December 2024

Fondazione Eni Enrico Mattei

Working Paper

28.2024

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

Paolo Agnolin, Italo Colantone, Piero Stanig

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

Paolo Agnolin (Bocconi University, Duke University and Dondena Research Centre; **Italo Colantone** (Bocconi University, Baffi Research Centre, GREEN Research Centre, CESifo and Fondazione Eni Enrico Mattei); **Piero Stanig** (Bocconi University, Yale-NUS and National University of Singapore - NUS)

Summary

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

JEL Classification: F1, F5

Corresponding Author:

Piero Stanig
Department of Social and Political Sciences - Bocconi University
Via Roentgen 1, 20136 Milan
e-mail: piero.stanig@unibocconi.it

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

Paolo Agnolin[†] Italo Colantone[‡] Piero Stanig[§]

November 29, 2024

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.

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 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,

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

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.

1 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 0.082* [0.040]	China shock -0.004 [0.033]	China shock 0.445* [0.210]	China shock 0.380 [0.210]	Family fin. -0.204** [0.078]	Family fin. -0.114 [0.075]
Cultural attitude:	-	Immigration -0.132** [0.002]	-	Cons. omnivore -1.074** [0.081]	-	Immigration -1.306** [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

[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.

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 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 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.,

[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.

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]

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

[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.

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.

2 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 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 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,

[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.

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)

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]

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.

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

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 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 [0.03]	-0.07 [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.

2.1 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.

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

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 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 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

[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.

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] 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.

3 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

[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 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).

from China on a large array of cultural 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:

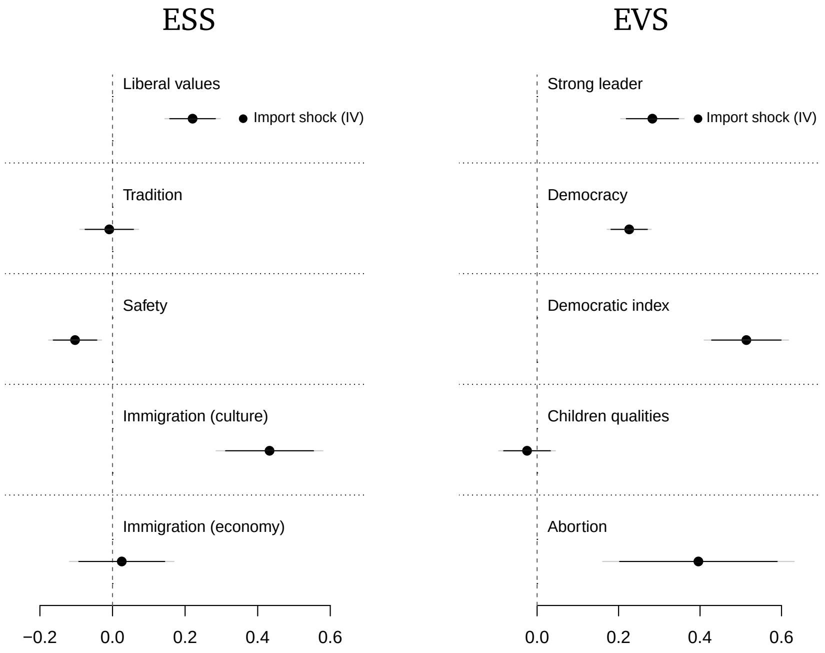
$$Cultural\ Attitude_{icrt} = \alpha_{ct} + \beta_1\ Import\ Shock_{cr(i)t} + \boldsymbol{Z_{it}}\gamma' + \epsilon_{icrt} \tag{1}$$

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 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

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

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.

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 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.

4 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 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

[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.

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.

Table A.3: Results

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.:	Leave	Leave	Omnivore	Paucivore	Omnivore	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]
$(\text{Age})^2$	-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 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). 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 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.

[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.

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. ΔIMPChina$_{cjt}$ 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:

$$Instrument\ Import\ Shock_{crt} = \sum_{j} \frac{L_{rj}^{pre-sample}}{L_{r}^{pre-sample}} * \frac{\Delta IMPChinaUSA_{jt}}{L_{cj}^{pre-sample}} \tag{7}$$

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.

[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.

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.

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.

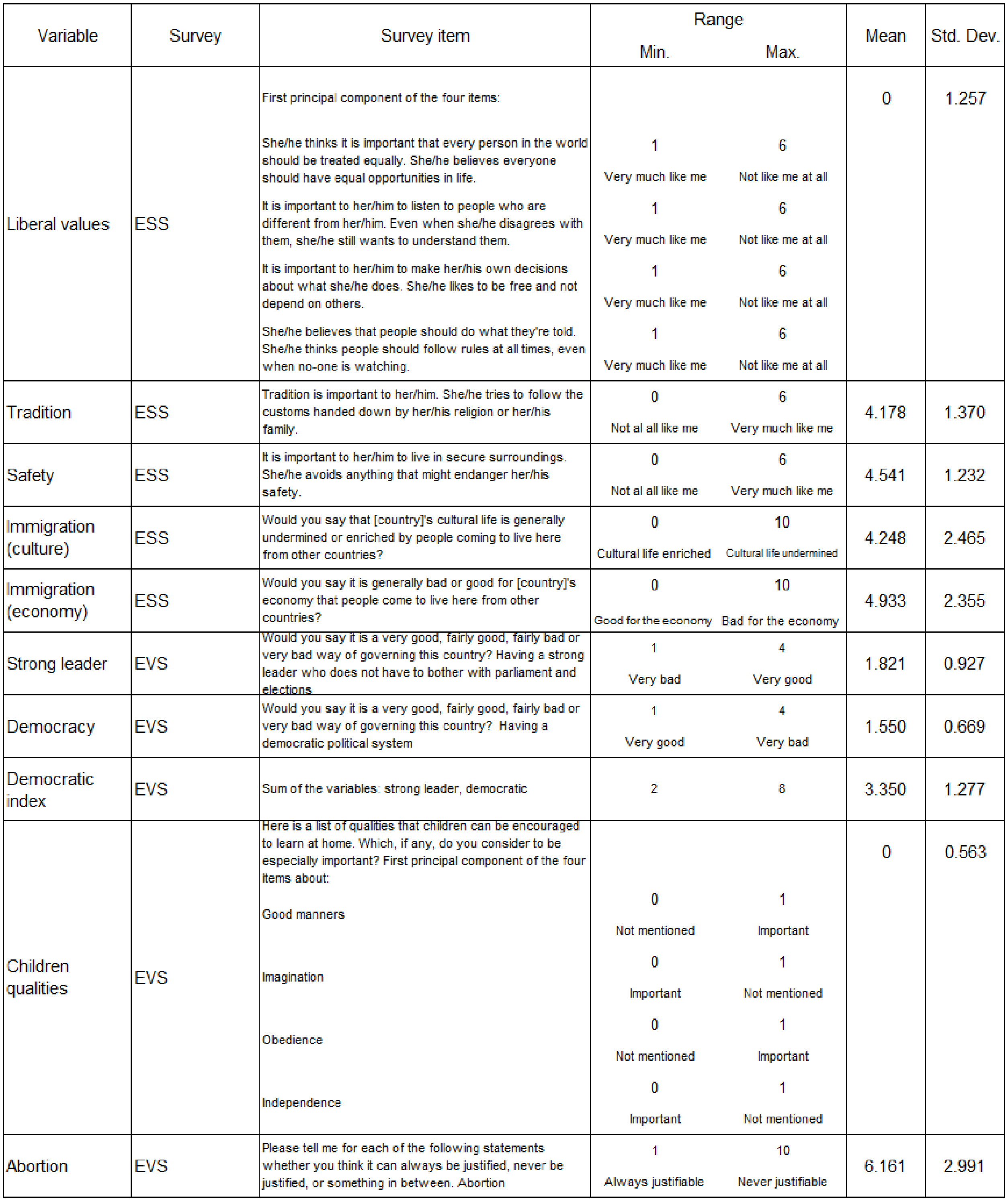
Table D.2: Cultural attitudes on import shock - EVS sample

	(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

D.4 Information on variables

Table D.3: Cultural attitudes



Variable	Survey	Survey item	Range		Mean	Std. Dev.
			Min.	Max.		
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].

FONDAZIONE ENI ENRICO MATTEI WORKING PAPER SERIES

Our Working Papers are available on the Internet at the following address:
https://www.feem.it/pubblicazioni/feem-working-papers/

"NOTE DI LAVORO" PUBLISHED IN 2024

1. A. Sileo, M. Bonacina, The automotive industry: when regulated supply fails to meet demand. The Case of Italy
2. A. Bastianin, E. Mirto, Y. Qin, L. Rossini, What drives the European carbon market? Macroeconomic factors and forecasts
3. M. Rizzati, E. Ciola, E. Turco, D. Bazzana, S. Vergalli, Beyond Green Preferences: Alternative Pathways to Net-Zero Emissions in the MATRIX model
4. L. Di Corato, M. Moretto, Supply contracting under dynamic asymmetric cost information
5. C. Drago, L. Errichiello, Remote work admist the Covid-19 outbreak: Insights from an Ensemble Community-Based Keyword Network Analysis
6. F. Cappelli, Unequal contributions to CO2 emissions along the income distribution within and between countries
7. I. Bos, G. Maccarrone, M. A. Marini, Anti-Consumerism: Stick or Carrot?
8. M. Gilli, A. Sorrentino, The Set of Equilibria in Max-Min Two Groups Contests with Binary Actions and a Private Good Prize
9. E. Bachiocchi, A. Bastianin, G. Moramarco, Macroeconomic Spillovers of Weather Shocks across U.S. States
10. T. Schmitz, I. Colantone, G. Ottaviano, Regional and Aggregate Economic Consequences of Environmental Policy
11. D. Bosco, M. Gilli, Effort Provision and Incentivisation in Tullock Group-Contests with Many Groups: An Explicit Characterisation
12. A. Drigo, Environmental justice gap in Italy: the role of industrial agglomerations and regional pollution dispersion capacity
13. P. I. Rivadeneyra García, F. Cornacchia, A. G. Martínez Hernández, M. Bidoia, C. Giupponi, Multi-platform assessment of coastal protection and carbon sequestration in the Venice Lagoon under future scenarios
14. T. Angel, A. Berthe, V. Costantini, M. D'Angeli, How the nature of inequality reduction matters for CO2 emissions
15. E. Bacchiocchi, A. Bastianin, T. Kitagawa, E. Mirto, Partially identified heteroskedastic SVARs
16. B. Bosco, C. F. Bosco, P. Maranzano, Income taxation and labour response. Empirical evidence from a DID analysis of an income tax treatment in Italy
17. M. M. H. Sarker, A. Gabino Martinez-Hernandez, J. Reyes Vásquez, P. Rivadeneyra, S. Raimondo, Coastal Infrastructure and Climate Change adaptation in Bangladesh: Ecosystem services insights from an integrated SES-DAPSIR framework
18. P. Maranzano, M. Pelagatti, A Hodrick-Prescott filter with automatically selected jumps
19. M. Bonacina, M. Demir, A. Sileo, A. Zanoni, The slow lane: a study on the diffusion of full-electric cars in Italy
20. C. Castelli, M. Castellini, C. Gusperti, V. Lupi, S. Vergalli, Balancing Climate Policies and Economic Development in the Mediterranean Countries
21. M. Gilli, A. Sorrentino, Characterization of the Set of Equilibria in Max-Min Group Contests with Continuous Efforts and a Private Good Prize
22. P. Pakrooh, M. Manera, Causality, Connectedness, and Volatility Pass-through among Energy-Metal-Stock-Carbon Markets: New Evidence from the EU
23. F. F. Frattini, F. Vona, F. Bontadini, Does Green Re-industrialization Pay off? Impacts on Employment, Wages and Productivity
24. A. Drigo, Breathing Inequality? Income, Ethnicity and PM2.5 Exposure in Bologna, Italy
25. D. Bosco, M. Gilli, A. Sorrentino, A Max-Min Two-Group Contest with Binary Actions and Incomplete Information à la Global Games
26. C. Amadei, C. Dosi, F. J. Pintus, Energy Intensity and Structural Changes: Does Offshoring Matter?
27. E. Ciola, E. M. Turco, M. Rizzati, D. Bazzana, S. Vergalli, Taking the green pill: Macro-financial transition risks and policy challenges in the MATRIX model

Fondazione Eni Enrico Mattei
Corso Magenta 63, Milano – Italia

Tel. +39 02 403 36934

E-mail: letter@feem.it
www.feem.it

Fondazione Eni
Enrico Mattei