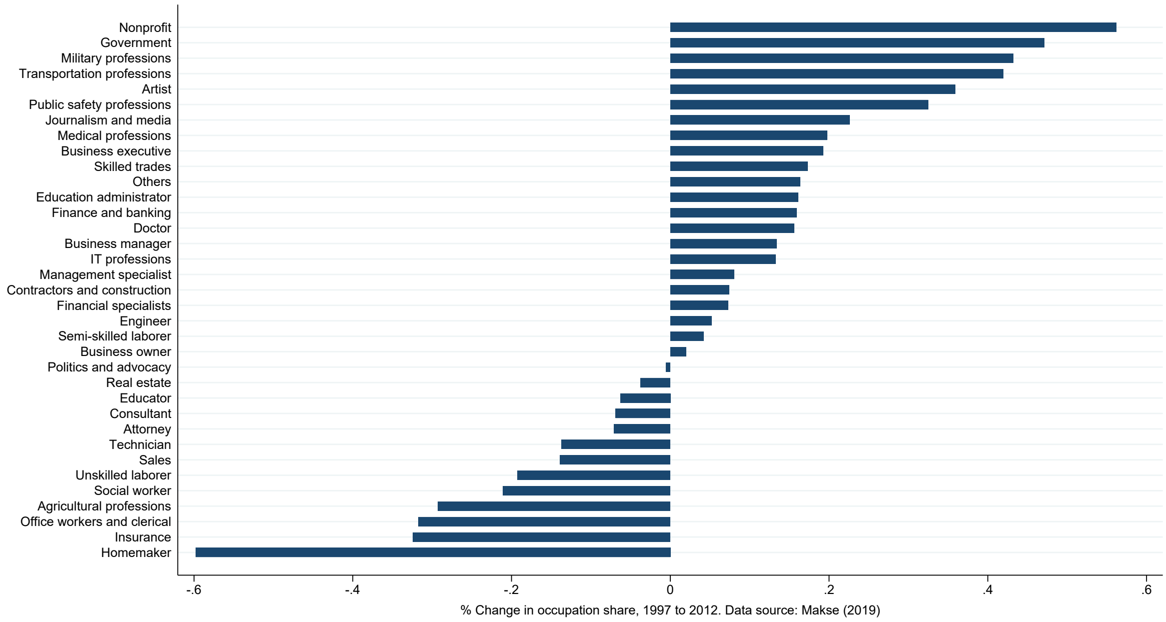
where: $\Delta R_{ct} = \sum_j \frac{L_{cj}^{\text{pre-sample}}}{L_c^{\text{pre-sample}}} * \frac{R_j^{t-1} - R_j^{t-n}}{L_j^{\text{pre-sample}}}$ represents CZ-level exposure to robot adoption over the past 4 years. The term $\widehat{Pr}(o_i = j | \text{age, gender, edu, race}, c)$ denotes individual i's predicted probability of working in occupation j, estimated using pre-shock data from the 1990 US census data. The parameter θ_j captures the automation threat associated to occupation j, based on the scores provided by Autor and Dorn (2013). I calculate these individual exposure scores for the CES respondents.

I now proceed to estimate the effect of individual exposure to automation on likelihood of having run for office. Specifically, I estimate the following equation:

$$Run\_office_{ict} = \alpha + \beta \ Individual \ Exposure_{ict} + \gamma \mathbf{X}_{ict} + \lambda_t + \lambda_c + \epsilon_{ict}$$

where i indexes individuals, c commuting zone, and t years. Individual Exposure$_{ict}$ represents our measure of individual exposure to automation. X_{it} is a vector of individual control variables, including age, gender, educational attainment and race. Standard errors are clustered by CZ-year. In the next section, I present the empirical results from the various analyses.

Figure 1: Change in occupation shares of state legislatures



4 Results

4.1 Descriptive analysis on state legislators

The empirical analysis begins with a descriptive examination of the changes in occupational backgrounds among U.S. state legislators from 1997 to 2012. Figure 1 presents the shift in occupation shares across various occupational categories within this elite group. It is evident that there have been significant changes, with some occupations becoming more prevalent, while others have experienced a decline. We observe a raise in legislators from non-profit, government, military, transportation, and artistic backgrounds. Conversely, occupations like homemaking, insurance, clerical work, and agriculture have witnessed a decline in representation. Some occupation experiencing a decline in their share can be traditionally seen as working-class, such as agricultural workers, unskilled laborers, and office workers, while roles in higher professional sectors, such as public safety professions, business executives, and nonprofit organizations, have gained ground.

However, while these descriptive results are insightful, they are limited by the nature of the state legislator dataset. The data on state legislators includes only elected representatives, meaning it captures only the final pool of candidates who successfully navigated the electoral process. This exclusion of candidates who did not make it past the screening stages or those who lost in the general election raises important questions about the extent to which the sample may reflect broader trends in the selection process. Additionally, the occupational categories in the state legislator dataset are broad and predefined, making it difficult to capture more granular occupational distinctions that could provide additional insights into how automation affects different types of workers.

To address these limitations, the next section of the analysis shifts focus to lower-level office candidates in California, spanning from 1998 to 2022. By examining these candidates, it is possible to overcome some of the constraints of the state legislator dataset and of most other datasets on politicians background in the US. Lower-level office candidates, as opposed to state legislators, represent a broader spectrum of potential political representatives and may offer a more comprehensive view of how automation and changes in occupational structures are shaping political representation. Moreover, because economic barriers to political participation tend to be lower at this level of office, analyzing this group of candidates offers a more nuanced understanding of the broader dynamics at play in political candidate selection.

4.2 Descriptive analysis on lower-level candidates in California

Figure 2 in this section provides a geographical perspective on the variation in candidates' occupational backgrounds across California. Using maps, we observe notable heterogeneity in the share of working-class candidates and the share of candidates from routine occupations, as well as differences in the average automability of candidates' occupations by county. This data focus in particular on candidates to city councils, although substantial heterogeneity by county also appears for other lower-level offices. Some counties, particularly in urban areas, exhibit a significantly lower proportion of working-class candidates, while rural counties tend to have a higher concentration of candidates from working-class backgrounds. Similarly, counties with a higher share of routine occupations tend to be more concentrated in rural regions. The significant level of heterogeneity remains both in the the period 1998-2001 and 2016-2021.

Figure 2: Occupational background by California Counties

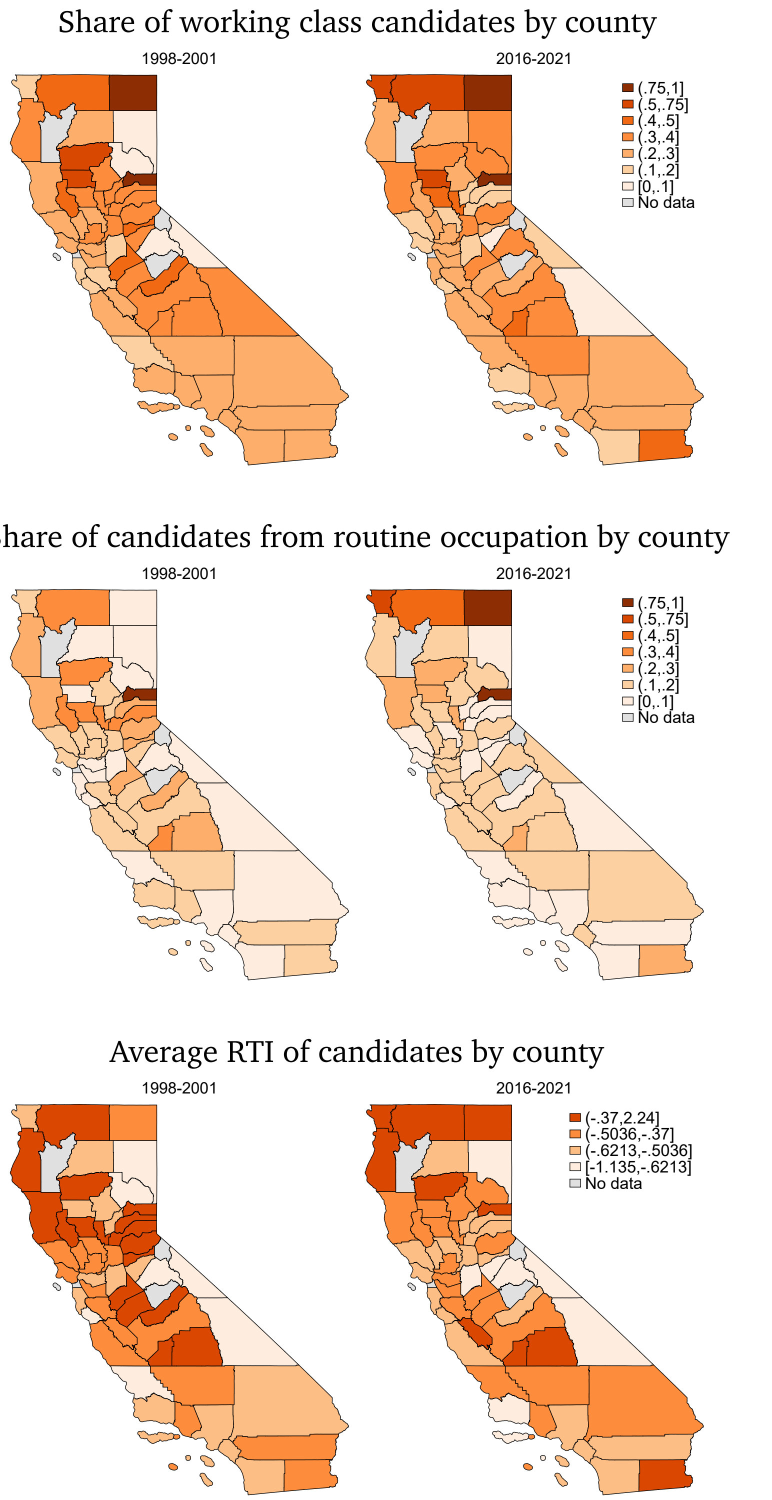
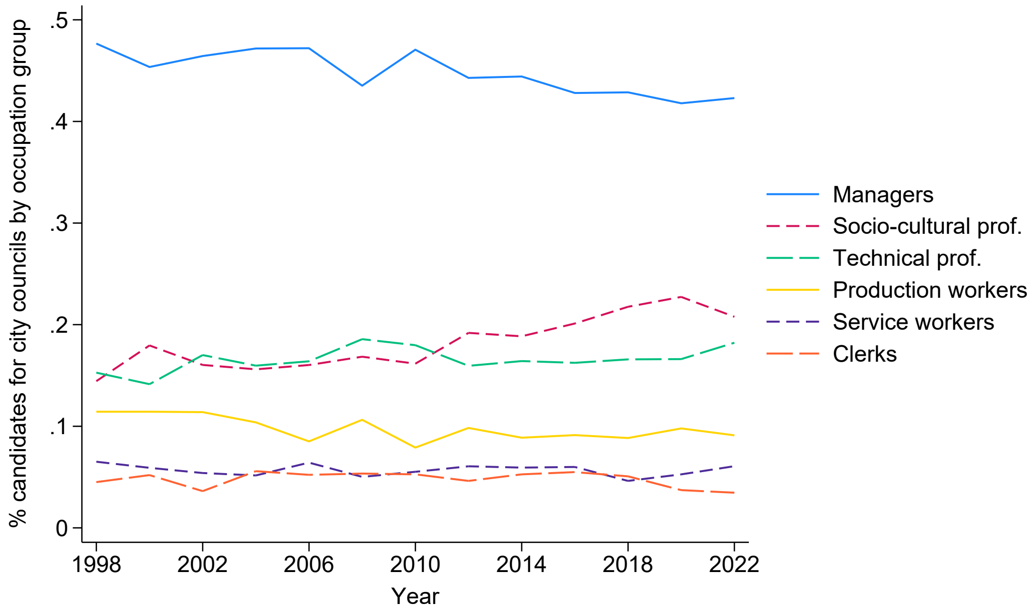


Figure 3 presents a graph depicting the trends in the share of candidates from various

Figure 3: Political candidates by social class

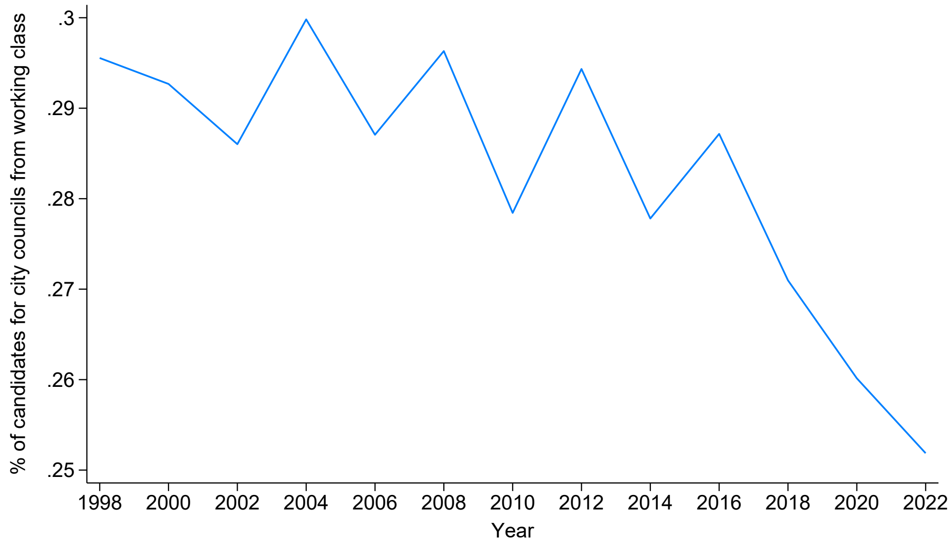


social classes, as defined by Oesch (2013). These social classes include managerial, sociocultural professionals, technical professionals, production workers, service workers, and clerks. The graph reveals some important trends over time. First, candidates from managerial occupations dominate the pool of lower-level office candidates, accounting for nearly 50% of all candidates. However, this share has slightly decreased over time. In contrast, there has been a substantial increase in the share of sociocultural professionals.

While there has been a slight increase in candidates from technical professions, their representation has remained relatively stable overall. On the other hand, the share of candidates from production worker backgrounds has decreased. Finally, the share of candidates from service and clerical occupations remains low but stable, with only minor fluctuations over the years. This pattern reflects the continued underrepresentation of individuals in lower-wage service sectors in political office, particularly at the city council level.

A further breakdown of the trends in the occupational backgrounds of California's lower-level office candidates is presented in Figure 4, which shows the decline in the share

Figure 4: Share of working class candidates

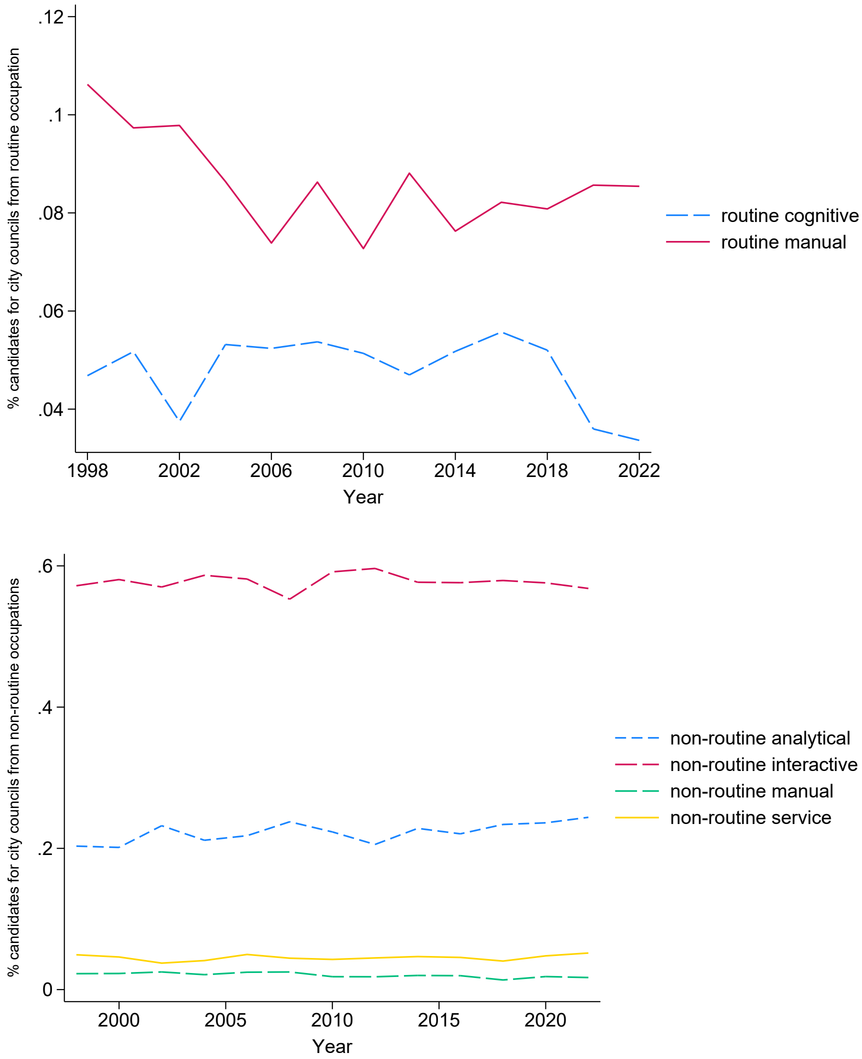


of working-class candidates over time. From around 29-30% in 1998, the proportion of working-class candidates has decreased to less than 26% by 2022. This suggests that individuals from manual or low-skilled occupations are becoming less likely to seek or succeed in running for office. This decline in working-class candidates may have significant implications for the representativeness of local governments, potentially diminishing the voices of traditionally underrepresented groups in political decision-making processes.

The two graphs in Figure 5 offer a closer look at the decline of routine occupations and the stability of non-routine occupations among lower-level office candidates. The first graph illustrates the decline in the share of candidates from routine occupations, especially those from routine manual jobs. Over time, these occupations have been increasingly underrepresented. Routine manual occupations, which are susceptible to automation, have seen the largest declines in candidate representation.

In contrast, the second graph illustrates the relative stability or slight increase in the share of candidates from non-routine occupations. Non-routine interactive occupations have remained stable in their share of political candidates, occupying nearly 60% of the total pool of candidates for city councils. This category, along that of non-routine interactive,

Figure 5: Share of candidates from routine and non-routine occupations



tipically includes individuals whose work involves problem-solving, communication, and interpersonal skills, and appear to be less vulnerable to the impacts of automation, as these occupations require human interaction and complex decision-making that cannot easily be replicated by machines.

4.3 The effect of exposure to automation on candidates' background

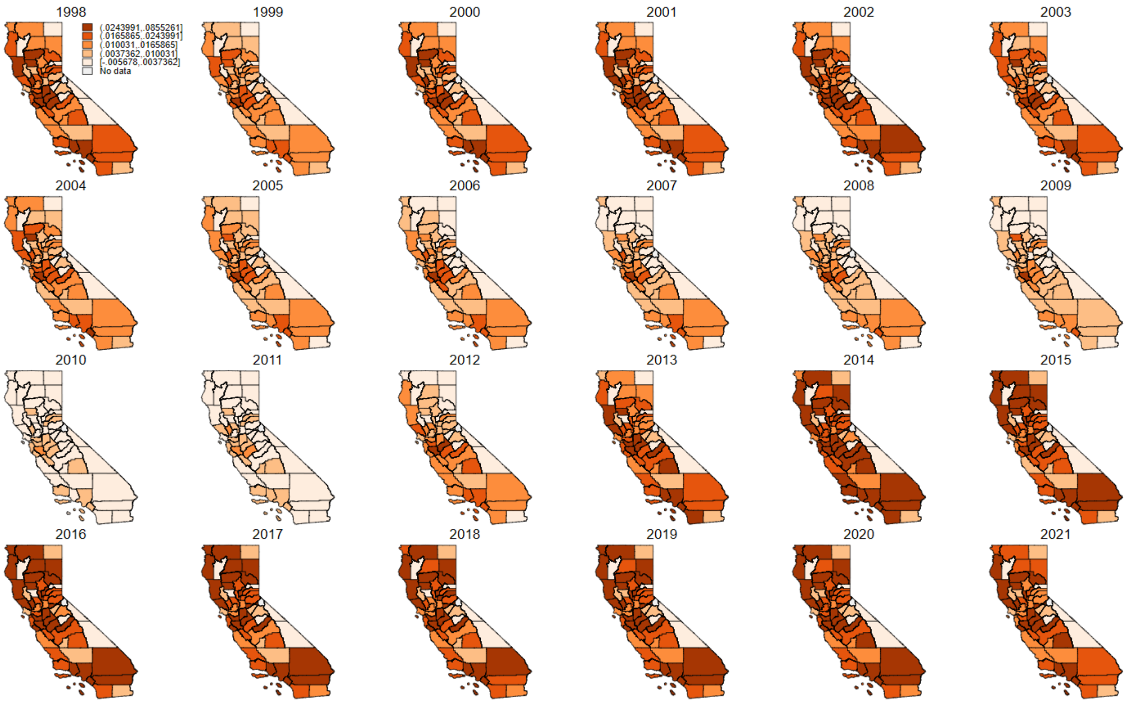
In this section, I examine the impact of county-level exposure to automation on the occupational background of candidates at lower-level offices in California. The focus is on understanding how different levels of exposure to automation—measured by a robot shock variable—affect the likelihood of observing candidates from the working class and those holding routine occupations.

Before diving into the regression analysis, it is important to highlight the variation in automation exposure across counties and how it evolves over time. Figure 6 presents a series of maps showing the exposure to automation for each California county from 1998 to 2021 (2022 values are imputed from 2021). The maps clearly illustrate that urban counties tend to experience higher exposure to automation compared to rural counties, reflecting a higher concentration of industries likely to be impacted by robots and automation technologies. However, there is notable heterogeneity even within urban areas, as some counties experience more rapid technological adoption than others. Additionally, exposure to automation fluctuates over time. This temporal and regional variation will be crucial as we move into the regression analysis, where I examine how these county-specific shocks explains candidate occupations.

Table 1 presents the main regression results examining the effect of county-level exposure to automation on the likelihood of having working-class candidates. I estimate the relationship using both Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS) estimations to address potential endogeneity concerns. I estimate four models in total: two OLS models (one with only year fixed effects and the other with both county and year fixed effects) and two 2SLS models using an instrumental variable for automation exposure.

The OLS results show that areas with greater exposure to automation tend to have a lower likelihood of electing working-class candidates. Specifically, an increase of one standard deviation in the robot shock is associated with a 4% decrease in the probability

Figure 6: Robot shock by county over time



of having a working-class candidate in the model specification with county and year fixed effects with 2SLS estimator. This result is statistically significant at the 5% level. This suggests that counties with greater exposure to automation are less likely to observe candidates from the working class.

Table 2 explores the automability of candidates' occupations, measured using two indices: the Routine+Manual Task Intensity by Autor and Dorn (2013) index and the Routine Task Intensity index of Goos *et al.* (2014). The automability of an occupation refers to the extent to which its tasks can be replaced by automation or robots.

The results show that, in counties with greater exposure to automation, candidates tend to have occupations with lower automability scores. For example, an increase in one standard deviation of robot shock is associated with a 0.05 decrease in the Routine+Manual Task Intensity index and a 0.66 decrease in the RTI index. Results are statistically significant at least at the 10% level for the first index and at the 5% level in the second. These findings

Table 1: Working class background and automation

VARIABLES	(1) Work class	(2) Work class	(3) Work class	(4) Work class
Robot shock	-0.022***	-0.001	-0.039***	-0.040**
	[0.003]	[0.004]	[0.005]	[0.020]
Observations	81,282	81,282	81,282	81,282
R-squared	0.018	0.034	0.016	0.013
Estimator	OLS	OLS	2SLS	2SLS
Controls	X	X	X	X
Year FE	X	X	X	X
County FE		X		X
rkf			246.7	23.24

Standard errors clustered by county-year
*** p<0.01, ** p<0.05, * p<0.1

suggest that, as counties experience more exposure to automation, the pool of candidates increasingly comes from occupations less susceptible to automation. This result aligns with the expectation that automation drives a shift in the types of occupations represented in local politics. Specifically, as automation impacts manual and routine occupations, candidates from these backgrounds may become less common. Conversely, candidates from occupations with lower automability are more likely to appear.

Tables A.1 and A.2 in the Appendix present results examining the effect of robot exposure on candidates from different social classes based on Oesch's classification, which includes categories such as production workers, clerks, service workers, managers, sociocultural professionals, and technical professionals. In the model with OLS estimator and year fixed effects of Table A.1, results show that in counties with greater exposure to automation, there is a noticeable shift in the composition of candidates, particularly a decline in production workers, clerks, and service workers, and an increase in sociocultural professionals, managers, and semi-technical professionals. More robuts models with 2SLS estimates from Table A.2, however, are typically not robust across the board.

Finally, Table 3 focuses on the binary outcome of whether a candidate comes from a

Table 2: Routine tasks and automation

VARIABLES	(1) Rout.-man.	(2) Rout.-man.	(3) RTI	(4) RTI
Robot shock	-0.045***	-0.050*	-0.033***	-0.066**
	[0.008]	[0.030]	[0.007]	[0.031]
Observations	78,128	78,128	70,683	70,683
R-squared	0.013	0.012	0.005	0.003
Estimator	2SLS	2SLS	2SLS	2SLS
Controls	X	X	X	X
Year FE	X	X	X	X
County FE		X		X
rkf	249.6	22.95	247.0	22.59

Standard errors clustered by county-year
*** p<0.01, ** p<0.05, * p<0.1

routine occupation (1 if routine, 0 otherwise), according to Oesch's classification. The results show a clear negative relationship: counties with higher exposure to robot shocks are less likely to field candidates from routine occupations. Specifically, an increase in one standard deviation of robot shock is associated with a 3.2 percentage point decrease in the probability of a candidate coming from a routine occupation. This result is statistically significant at the 5% level. This finding reinforces the earlier results and highlights that automation is likely reshaping the pool of candidates, driving a decrease in the representation of those from routine occupations.

4.4 Evidence from survey data

This study highlights the importance of focusing on early selection into the political process, specifically through access to lower-level offices. These positions serve as initial indicators of political ambition and capacity, traits that are also essential for advancing to higher stages of political careers. While the proportion of individuals pursuing higher office in the United States is undoubtedly small, a significant number of people, at some point in their lives, choose to enter the political arena by running for an office—whether it be a school or

Table 3: Routinary occupations and automation

VARIABLES	(1) Routine (Oesch)	(2) Routine (Oesch)
Robot shock	-0.029*** [0.004]	-0.032** [0.016]
Observations	81,211	81,211
R-squared	0.011	0.007
Estimator	2SLS	2SLS
Controls	X	X
Year FE	X	X
County FE		X
rkf	246.6	23.46

Standard errors clustered by county-year
*** p<0.01, ** p<0.05, * p<0.1

local board, a city council, or similar roles.

Survey data from the CES enables an analysis of the impact of robotization on individuals' propensity to run for office (i.e., political candidancy) across different skill levels. Table 4 presents the results of the model in which individual candidacy experience at any level is regressed on commuting zone exposure to robot adoption, interacted with educational attainment. Robot adoption is instrumented using robot adoption in other advanced economies, and the model controls for educational attainment, gender, age, and race, in addition to country-year and commuting zone fixed effects. Column (2) further includes full interactions of the robot shock with other individual covariates.

The results indicate significant polarization in political candidacy across skill levels. Specifically, higher exposure to technological change is associated with a decrease in political candidacy among individuals without a high school diploma, while the interactions with higher educational attainment are positive and increase as education level rises. On average, individuals with a college degree are more likely to run for office, but this gap widens as exposure to robotization increases. The interaction terms between robot adoption and education level are all statistically significant at the 95% level in the more robust model, which include full interactions (column 2). The full results are displayed in Table A.3

To better understand the magnitude of the effect of robotization across different education levels, I estimate the marginal effects, which are shown in Figure 7. The marginal effect of exposure to robotization is negative and statistically significant for individuals at the lower end of the skill distribution, while it is positive and significant for those with the highest educational attainment. Therefore, in areas with higher exposure to robot adoption, there is increasing polarization in political candidacy at the extremes of the skill distribution, whereas individuals in the middle—representing the majority of the population—are largely unaffected. Overall, this evidence demonstrates how technological change can exacerbate inequality in political participation.

The distributional effects of technological change, particularly robotization, can shape the propensity of different social groups to run for office by influencing material conditions and socioeconomic status. As highlighted by Carnes (2020), the decision to pursue a political career or run for office can be especially burdensome—and often appears unfeasible—for individuals with limited resources. Consistent with the literature on political inequality and participation (Ansell and Gingrich, 2024; Beramendi *et al.*, 2024), the CES data reveal a positive association between higher income and political candidacy experience. Evidence from a model in which political candidacy experience is regressed on income levels—controlling for individual demographics, CZs and year fixed effects- shows that the likelihood of running for office increases significantly as income rises (see Table A.4).

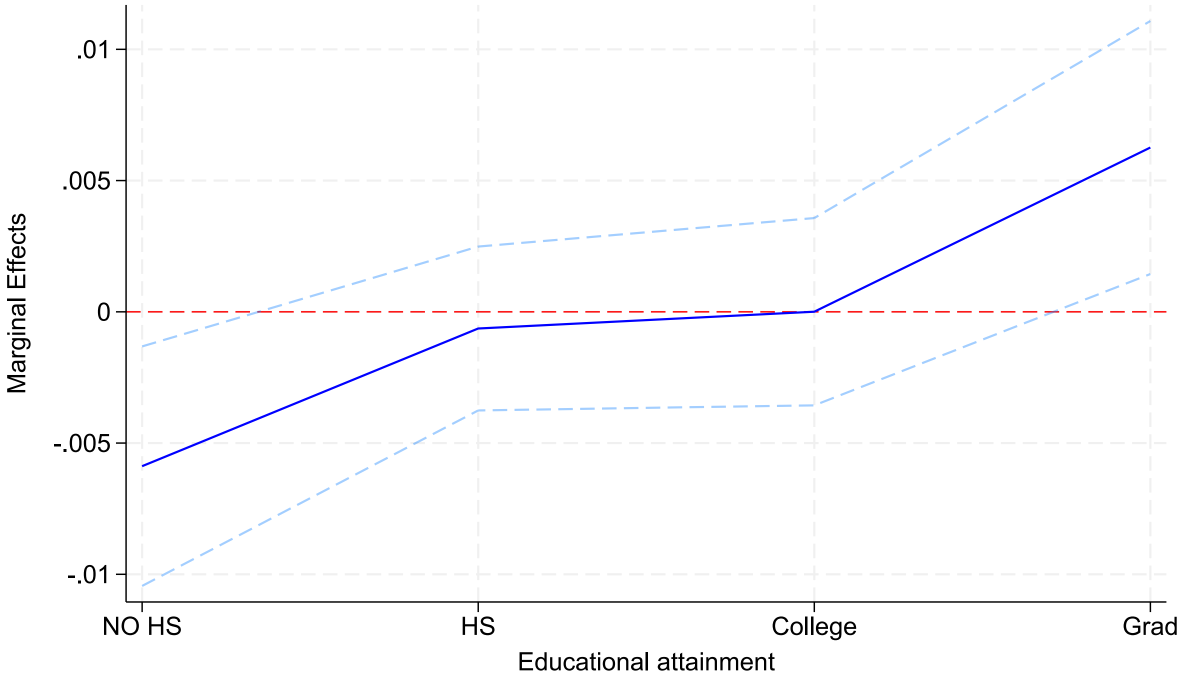
To provide additional evidence on the mechanisms through which automation impacts political participation, I re-estimate the model on the impact of the robot shock, now with family income as the dependent variable. Figure 8 illustrates the magnitude and statistical significance of the marginal effects of robotization for different skill levels. The results show that the same skill levels for which automation influences political candidacy are also affected in terms of income. Specifically, in areas with higher exposure to robotization, automation reduces the income of individuals at the lower end of the skill distribution,

Table 4: Political candidancy and exposure to automation (CES)

VARIABLES	(1) Run office	(2) Run office
Robot shock	-0.005*	-0.012***
	[0.003]	[0.004]
Robot shock * HS	0.005*	0.005**
	[0.002]	[0.002]
Robot shock * College	0.005*	0.006**
	[0.003]	[0.003]
Robot shock * Grad	0.012***	0.012***
	[0.003]	[0.003]
HS	0.007***	0.008***
	[0.002]	[0.002]
College	0.023***	0.023***
	[0.002]	[0.002]
Grad	0.041***	0.041***
	[0.002]	[0.002]
Female	-0.027***	-0.026***
	[0.001]	[0.001]
Age	0.000***	0.000***
	[0.000]	[0.000]
Observations	364,176	364,176
R-squared	0.012	0.012
Estimator	2SLS	2SLS
Controls	X	X
Full inter.		X
Year FE	X	X
Czone FE	X	X
rkf	34.94	15.74

Standard errors clustered by commuting zone-year.
*** p<0.01, ** p<0.05, * p<0.1

Figure 7: Marginal Effects of robot shock on political candidancy by educational attainment

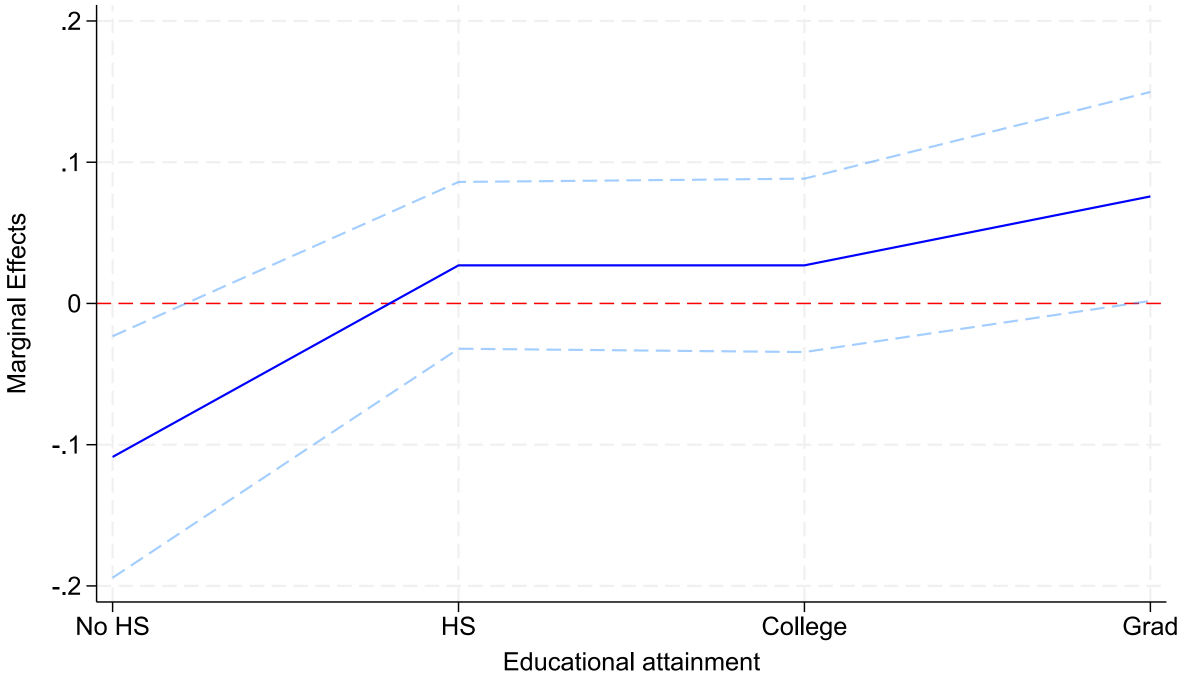


Note: The plot illustrates the marginal effects of the robot shock on the propensity to run across various levels of educational attainment, accompanied by 90% confidence intervals. The marginal effect at the lowest and highest educational attainment would also appear significant with 95% CI. The model controls for educational attainment, gender, age, race, and includes fixed effects for year and commuting zone. It also accounts for interactions between the robot shock and individual covariates. The analysis uses a 2SLS estimator, with the robot shock instrumented by robot adoption in other advanced economies, particularly South Korea and Japan.

while increasing the income of the most educated individuals. The marginal effects are statistically significant at the 95% level for respondents without a high school diploma, and at the 90% level for those with education beyond college, thus individuals at the extreme of the skill distribution.

Interestingly, automation does not appear to affect the propensity to run for office among individuals with intermediate levels of educational attainment. This may suggest that the impact of robotization on the middle segment of the distribution is likely to be shaped by heterogeneity across other dimensions, such as routine task intensity, which cannot be fully captured by the skill dimension alone.

Figure 8: Marginal Effects of robot shock on family income by educational attainment



*Note:*The plot illustrates the marginal effects of the robot shock on family income across various levels of educational attainment, accompanied by 90% confidence intervals. The marginal effect of the lowest educational attainment is significant at the 95% CI. The model controls for educational attainment, gender, age, race, and includes fixed effects for year and commuting zone. It also accounts for interactions between the robot shock and individual covariates. The analysis uses a 2SLS estimator, with the robot shock instrumented by robot adoption in other advanced economies, particularly South Korea and Japan. Full results are shown in Table A.5.

4.5 Individual level analysis based on counterfactual automability

A major difficulty in analyzing the effects of automation on political candidacy is that the current occupations of candidates may not reflect individuals' true exposure. Automation reshapes the labor market over time, displacing some workers and steering others into less automatable roles. As a result, individuals who appear to face low exposure today may have been affected by earlier waves of displacement or may have never accessed the occupations most at risk. Relying solely on present occupation risks overlooking precisely those individuals most impacted by technological change.

Table 5: Effects of individual exposure to automation (CES)

	(1)	(2)	(3)	(4)
VARIABLES	Run Office		Fam. income	
Individual Exposure	-0.081***	-0.129***	-2.712***	-3.714***
	[0.013]	[0.020]	[0.154]	[0.282]
Education	0.008***	0.007***	0.741***	0.720***
	[0.000]	[0.000]	[0.006]	[0.007]
Gender	-0.032***	-0.033***	-0.783***	-0.798***
	[0.001]	[0.001]	[0.009]	[0.009]
Age	0.001***	0.001***	0.011***	0.010***
	[0.000]	[0.000]	[0.000]	[0.000]
Estimator	OLS	2SLS	OLS	2SLS
Observations	364,316	364,316	547,117	547,117
Year FE	X	X	X	X
Czone FE	X	X	X	X
Std dev. Y	0.191	0.191	2.859	2.859
Std dev. X	0.0331	0.0331	0.0331	0.0331
Magnitude	-0.0139	-0.0223	-0.0314	-0.0430
R-squared	0.029	0.017	0.212	0.163
Kleibergen-Paap F-stat		1318		1567

Standard errors clustered by commuting zone-year
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

To address this issue, I employ a counterfactual measure of exposure that captures latent vulnerability based on pre-automation labor market conditions, following Anelli *et al.* (2021). An additional advantage of this approach is that, unlike the analysis of heterogeneous effects by skill level, it captures variation in exposure within local areas and among individuals with similar levels of education. This is particularly important for identifying differences within the middle class, especially among those with intermediate educational attainment, but who may face very different levels of automation risk.

Table 5 reports the results of the regression analysis examining the relationship between individual exposure to automation and political candidacy experience. Columns 1–2 show that the coefficient on individual exposure to automation is negative and statistically significant in both the OLS and 2SLS specifications. In the instrumental variable model, a one standard deviation increase in exposure is associated with a 2 percentage point reduction in the probability of having run for office, after controlling for fixed effects and covariates. This effect, while modest in size, is substantively meaningful, especially when considering the possibility of cumulative impacts over time and across persistently exposed areas. These results suggest that automation contributes to reducing the political participation of individuals who are adversely affected by technological change, not based on their current occupation but rather on their latent vulnerability as defined by pre-automation labor market conditions.

Columns 3–4 turn to family income as the outcome. In the IV model, a one standard deviation increase in individual exposure is associated with a 4% of a standard deviation decrease in family yearly income (approximately $1,000), after accounting for controls and fixed effects. The negative and statistically significant relationship confirms that individuals most exposed to automation also experience a material decline in economic well-being. Taken together, these findings indicate that automation is linked to a decline in both political and economic resources for a segment of the population, i.e. those disproportionately exposed to technological displacement, highlighting the presence of a group that is simultaneously losing ground in both economic and political spheres.

5 Conclusion

This paper has examined the impact of technological change, particularly automation, on the composition of political candidates, focusing on how these changes affect the socio-economic backgrounds of individuals who enter politics. The findings suggest a significant transformation in the occupational background of political candidates, with counties exposed to higher levels of automation showing a marked decline in working-class candidates and those from routine occupations. These results support the hypothesis that technological change, through its impact on labor markets, also shapes political representation by influencing which groups are more likely to pursue political office.

The evidence from the regression analysis suggests that as automation increases in an electoral district, the likelihood of candidates from working-class or routine occupations diminishes. This shift suggests that automation not only disrupts labor markets but also impacts the pool of political candidates. Specifically, counties that experience higher levels of automation tend to see a reduced representation of working-class candidates, who are typically more exposed to automation-induced job losses, while those from higher-skilled and less automatable occupations are more likely to enter the political arena.

These findings highlight an important dimension of the intersection between technological change, labor markets, and political representation. They suggest that the winners of automation, those who are not as vulnerable to technological disruption, are more likely to participate in politics, while the losers—those who see their economic prospects diminished—are less likely to run for office. This dynamic exacerbates existing inequalities in political representation, as it leads to a political elite that is increasingly composed of individuals from higher socio-economic backgrounds, who are better positioned to take advantage of the opportunities created by automation.

This study contributes to the literature on political recruitment and representation by offering a structural explanation for the changing composition of political elites. It also emphasizes the broader implications of technological change on democratic systems, highlighting how shifts in the labor market can influence the diversity of political candidates and, ultimately, the interests represented in policymaking. As political elites become more homogenous in terms of their socio-economic background, the risk grows of a disconnect between the interests of the political class and those of the broader electorate, especially the working-class and routine workers who face greater economic insecurity due to automation.

The implications of these findings are significant. In an era of rapid technological advancement, understanding how automation reshapes the political candidate pool is essential for ensuring a more representative and inclusive political system. As automation continues to evolve and expand, future research should explore how other forms of technological change—such as artificial intelligence and green transitions—affect political participation and the diversity of political elites. This research not only provides a novel perspective on the transformation of political elites but also serves as a crucial step in understanding the broader implications of technological progress for democratic governance.

References

ACEMOGLU, D. and AUTOR, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, vol. 4, Elsevier, pp. 1043–1171.

— and RESTREPO, P. (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy*, **128** (6), 2188–2244.

ANELLI, M., COLANTONE, I. and STANIG, P. (2021). Individual vulnerability to industrial robot adoption increases support for the radical right. *Proceedings of the National Academy of Sciences*, **118** (47).

—, GIUNTELLA, O. and STELLA, L. (2024). Robots, marriageable men, family, and fertility. *Journal of Human Resources*, **59** (2), 443–469.

ANSELL, B. and GINGRICH, J. (2024). Political inequality. *Oxford Open Economics*, **3** (S1).

AUTOR, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, **344** (6186), 843–851.

— and DORN, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, **103** (5), 1553–1597.

—, KATZ, L. F. and KEARNEY, M. (2006). Measuring and interpreting trends in economic inequality. In *AEA Papers and Proceedings*, vol. 96, pp. 189–194.

—, LEVY, F. and MURNANE, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, **118** (4), 1279–1333.

BERAMENDI, P. (2015). *The politics of advanced capitalism*. Cambridge University Press.

—, BESLEY, T. and LEVI, M. (2024). Political equality: what is it and why does it matter? *Oxford Open Economics*, **3** (Supplement_1), i262–i281.

BORCHERT, J. (2003). Professional politicians: Towards a comparative perspective. *The political class in advanced democracies*, pp. 1–25.

BURDEN, B. C. (2015). *Personal roots of representation*. Princeton University Press.

CARNES, N. (2020). *The cash ceiling: why only the rich run for office–and what we can do about it*. Princeton University Press.

— and LUPU, N. (2023). The economic backgrounds of politicians. *Annual Review of Political Science*, **26** (Volume 26, 2023), 253–270.

CASELLI, F. and MORELLI, M. (2004). Bad politicians. *Journal of public economics*, **88** (3-4), 759–782.

COLANTONE, I. and STANIG, P. (2018). Global competition and brexit. *American political science review*, **112** (2), 201–218.

DAL BÓ, E., DAL BÓ, P. and SNYDER, J. (2009). Political dynasties. *The Review of Economic Studies*, **76** (1), 115–142.

—, FINAN, F., FOLKE, O., PERSSON, T. and RICKNE, J. (2017). Who becomes a politician? *The Quarterly Journal of Economics*, **132** (4), 1877–1914.

— and ROSSI, M. A. (2011). Term length and the effort of politicians. *The Review of Economic Studies*, **78** (4), 1237–1263.

DANCYGIER, R., LINDGREN, K.-O., NYMAN, P. and VERNBY, K. (2021). Candidate supply is not a barrier to immigrant representation: A case–control study. *American Journal of Political Science*, **65** (3), 683–698.

EVANS, G. and TILLEY, J. (2017). *The new politics of class: The political exclusion of the British working class*. Oxford University Press.

FERRAZ, C. and FINAN, F. (2009). *Motivating politicians: The impacts of monetary incentives on quality and performance*. Tech. rep., National Bureau of Economic Research.

FOLKE, O. and RICKNE, J. (2024). The class ceiling in politics. *American Political Science Review*, pp. 1–18.

FOX, R. L. and LAWLESS, J. L. (2005). To run or not to run for office: Explaining nascent political ambition. *American journal of political science*, **49** (3), 642–659.

GAGLIARDUCCI, S. and PASERMAN, M. D. (2012). Gender interactions within hierarchies: evidence from the political arena. *The Review of Economic Studies*, **79** (3), 1021–1052.

GALASSO, V. and NANNICINI, T. (2011). Competing on good politicians. *American political science review*, **105** (1), 79–99.

GOLDIN, C. and KATZ, L. F. (2007). Long-run changes in the wage structure: Narrowing, widening, polarising. *Brookings Papers on Economic Activity*, **38**, 135–168.

GOOS, M. and MANNING, A. (2007). Lousy and lovely jobs: The rising polarization of work in britain. *The review of economics and statistics*, **89** (1), 118–133.

—, — and SALOMONS, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American economic review*, **104** (8), 2509–2526.

GULZAR, S. (2021). Who enters politics and why? *Annual Review of Political Science*, **24** (1), 253–275.

KATZ, L. F. and MURPHY, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The quarterly journal of economics*, **107** (1), 35–78.

KITSCHELT, H. and REHM, P. (2014). Occupations as a site of political preference formation. *Comparative Political Studies*, **47** (12), 1670–1706.

KITSCHELT, H. P. and REHM, P. (2023). Polarity reversal: The socioeconomic reconfiguration of partisan support in knowledge societies. *Politics & Society*, **51** (4), 520–566.

LAMPRINAKOU, C., MORUCCI, M., CAMPBELL, R. and VAN HEERDE-HUDSON, J. (2017). All change in the house? the profile of candidates and mps in the 2015 british general election. *Parliamentary Affairs*, **70** (2), 207–232.

MACKENZIE, S. A. (2015). Life before congress: using precongressional experience to assess competing explanations for political professionalism. *The Journal of Politics*, **77** (2), 505–518.

MAKSE, T. (2019). Professional backgrounds in state legislatures, 1993–2012. *State Politics & Policy Quarterly*, **19** (3), 312–333.

MERLO, A., GALASSO, V., LANDI, M. and MATTOZZI, A. (2010). The labor market of italian politicians.(2010). *The Ruling Class: Management and Politics in Modern Italy*, pp. 7–95.

NORRIS, P. and LOVENDUSKI, J. (1993). 'if only more candidates came forward': Supply-side explanations of candidate selection in britain. *British Journal of Political Science*, **23** (3), 373–408.

OESCH, D. (2013). *Occupational change in Europe: how technology and education transform the job structure*. Oxford University Press.

O'GRADY, T. (2019). Careerists versus coal-miners: welfare reforms and the substantive representation of social groups in the british labour party. *Comparative Political Studies*, **52** (4), 544–578.

YZERBYT, V. Y., MULLER, D. and JUDD, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of experimental social psychology*, **40** (3), 424–431.

The Candidate Factory: technological change and political supply

Appendix

Paolo Agnolin

A Additional analyses

Figure 9: Number of candidates to lower-level offices by county

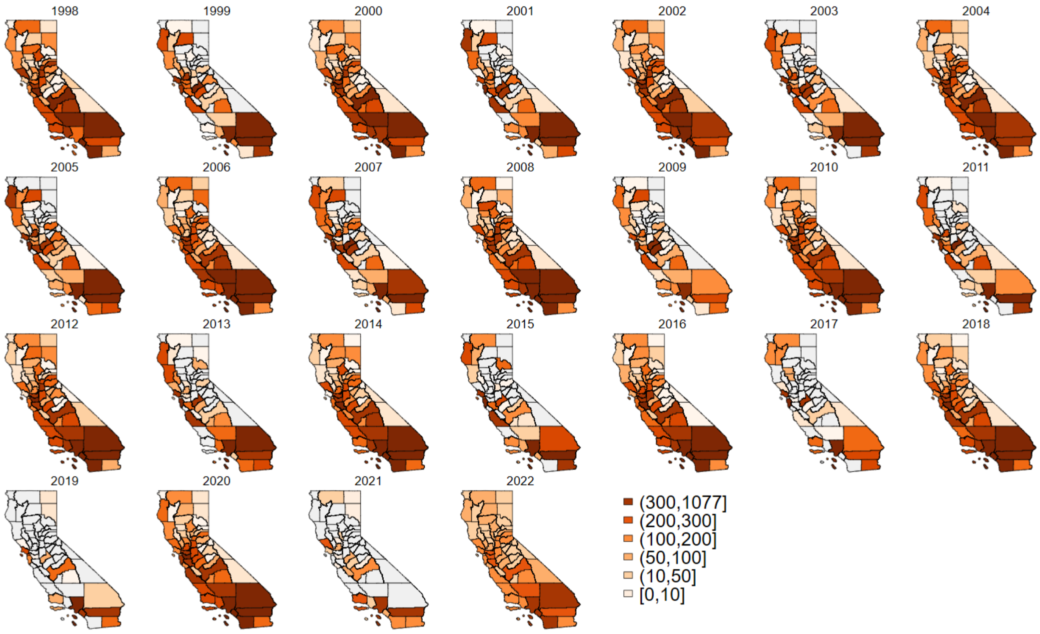


Table A.1: Occupation groups and automation (OLS)

VARIABLES	(1) Tech prof	(2) Product	(3) Managers	(4) Clerks	(5) Soc-cul prof	(6) Service
Robot shock	0.020***	-0.025***	0.018***	-0.005***	0.002	-0.010***
	[0.004]	[0.004]	[0.005]	[0.002]	[0.003]	[0.002]
Observations	81,549	81,549	81,549	81,549	81,549	81,549
R-squared	0.011	0.011	0.074	0.004	0.039	0.009
Estimator	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Controls	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
County FE						
rkf	246.7	246.7	246.7	246.7	246.7	246.7

Standard errors clustered by county-year
*** p<0.01, ** p<0.05, * p<0.1

Table A.2: Occupation groups and automation (IV)

VARIABLES	(1) Tech prof	(2) Product	(3) Managers	(4) Clerks	(5) Soc-cul prof	(6) Service
Robot shock	0.016	-0.018	0.012	-0.015*	0.013	-0.004
	[0.018]	[0.017]	[0.021]	[0.008]	[0.016]	[0.010]
Observations	81,549	81,549	81,549	81,549	81,549	81,549
R-squared	0.009	0.009	0.071	0.003	0.040	0.008
Estimator	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Controls	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Ccounty FE	X	X	X	X	X	X
rkf	23.42	23.42	23.42	23.42	23.42	23.42

Standard errors clustered by county-year
*** p<0.01, ** p<0.05, * p<0.1

Table A.3: Political candidancy and exposure to automation (CES)

VARIABLES	(1) Run office	(2) Run office
Robot shock	-0.005*	-0.012***
	[0.003]	[0.004]
Robot shock * HS	0.005*	0.005**
	[0.002]	[0.002]
Robot shock * College	0.005*	0.006**
	[0.003]	[0.003]
Robot shock * Grad	0.012***	0.012***
	[0.003]	[0.003]
Robot shock * female		-0.003**
		[0.001]
Robot shock * age		0.000***
		[0.000]
Robot shock * black		0.001
		[0.002]
Robot shock * hispanic		0.003
		[0.004]
Robot shock * other		0.002
		[0.003]
HS	0.007***	0.008***
	[0.002]	[0.002]
College	0.023***	0.023***
	[0.002]	[0.002]
Grad	0.041***	0.041***
	[0.002]	[0.002]
Female	-0.027***	-0.026***
	[0.001]	[0.001]
Age	0.000***	0.000***
	[0.000]	[0.000]
Black	-0.001	-0.001
	[0.002]	[0.002]
Hispanic	0.001	0.001
	[0.002]	[0.002]
Other	-0.000	-0.000
	[0.002]	[0.002]
Observations	364,176	364,176
R-squared	0.012	0.012
Estimator	2SLS	2SLS
Controls	X	X
Full inter.		X
YEAR FE	X	X
CZONE FE	X	X
rkf	34.94	15.74

Standard errors clustered by commuting zone-year
*** p<0.01, ** p<0.05, * p<0.1

Table A.4: Political candidancy and family income

VARIABLES	(1) Run office
Income levels:	
Less than 10k	-0.000
	[0.003]
10k - 20k	-0.011***
	[0.002]
20k - 30k	-0.010***
	[0.002]
30k - 40k	-0.009***
	[0.002]
40k - 50k	-0.006***
	[0.002]
50k - 60k	-0.000
	[0.002]
60k - 70k	-0.003
	[0.002]
70k - 80k	0.007***
	[0.002]
80k - 100k	0.008***
	[0.002]
100k - 120k	0.012***
	[0.002]
120k - 150k	0.011***
	[0.002]
150k+	0.017***
	[0.002]
Observations	365,827
R-squared	0.023
Estimator	OLS
Controls	X
YEAR FE	X
CZONE FE	X

Std errors clustered by CZ-year.
Ind. controls:gender, age and race.
Baseline income category represents non-respondents to item on income.
*** p<0.01, ** p<0.05, * p<0.1

Table A.5: Family income and exposure to automation (CES)

VARIABLES	(1) Fam. income	(2) Fam. income
Robot shock	-0.122**	-0.056
	[0.051]	[0.064]
Robot shock * HS	0.140***	0.136***
	[0.041]	[0.041]
Robot shock * College	0.143***	0.136***
	[0.044]	[0.044]
Robot shock * Grad	0.179***	0.184***
	[0.049]	[0.049]
Robot shock * female		0.058***
		[0.018]
Robot shock * age		-0.002***
		[0.001]
Robot shock * black		-0.041
		[0.031]
Robot shock * hispanic		0.073
		[0.059]
Robot shock * other		0.028
		[0.053]
HS	1.296***	1.295***
	[0.029]	[0.029]
College	2.834***	2.832***
	[0.031]	[0.031]
Grad	4.355***	4.354***
	[0.035]	[0.035]
Female	-0.666***	-0.667***
	[0.011]	[0.012]
Age	0.009***	0.009***
	[0.000]	[0.000]
Black	-1.245***	-1.246***
	[0.023]	[0.023]
Hispanic	-0.644***	-0.625***
	[0.030]	[0.033]
Other	-0.301***	-0.296***
	[0.028]	[0.028]
Observations	546,991	546,991
R-squared	0.174	0.174
Estimator	2SLS	2SLS
Controls	X	X
Full inter.		X
YEAR FE	X	X
CZONE FE	X	X
rkf	45.98	20.45

Standard errors clustered by commuting zone-year.
*** p<0.01, ** p<0.05, * p<0.1

Figure 10: Number of candidates to city councils by county

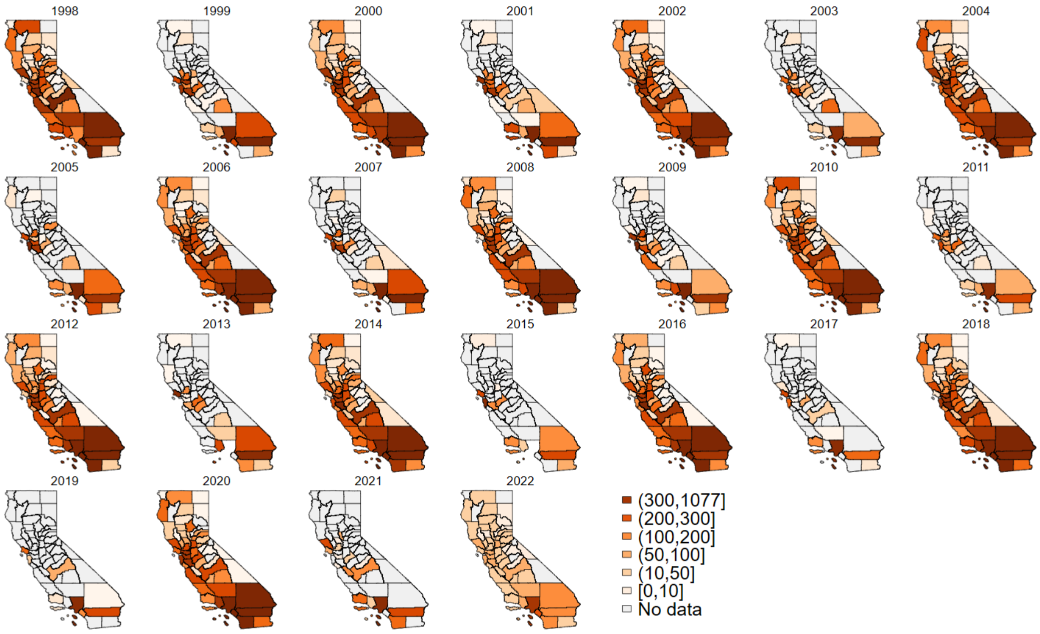


Figure 11: Share of working class candidates by county

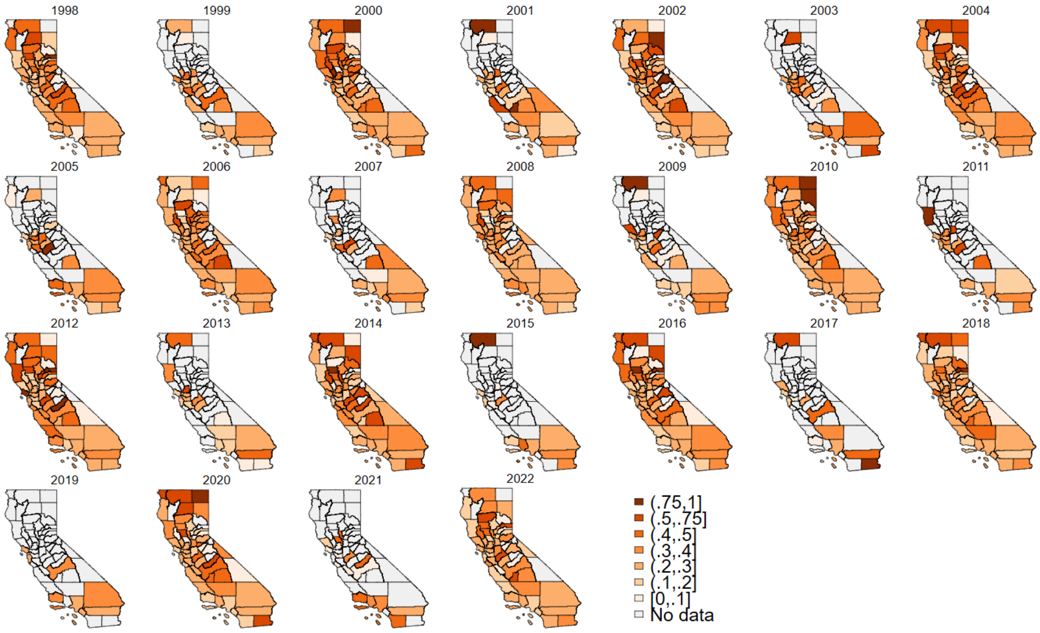


Figure 12: Share of candidates from routine occupations by county