

NBER WORKING PAPER SERIES

INSTRUMENTAL VARIABLES AND CAUSAL MECHANISMS:
UNPACKING THE EFFECT OF TRADE ON WORKERS AND VOTERS

Christian Dippel
Robert Gold
Stephan Heblich
Rodrigo Pinto

Working Paper 23209
<http://www.nber.org/papers/w23209>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2017, Revised January 2018

We thank David Autor, Sascha Becker, Gilles Duranton, Jon Eguia, Andreas Ferrara, Paola Giuliano, Tarek Hassan, Kosuke Imai, Ralf Meisenzahl, Ralph Ossa, Anne Otto, David Pacini, Nicola Persico, Giacomo Ponzetto, Daniel Sturm, Peter Schott, Dan Trefler, Dan Treisman, Nico Voigtländer, Wouter Vermeulen, Till von Wachter, Romain Wacziarg, Frank Windmeijer, Yanos Zylberberg, and seminar participants at Bristol, Kiel, the LMU, the LSE, Toronto, UCLA, Warwick, the 2013 Urban Economics Association and German Economists Abroad Meetings, the 2015 Quebec Political Economy conference, and the 2016 EEA conference for valuable comments and discussions. We thank Wolfgang Dauth for sharing the crosswalk from product classifications to industry classifications in the German IAB data. Dippel acknowledges financial support from UCLA's Center for Global Management. This paper builds on an earlier working paper titled "Globalization and its (Dis-)Content." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Christian Dippel, Robert Gold, Stephan Heblich, and Rodrigo Pinto. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Instrumental Variables and Causal Mechanisms: Unpacking The Effect of Trade on Workers and Voters

Christian Dippel, Robert Gold, Stephan Heblich, and Rodrigo Pinto

NBER Working Paper No. 23209

March 2017, Revised January 2018

JEL No. F1,F6,J2

ABSTRACT

It is often the case that an endogenous treatment variable causally affects an intermediate variable that in turn causally affects a final outcome. Using an Instrumental Variable (IV) identifies the causal effect of the endogenous treatment on both the intermediate and the final outcome variable, but not the extent to which the intermediate variable affects the final outcome. We present a new testable framework in which a single IV suffices to also estimate the causal effect of the intermediate variable on the final outcome. We use this framework to investigate to what extent German voters responded to the labor market turmoil caused by increasing trade with low-wage manufacturing countries. We first establish that import competition increased voters' support for only extreme (right) parties. We then decompose this populist 'total effect' into a 'mediated effect' running through labor market adjustments and a 'direct effect' of trade exposure on voting behavior. We find the total consists of a large populist effect driven by labor markets and a relatively smaller but moderating direct effect. Our approach provides a template that may be useful in a broad range of empirical applications studying causal mechanisms in observational data.

Christian Dippel
UCLA Anderson School of Management
110 Westwood Plaza, C-521
Los Angeles, CA 90095
and NBER
christian.dippel@anderson.ucla.edu

Robert Gold
Kiel Institute for the World Economy
Kiellinie 66
24105 Kiel
Germany
robert.gold@ifw-kiel.de

Stephan Heblich
Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN
UK
stephan.heblich@bristol.ac.uk

Rodrigo Pinto
Department of Economics
8283 Bunche Hall
Los Angeles, CA 90095
rodrig@econ.ucla.edu

1 Introduction

International trade between high and low-wage countries has risen dramatically in the last thirty years (Krugman, 2008). While consumers in high-wage countries have benefited from such import exposure through cheaper manufactured goods, there has also been real wage stagnation and substantial losses of manufacturing jobs (Autor, Dorn, and Hanson, 2013; Dauth, Findeisen, and Suedekum, 2014; Pierce and Schott, 2016; Malgouyres, 2017). Import exposure also appears to have impacted politics, i.e. by increasing support for parties and politicians with protectionist, populist, and nationalist agendas (Malgouyres, 2014; Feigenbaum and Hall, 2015; Autor, Dorn, Hanson, and Majlesi, 2016; Che, Lu, Pierce, Schott, and Tao, 2016).

We identify the causal mechanisms that link the nexus of import exposure, labor market adjustments and political populism. For this purpose we develop a novel framework for mediation analysis in IV settings. In particular, we are interested in the extent to which import exposure turns voters towards political populism *because* it affects labor markets. The first step towards an answer is to separately estimate the causal effect of import exposure on labor markets and on voting. The well-understood identification challenge with this is that import exposure may be endogenous because of unobserved confounding variables such as local demand shocks. The solution involves using an instrumental variable (IV), which allows for the identification of the causal effect of import exposure on labor markets and on voting.¹ However, the standard IV model does not identify the extent to which import exposure affects voting *because* it affects labor markets. We provide a novel solution to this problem.

For clarity, we label import exposure as the ‘treatment’ T , labor market adjustments as a potential ‘mediator’ M , voting as the ‘final outcome’ Y , and the instrument as Z . *Models I–II* in Table 1 depict the standard IV model where confounder V makes T endogenous, and the usual exclusion restrictions make Z a valid instrument: $M(t) \perp\!\!\!\perp Z$ and $Y(t) \perp\!\!\!\perp Z$.² Without added assumptions, the extent to which T causes Y *through* M remains unidentified in *Models I–II*.

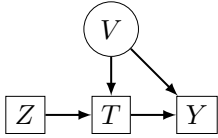
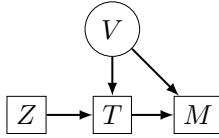
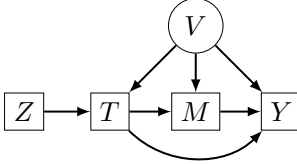
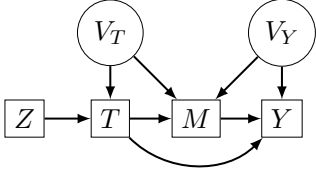
Our question requires a *mediation model*, i.e. one where T (import exposure) causes an inter-

¹The same IV strategy may be employed to identify the causal effect of import exposure on several outcomes. For example, three pairs of papers in the literature above each use the same identification strategy to separately investigate the effect of import exposure on labor markets and on some form of political outcomes; e.g. Autor et al. (2013) and Autor et al. (2016), Malgouyres (2017) and Malgouyres (2014), as well as Pierce and Schott (2016) and Che et al. (2016).

²We use $\perp\!\!\!\perp$ to denote statistical independence, $Y(t)$ for the ‘counterfactual outcome’ and $M(t)$ for the ‘counterfactual mediator’ when T is fixed at value t . See section 4 for detailed notation.

Table 4: General Mediation Model and Our Solution to the Identification Problem
Table 1: The Identification Problem of Mediation Analysis with IV

A. Directed Acyclic Graph (DAG) Representation

Model I: IV for Voting Y	Model II: IV for Labor M	Model III: General Mediation	Model IV: Restricted Mediation
			
B. Structural Equations			
$T = f_T(Z, V, \epsilon_T),$	$T = f_T(Z, V, \epsilon_T),$	$T = f_T(Z, V, \epsilon_T)$	$T = f_T(Z, V_T, \epsilon_T)$
$Y = f_Y(T, V, \epsilon_Y),$	$M = f_M(T, V, \epsilon_M),$	$M = f_M(T, V, \epsilon_M)$	$M = f_M(T, V_T, V_Y, \epsilon_M)$
$Z, V, \epsilon_T, \epsilon_Y$ Stat. Indep.	$Z, V, \epsilon_T, \epsilon_M$ Stat. Indep.	$Y = f_Y(T, M, V, \epsilon_Y)$	$Y = f_Y(T, M, V_Y, \epsilon_Y)$
		$Z, V, \epsilon's$ Stat. Indep.	$Z, V_Y, V_T, \epsilon's$ Stat. Indep.

Panel A gives the directed acyclic graph (DAG) representation of the models. *Model I* and *Model II* are the standard IV models, which enable the identification of the causal effects of T on Y and T on M . *Model III* is the General Mediation Model with an instrumental variable Z . Identification can only be achieved with an additional designated instrument for M (not depicted). *Model IV* is the Restricted Mediation Model. This model enables the identification of the total, the direct and the indirect effect of T on Y , with only one instrumental variable Z . The model also enables the identification of the causal effect of M on Y . Panel B presents the nonparametric structural equations of each model. Conditioning variables are suppressed for sake of notational simplicity. More detailed versions of these DAGs are in section 4. We refer to Heckman and Pinto (2015a) for a recent discussion on causality and directed acyclic graphs.

mediate outcome M (labor markets) which in turn causes a final outcome Y (voting). Mediation analysis decomposes the *total effect* of T on Y into the *mediated effect* of T on Y that operates through M and the *residual effect* that does not.³ *Model III* of Table 1 shows the main identification challenge in combining the two IV models into a *General Mediation Model*: the relation among T, M, Y may be tempered by confounding variables V that jointly cause all three of T, M , and Y .⁴

Existing approaches to achieving identification in *Model III* either assume that T is as good as randomly assigned, i.e. there is no V and therefore no need for an IV (e.g. Imai, Keele, Tingley, and Yamamoto 2011a), or require separate instruments which need to be dedicated to M and require additional exogeneity assumptions (e.g. Jun, Pinkse, Xu, and Yildiz 2016; Frolich and Huber 2017).

³The mediated effect may alternatively be labeled as the ‘indirect effect’, and the residual effect as the ‘direct effect’. For recent works on this literature, see Heckman and Pinto (2015b); Pearl (2014); Imai, Keele, and Tingley (2010).

⁴ A second challenge is that T may cause other unobserved mediators U that impact not only Y but also M . A pertinent example for U are *Trade Adjustment Assistance* (TAA) programs in the U.S., which are specifically designed to cushion the effect of import exposure on labor markets and which may well impact the vote share of the Democratic party that has championed these programs. For expositional clarity, we include U only into later DAGs, and with a more detailed discussion. See Tables 4 and 5 in section 4.

However, there are in fact many settings in which the concerns about confounders V suggest a solution that requires neither restrictive assumptions nor additional instruments. Specifically, this is true if unobservables that confound the relationship between T and M are thought to affect Y only through T and M . For instance, if negative domestic demand shocks reduced a region’s import exposure (T) and employment (M), then it would seem eminently reasonable that this affects voting (Y) only through T and M . In turn, while local political decisions such as zoning may affect local employment (M) and voting behavior (Y), they should not impact local import exposure (T) as it is measured in the literature. This argument gives rise to the structure represented by *Model IV* of Table 1, which replaces the confounding variable V with two unobserved confounding variables: V_T that affects T and M and V_Y that affects M and Y . This structure does not assume away confounding effects; indeed variables T , M , Y remain endogenous. Yet, section 4 shows that this assumption generates the identifying variation and the exclusion restriction that allow us to identify the causal effect of M on Y , thus allowing to identify the extent to which T causes Y through M . See section 4.2 for a discussion of the framework’s intuition.⁵

In section 4.3 we show that under linearity, the model is straightforwardly estimated using three separate Two-Stage Least Squares (2SLS) estimations of the causal effects of T on M , T on Y , and M on Y .⁶ In section 4.4 we derive a simple model specification test of our assumptions. In section 4.5 we explicitly contrast our model with an alternative approach that uses a separate dedicated instrument for M . We use automation as an example that clarifies the restrictive exogeneity assumptions required in this approach. Our method is not specific to our empirical setting. It may be applied to enrich causal analysis wherever IV is used to evaluate the causal effect of a treatment T on multiple outcomes, and where one observed outcome (M) in turn is thought to influence another (Y).

Our data combine changes in sector-specific trade flows with local labor markets’ initial industry mix to determine regional import exposure (T). We then instrument T with a measure based

⁵We also discuss some research applications for which we view our identifying assumptions as less suitable.

⁶ Our main focus is to estimate the product of the effect T on M and the effect M on Y , i.e. the *mediated effect*, and to compare it to the estimated *total effect* of T on Y . This objective is naturally similar to traditional approaches to mediation analysis, which assume that both T and M are exogenous, and apply OLS to estimate three equations

$$Y_{it} = \beta_T^Y T_{it} + \epsilon_{it}^Y, \quad M_{it} = \beta_T^M T_{it} + \epsilon_{it}^M, \quad Y_{it} = \beta_T^{Y|T} T_{it} + \beta_M^{Y|T} M_{it} + \epsilon_{it}^{Y|T},$$

and then compare the total effect β_T^Y to the mediated effect $\beta_T^M \times \beta_M^{Y|T}$. See [Baron and Kenny \(1986\)](#) and [MacKinnon \(2008\)](#) for an overview.

on other high-wage countries' sector-specific trade flows (Z).⁷ The data is organized as a stacked panel of two first differences for the periods 1987–1998 and 1998–2009, with specific start- and end-points dictated by national election dates. The analysis precedes the European debt crisis and each period includes a large international trade shock: In 1989, the fall of the Iron Curtain opened up the Eastern European markets, and in 2001 China's accession to the WTO led to another large increase in import exposure.

We use German data because it offers several advantages for the question at hand: (i) Unlike the U.S., where voters only chose between two parties, Germany's multi-party system straddles the entire political spectrum from the far-left to the extreme right so that we can consistently measure changes in political preferences over time. (ii) Unlike the U.S., where individual congressmen adjust their policy stance to local conditions because they are elected by district, Germans cast their main vote for a party *at large* so that local voting patterns are un-confounded by local variation in political messaging. (iii) Unlike the U.S., where political boundaries and economic statistics do not overlap well, we are able to measure vote shares, regional import exposure and labor market conditions all at the same statistical unit of 408 *Landkreise*.⁸ (iv) Unique amongst attitudinal socioeconomic surveys, the German *Socio-Economic Panel's* (SOEP) long-running panel structure allows us to cross-validate the aggregate results with an individual-level panel-analysis, relating decadal changes in individual workers' stated party preferences to changes in their local labor markets' import exposure over the same time.

Estimating *Model I* in Table 1, we find that import exposure (T) increased voters' support (Y) for only the narrow segment of the nationalist extreme right, with its highly protectionist agenda. There is a positive but insignificant effect on turnout, and no effects on any of the mainstream parties, small parties, or the far left.⁹ These findings are corroborated by the SOEP's individual-level data, where we can additionally show that the effects are entirely driven by low-skill workers employed in manufacturing, i.e. those most affected by the labor market adjustments to increasing

⁷In this, we follow the work of Autor et al. (2013) who suggest the import exposure of high-wage countries other than Germany as an IV for Germany's import exposure T . The resulting identifying variation is driven by supply changes (productivity or market access increases) in low-wage countries instead of fluctuations in German domestic conditions. We refer the reader to the literature above and to section 2.4 for more details.

⁸In U.S. data, one observes vote-shares in 3,007 counties, politicians in 435 congressional districts, and trade shocks in 741 commuting zones.

⁹Election outcomes are divided into changes in the vote-share of (i) four mainstream parties: the CDU, the SPD, the FDP and the Green party, (ii) extreme-right parties, (iii) far-left parties, (iv) other small parties, and (v) turnout, see Falck, Gold, and Heblich (2014).

import exposure. Estimation based on gravity residuals instead of our IV strategy yields similar results.¹⁰

Estimating *Model II* in Table 1, we corroborate existing results that import exposure significantly reduces employment (M), particularly in manufacturing.

It seems likely that import exposure affects voting because it affects employment, but this is far from the only potential reason voters may respond: On the one hand, import exposure may make people better off because of output price reductions and increases in internationally-made varieties. It may also increase transfers through government programs like TAA that target trade-exposed regions. In addition, import exposure may lead German manufacturers to shift production towards more differentiated and higher mark-up varieties, as it did in the U.S. (Holmes and Stevens, 2014). In Germany, there is also strong evidence of trade-induced task-upgrading within industries and occupations (Becker and Muendler, 2015). All of these effects may reinforce support for mainstream policies of trade liberalization. On the other hand, if import exposure creates anxiety about the future it may turn voters away from the mainstream beyond its measurable labor market effects (Mughan and Lacy, 2002; Mughan, Bean, and McAllister, 2003). Depending on these factors' relative importance, the residual (i.e. not mediated by employment) effect of import exposure on voting for the extreme right may be positive or negative. Unpacking these causal links is important because it determines the degree to which political populism can be combated with labor market policies.

Estimating *Model IV* in Table 1, we find that import exposure does indeed affect extreme-right voting by means of reducing employment.¹¹ This effect of labor market adjustments caused by import exposure is even larger than the overall effect of import exposure on extreme-right voting. Stated differently, the mediated effect is larger than the total effect, which implies that other channels that link import exposure to voting (the 'residual effect') are moderating in the aggregate. We extend the analysis to the set of all labor market variables we observe.¹² Several of these correlate with total employment and could plausibly mediate between import exposure and voting, e.g. manufacturing employment, wages, and unemployment. Without additional dedicated instru-

¹⁰We report these for completeness as this is standard (Autor et al., 2013; Dauth et al., 2014). However, our focus is naturally on the IV setting to which our identification framework applies.

¹¹We also pass the specification test we develop in section 4.4.

¹²This is of more general interest because researchers will often be in the situation of having a number of observed variables that potentially link a treatment T to an outcome Y .

ments, our framework estimates the mediating effect of a single variable. We therefore need to aggregate the additional labor market variables into one index. To do this, we perform a *principal component* (PC) analysis.¹³ This approach is appealing as long as the mediating effects are sharply concentrated in one PC, and this PC has a clear interpretation. Indeed, we find that changes in manufacturing employment and wages are concentrated in a single PC in our data, and that this PC is solely responsible for linking import competition to extreme-right voting. Quantitatively, this analysis confirms that labor market adjustments, concentrated in manufacturing, are the main reason for the political backlash against free trade in the data we study.

Our paper’s contribution is two-fold: It answers a relevant substantive question, and in order to do so it makes a methodological contribution to the literature on causal mechanisms and on IV, which we discuss in detail in section 4.5. On the substantive side, our paper sits at the nexus of the literatures on trade, local labor markets and politics. Our findings provide a first causal estimate of the importance of labor market adjustments in explaining the effect of import exposure on voting. More broadly, we relate to a literature on the effects of economic shocks on voters (Scheve and Slaughter, 2001; Bagues and Esteve-Volart, 2014; Jensen, Quinn, and Weymouth, 2016; Charles and Stephens, 2013; Brunner, Ross, and Washington, 2011; Giuliano and Spilimbergo, 2014) and political cleavages (Rogowski, 1987; Hiscox, 2002).

On the methodological side, we offer a mediation model which relies on a single instrumental variable Z that directly causes T to identify three causal effects, while allowing for endogenous variables caused by confounders and for unobserved mediators. This parsimonious feature is particularly useful for empirical applications in which good instrumental variables are scarce, i.e. most. Our model is testable, can be estimated by well-known 2SLS methods and applied to a broad range of empirical research questions.

In the following, section 2 describes the data. Section 3 presents the IV results for *Models I–II*, establishing the causal effects of import exposure on voting and on labor markets. Section 4 explains our mediation model and lays out our identification approach to unpack the causal mechanism by which import exposure changes voting. Section 5 applies our mediation *Model IV* to estimates the causal links between trade exposure, labor market adjustments and voting behavior. Section 6 concludes.

¹³PC analysis is attractive in this context because it generates orthogonal indices that are purely statistical.

2 Data

Our data is organized as stacked panel of first differences between election dates, 1987 to 1998 (period 1) and 1998 to 2009 (period 2), staying as close as possible to the decadal changes usually studied in the literature. We study regional exposure to German trade with Eastern Europe and China, that was exogenously affected by the fall of Communism and China’s WTO accession. In Germany, imports from and exports to China and Eastern Europe roughly tripled over the period 1987 to 1998 (from about 20 billion to about 60 billion Euros each),¹⁴ and again tripled between 1998 and 2009.

Our data is observed at the county (*Landkreis*) level.¹⁵ We drop all city states from the sample, and follow [Dauth et al. \(2014\)](#) in excluding East-German counties from the first period of analysis, but including them in the second period. We observe 408 counties in our data, 86 of which are in East Germany. Over two periods, we have 730 ($= (408 - 86) + 408$) observations in total. For reference, we represent the data as 2 separate *Landkreise*-maps for periods 1 and 2 in [Appendix A](#). We include period-specific fixed effects for four broad regions in all estimations. These imply that the imbalanced nature of the panel has no bearing on any coefficient estimates. Indeed, none of our results are affected at all by dropping East Germany altogether, and we include it primarily to stay close to the existing literature.

With a view towards the mediation framework we develop in section 4, we need the following variables: *Treatment* T_{it} is our measure of local labor market i ’s trade exposure in period t . *Mediators* M_{it} are labor market variables, and *Final Outcome* Y_{it} refers to voting outcomes. Finally, we construct Z_{it} as an *Instrument* for T_{it} . We now explain how these variables are measured.¹⁶

¹⁴Throughout the paper, we report values in thousands of constant-2005 Euros using exchange rates from the German Bundesbank.

¹⁵ We follow [Dauth et al. \(2014\)](#) in using counties as a representation of German local labor markets. [Dauth et al. \(2014\)](#) show that results are qualitatively identical when using broader ‘functional labor markets’ but at the cost of econometric precision.

¹⁶ Conditioning variables X_{it} are discussed with the results in section 3.1.

2.1 Import Exposure (Treatment T)

We follow [Autor et al. \(2013\)](#) and [Dauth et al. \(2014\)](#) and calculate T as *net import exposure*:

$$T_{it} = \sum_j \frac{L_{ijt}}{L_{jt}} \frac{\Delta IM_{Gjt} - \Delta EX_{Gjt}}{L_{it}}. \quad (1)$$

ΔIM_{Gjt} denotes changes in Germany’s imports in industry j in period t . Local labor market i ’s composition of employment at the beginning of period t determines its exposure to changes in industry-specific trade flows ΔIM_{Gjt} over the ensuing decade.¹⁷ Sector j receives more weight if region i ’s national share of that sector $\frac{L_{ijt}}{L_{jt}}$ is high, but a lower weight if i ’s overall workforce L_{it} is larger. [Autor et al. \(2013\)](#) focus on imports (ΔIM_{Gjt}) and consider the net of imports (ΔIM_{Gjt}) minus exports (ΔEX_{Gjt}) only in their appendix. [Dauth et al. \(2014\)](#) show that in Germany imports from and exports to low-wage manufacturers are not only more balanced in the aggregate than in the U.S. but also correlate positively at the industry level. As a result, we rely on a local labor market’s *net* import exposure throughout the paper.

One concern with the measure of trade exposure in equation (1) is that it is a composite effect of the relative importance of trade-intensive industries *and* the relative importance of manufacturing employment in a region (i.e. $\frac{1}{L_{it}}$ relative to $\sum_j L_{ijt}$). The share of manufacturing employment might independently shape subsequent labor-market and voting changes. This problem is well known, and is solved by always conditioning on region i ’s initial share of manufacturing employment in all our regressions ([Autor et al., 2013](#)).

2.2 Labor Market Variables (Mediator M)

From the *Institut für Arbeitsmarkt- und Berufsforschung* (IAB)’s Historic Employment and Establishment Statistics (HES) database we glean information on workers’ industry of employment, occupation, and place of work for all German workers subject to social insurance.¹⁸ From the individual-level data we aggregate up to the *Landkreis* level to match our voting data. We focus on

¹⁷The *Institut für Arbeitsmarkt- und Berufsforschung* (IAB) reports industries of employment L_{ij} in standard international trade classification (SITC), and we link these to the UN Comtrade trade data using the crosswalk described in [Dauth et al. \(2014\)](#), which covers 157 manufacturing industries.

¹⁸see [Bender, Haas, and Klose 2000](#) for a detailed description. Civil servants and self-employed individuals are not included in the data. Furthermore, we exclude workers younger than 18 or older than 65 and we exclude all individuals in training and in part-time jobs because their hourly wages cannot be assessed.

decadal changes in (i) total employment. In addition, section 5.2 extends the analysis to a number of additional labor market outcomes that potentially connect T to Y : (ii) manufacturing’s employment share, (iii) manufacturing wages, (iv) non-manufacturing wages, and (v) unemployment, with data for the last one coming from the *German Statistical Office*. [Online Appendix A](#) provides additional information on data sources and variable construction.

2.3 Voting (Final Outcome Y)

To measure how import exposure affects voting behavior, we focus on party-votes in federal elections in Germany (*Bundestagswahlen*).¹⁹ Due to its at-large voting system Germany, like most continental European countries, has consistently had a multi-party system that spans the full spectrum from far-left to extreme-right parties. This allows us to contrast the effect of import exposure on populists parties’ vote share with that for moderate parties. There are four parties that we label ‘established’ in that they were persistently represented in parliament over the 25 years we study. There is also a large number of small parties. The average vote share of these small parties is far below the 5% threshold of party votes needed to enter the federal parliament.²⁰ We collected these data to create a novel dataset of party vote shares at the county level. We group the small parties into three categories: far-left parties, extreme-right parties, and a residual category of other small parties. Altogether, *Landkreis*-level voting outcomes are divided into changes in the vote-share of (i) four mainstream parties: the CDU, the SPD, the FDP and the Green party, (ii) extreme-right parties, (iii) far-left parties, (iv) other small parties, and (v) turnout, see [Falck et al. \(2014\)](#). [Online Appendix B](#) provides additional background on the German political system and party landscape.

2.4 Others’ Import Exposure (Instrument Z)

Endogeneity concerns in estimating the effect of import exposure on labor markets and voting come from the fact that domestic demand and supply shocks may simultaneously affect T_{it} , local

¹⁹The party vote, called (*Zweitstimme*), mainly determines a party’s share of parliamentary seats. German voters also cast a second vote for individual candidates, called (*Erststimme*). This vote for individuals affects the very composition of party factions in the parliament, but has no significant influence on their overall parliamentary share. Moreover, the decision on individual candidates might be strategic. We thus follow [Falck et al. \(2014\)](#) and focus on the party vote.

²⁰This threshold is not binding if a party wins at least three seats through the vote for individual candidates (*Erststimme*). During our period of analysis, this occurred once in 1994. The individual candidates of the party PDS won 4 seats by *Erststimme*. As a result, the party received 30 seats in total, according its 4.4% of party votes (*Zweitstimme*) received.

labor market outcomes, and local voting behavior.

To overcome this problem, we follow the approach in [Autor et al. \(2013\)](#) and instrument Germany's imports from (exports to) China and Eastern Europe, ΔIM_{Gjt} (ΔEX_{Gjt}), with the average imports from (exports to) a set of similar high-wage economies ΔIM_{Ojt} (ΔEX_{Ojt}).²¹

$$Z_{it}^{IM} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta IM_{Ojt}}{L_{it-1}}, \quad Z_{it}^{EX} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta EX_{Ojt}}{L_{it-1}}. \quad (2)$$

Also following [Autor et al. \(2013\)](#), we lag the initial employment shares by one decade to address reverse causality concerns, denoting the lag by the subscript $t - 1$.

3 Baseline Results

3.1 Model I: Estimating the Total Effect of T (Trade Exposure) on Y (Voting)

We now estimate *Model I* of Table 1, using standard IV and the following Second Stage equation:

$$Y_{it} = \Gamma_T^Y \cdot T_{it} + \Gamma_X^Y \cdot X_{it} + \epsilon_{it}^Y \quad (3)$$

Treatment T_{it} represents trade exposure as defined in equation (1). Outcome Y_{it} denotes changes in voting behavior in county i over period t . While we are ultimately interested in political support for populists, our data allow us to study effects on the entire political spectrum, i.e. changes in the vote-shares of moderate, small, extreme-right, and far-left parties, as well as turnout.

X_{it} denotes a selection of control variables. These are i 's start-of-period manufacturing employment share; the start-of-period employment share that is college educated, foreign born, or female; the employment share in the largest sector;²² along with separate controls for the employment share in car manufacturing and the chemical industry;²³ start-of-period vote-shares for

²¹We choose the same countries as [Dauth et al. \(2014\)](#) to instrument German imports and exports: Australia, Canada, Japan, Norway, New Zealand, Sweden, Singapore, and the United Kingdom. This set of countries excludes Eurozone countries because their demand conditions are likely correlated with Germany's.

²²It is a feature of the German economy that some regions are dominated by one specific industry. In such regions, individual firms (e.g. Daimler-Benz, Volkswagen, or Bayer) are likely to have political bargaining power, and as a result politicians may help buffer trade shocks to limit adverse employment effects.

²³The latter account for those industries' outstanding importance for the German economy.

all parties; voter turnout, start-of-period unemployment rate, and population-share of retirement age. X_{it} further includes a set of period-specific region fixed effects (North, West, South, and East Germany) with the regions being comparable to U.S. Census divisions (Dauth et al., 2014).²⁴ The regional fixed effects are period-specific to allow for different trends by period. Standard errors ϵ_{it} are clustered at the level of 96 larger economic zones defined by the Federal Office for Building and Regional Planning (BBR). The First Stage equation is

$$T_{it} = \Gamma_{IM}^T \cdot Z_{it}^{IM} + \Gamma_{EX}^T \cdot Z_{it}^{EX} + \Gamma_X^T \cdot X_{it} + \epsilon_{it}^T. \quad (4)$$

Instruments Z_{it}^{IM} and Z_{it}^{EX} are defined in (2). Table 2 presents our baseline results. Each cell reports results from a different regression. Rows specify different outcome variables, and columns refer to different regression specifications.²⁵

In our least conservative specification (column 1 of table 2), we consider the start-of-period manufacturing employment share as the only control. We always control for a region's start-of-period manufacturing share in employment because it inherently drives part of the variation in T_{it} ; see the discussion in 2.4. In column 2, we add controls for the structure of the workforce, i.e., the start-of-period employment share that is college educated, foreign born, or female. In column 3, we account for the disproportionate regional employment share of some firms by including a control for the employment share in the largest sector, along with separate controls for the employment share in car manufacturing and the chemical industry. In column 4, we add start-of-period vote-shares for all party outcomes and turnout. Finally, in column 5, we add the start-of-period unemployment rate and the population-share of retirement age. This is the most conservative specification, and our preferred one. In this specification, a one-standard-deviation increase in T_{it} (€1,350 per worker) increases the extreme-right vote share by 0.12 ($0.09 \cdot 1.35$) percentage points, roughly 28 percent of the average per-decade increase of 0.43 percentage points during the 22 years we study. Column 6 reports the results from our preferred specification as beta coefficients to facilitate comparison between the effects on election outcomes.

The effects are broadly consistent across all five specifications, though we see that the stepwise inclusion of controls reduces the effect size. Our findings suggest no effect on turnout; and looking

²⁴ Each of Germany's 16 states (*Bundesländer*) is fully contained inside one of these four regions.

²⁵ Results for the coefficients on all control variables are reported in Online Appendix D (table 2).

Table 2: Effect of Trade Exposure (T_{it}) on Voting (Y_{it})

		(1)	(2)	(3)	(4)	(5)	(6)
		Baseline IV	+ Structure IV	+ Industry IV	+ Voting IV	+Socio IV	Standard. IV
$\widehat{\Gamma}_T^Y$:	Δ Turnout	0.002 (0.939)	0.003 (1.192)	0.004 (1.455)	0.002 (1.095)	0.002 (1.223)	0.036 (1.223)
	<u>Established Parties:</u>						
	Δ Vote Share CDU/CSU	-0.128 (0.744)	-0.130 (0.808)	-0.180 (0.993)	-0.062 (0.475)	-0.066 (0.501)	-0.016 (0.501)
	Δ Vote Share SPD	-0.020 (0.129)	0.004 (0.030)	-0.006 (0.039)	-0.011 (0.090)	-0.009 (0.073)	-0.001 (0.073)
	Δ Vote Share FDP	0.215*** (2.788)	0.176** (2.384)	0.170** (2.197)	0.109 (1.377)	0.119 (1.583)	0.022 (1.583)
	Δ Vote Share Green Party	-0.132** (2.294)	-0.055 (1.309)	-0.030 (0.612)	-0.025 (0.551)	-0.018 (0.413)	-0.006 (0.413)
	<u>Non-established Parties</u>						
	Δ Vote Share Extreme-Right Parties	0.118*** (3.370)	0.099*** (3.118)	0.113*** (2.845)	0.086** (1.980)	0.089** (2.055)	0.044** (2.055)
	Δ Vote Share Far-Left Parties	-0.037 (0.289)	-0.078 (0.643)	-0.080 (0.639)	-0.068 (0.588)	-0.092 (0.859)	-0.024 (0.859)
	Δ Vote Share Other Small Parties	-0.015 (0.391)	-0.017 (0.458)	0.013 (0.327)	-0.028 (0.687)	-0.024 (0.564)	-0.018 (0.564)
	<u>First Stage:</u>						
$\widehat{\Gamma}_{IM}^T$:	Z_{it}^{IM}	0.225*** (8.220)	0.234*** (8.350)	0.221*** (7.816)	0.220*** (7.966)	0.220*** (7.971)	0.220*** (7.971)
$\widehat{\Gamma}_{EX}^T$:	Z_{it}^{EX}	-0.211*** (8.519)	-0.212*** (8.251)	-0.208*** (8.065)	-0.201*** (7.660)	-0.202*** (7.568)	-0.202*** (7.568)
F-Stat. of excluded Instruments		43.81	43.64	40.15	38.77	38.21	38.21
Period-by-region F.E.		Yes	Yes	Yes	Yes	Yes	Yes
Observations		730	730	730	730	730	730

Notes: (a) Each cell reports results from a separate instrumental variable regression. The data is a stacked panel of first-differences at the *Landkreis* level. Each regression has 730 observations, i.e. 322 *Landkreise* in West Germany, observed in 1987–1998 and 1998–2009, and 86 *Landkreise* in East Germany, observed only in 1998–2009. We drop three city-states (Hamburg, Bremen, and Berlin in the East). (b) All specifications include region-by-period fixed effects. Column 1 controls only for start-of-period manufacturing. Column 2 adds controls for the structure of the workforce (share female, foreign, and high-skilled). Column 3 adds controls for dominant industries (employment share of the largest industry, in automobiles, and chemicals). Column 4 adds start-of-period voting controls. Column 5 is our preferred specification, adding start-of-period socioeconomic controls (population share unemployed, and individuals aged 65+). Finally, Column 6 presents our preferred specification with standardized outcome variables to facilitate comparison. (c) The bottom panel presents the first stage results. It reports coefficients for only the two instruments, but includes the full set of controls from the top-panel. Results for the coefficients on all control variables are reported in [Online Appendix D](#) (table 2). (d) *t*-statistics are reported in round brackets, standard errors are clustered at the level of 96 commuting zones. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

at reactions across the political spectrum, we see no significant effects on established, small, or far-left parties in our preferred specification in column 5. The only segment of the party spectrum that responds consistently to trade shocks across all specifications is the vote-share of extreme-right parties.²⁶ Looking at the beta coefficients reported in column 6, we see that the estimated effects for all parties except the extreme right are not only insignificant but also small compared to the effect on extreme-right parties. For a better understanding of potential biases, we present corresponding OLS estimates in table 3 of [Online Appendix D](#).²⁷

Interpreting the Effects: Whether far-left or extreme-right populists capture the anti-globalization sentiment is ultimately country-specific. To argue with a recent headline in the *The Economist* (2016) “Farewell, left versus right. The contest that matters now is open against closed.” Nonetheless, political scientists argue that the political left in Europe has found it difficult to take a coherent position against globalization in the last two decades, often hampered by internal intellectual conflicts ([Arzheimer, 2009](#)). [Sommer \(2008, p. 312\)](#) argues that “in opposing globalization, the left-wing usually criticizes an unjust and profit-oriented economic world order. [It] does not reject globalization per se but rather espouses a different sort of globalization. In contrast, the solutions proposed by the extreme right keep strictly to a national framework. The extreme right’s claim, therefore, that it is the only political force that opposes globalization fundamentally [...] rings true.”²⁸ In [Appendix A](#) we provide anecdotal evidence linking local import exposure to increasing support for the extreme right in two regions of Germany.

While it is clear that import exposure has increased the extreme right’s vote share in our data, the overall effect is small. This is partly mechanical because in our setup the fixed effects absorb bigger shifts in voting behavior. More importantly, however, Germany did not have a pop-

²⁶However, the coefficient for the market-liberal FDP shows a marginally insignificant t-statistic of 1.58, and for turnout we see a t-statistic of 1.22. The latter indicates that turnout might increase with trade exposure. This would complement [Charles and Stephens \(2013\)](#), who find that positive economic shocks decrease voter turnout. One possible explanation for the positive though marginally insignificant effect on votes for the pro-market FDP is that regions hit by a trade shock may face increasing demand for redistribution or government intervention in markets ([Rodrik, 1995](#)). As a result, those who do not approve such policies may choose to vote for the FDP. Based on our reading of German politics, we take this as a hint for possible polarization, if the economically liberal FDP became an attractive choice for voters who position themselves against growing anti-globalization sentiments in their region.

²⁷A comparison between IV and OLS estimates for the effect on extreme-right parties shows that the OLS coefficient is consistently smaller than the IV coefficient. This result is in line with the concern that trade exposure partly reflects domestic sectoral demand shifts.

²⁸For illustration, we provide an excerpt from the extreme-right NPD’s ‘candidate manual’: “Globalization is a planetary spread of the capitalist economic system under the leadership of the Great Money. Despite by its very nature being Jewish-nomadic and homeless, it has its politically and militarily protected locus mainly on the East Coast of the United States” ([Grumke, 2012, p. 328](#)).

ulist party with broad appeal during our study period. All anti-globalization parties at the right fringe were extremist parties with neo-Nazi ties and associations to the *Third Reich*, which made them anathema to most Germans. Where populist leaders have broad appeal, like Marine Le Pen in France, or Donald Trump in the U.S., the political backlash to import exposure seems to be stronger. The coefficient size is thus specific to the political context, while the nexus of import exposure, labor market adjustments and political populism seems to be general. Our focus is therefore not on the magnitude of the effect of trade exposure on voting behavior but on the causal mechanisms underlying it.²⁹

Gravity: We also estimate results based on gravity residuals. This approach does not use IV but instead estimates the exogenous evolution of industry-specific Chinese and Eastern European comparative advantage over Germany based on a comparison of bilateral trade flows of Germany and ‘China plus Eastern Europe’ vis-a-vis the same set of destination markets.³⁰ The gravity results are reported in [Appendix B](#) and [Online Appendix E](#) and are in line with those in [table 2](#).

Individual-Level Analysis: One benefit of using German data is that the SOEP has a long-run panel structure that is unique amongst attitudinal socio-economic surveys, starting in 1984 ([GSOEP, 2007](#)).³¹ Importantly, we can locate individuals inside their local labor markets. As a result, we can associate individual workers w with their local labor market i ’s trade exposure (T), instrument T with Z as before, and add the same set of regional controls. This allows us to track decadal changes in individuals’ party preferences in a way that mirrors our main local labor market analysis.³² For our purpose, the relevant GSOEP question asks: “*If there was an election today, who would you vote for?*” We translate this question into a series of dummies that reflect the full party spectrum also observed in [table 2](#), e.g. one dummy if the individual would you vote for the CDU, one if the individual would vote for the SPD, etc. The results, reported in [Appendix C](#), mimic closely our main [table 2](#). A county’s import exposure shifts individuals’ preferences to the

²⁹ Aside from the size of estimated coefficients we also note that Germany had relatively balanced trade with low-wage manufacturing countries during our study period and did not experience the “China Shock” in the same way as the U.S. and other high-wage countries. See [Online Appendix C](#).

³⁰ See [Autor et al. \(2013\)](#) and [Dauth et al. \(2014\)](#) for a discussion of the gravity residuals approach relative to the IV approach.

³¹ In the U.S., the *General Social Survey* (GSS) for example only added a panel component in 2008.

³² Because the SOEP only started to ask about voting intentions for the full party spectrum in 1990 we use the time windows 1990-1998 and 1998-2009, i.e. a slightly shorter period 1 compared to our main results.

Table 3: Effect of Trade Exposure T_{it} on Total Employment

	(1)	(2)	(3)	(4)	(5)
	Baseline IV	+ Structure IV	+ Industry IV	+ Voting IV	+Socio IV
$\widehat{\Gamma}_T^M : \Delta \log(\text{Total Employment})$	-0.023*** (2.853)	-0.024*** (3.131)	-0.025*** (3.203)	-0.025*** (3.239)	-0.024*** (3.295)
Period-by-region FE	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730

Notes: (a) Each cell reports results from a separate instrumental variable regression. The data is a stacked panel of first-differences at the *Landkreis* level. Each regression has 730 observations, i.e. 322 *Landkreise* in West Germany, observed in 1987–1998 and 1998–2009, and 86 *Landkreise* in East Germany, observed only in 1998–2009. We drop three city-states (Hamburg, Bremen, and Berlin in the East). (b) All specifications include region-by-period fixed effects. Columns incrementally include added controls in the identical fashion as table 2. Column 5 is our preferred specification, which includes socioeconomic controls. The reported coefficient is a semi-elasticity: For example, a one-standard-deviation increase in T_{it} (€1,350 per worker) decreased total employment by about 3 percent, ($e^{-0.024 \cdot 1.35} - 1 = -0.032$). The unreported first stage is identical to that in table 2. *t-statistics* are reported in round brackets, standard errors are clustered at the level of 96 commuting zones. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

extreme right. Splitting the sample by worker types, we find the results to be entirely driven by low-skill workers, and more specifically those in manufacturing sectors, who are also most likely to experience adverse labor market effects from import exposure.

3.2 Model II: The Effect of Trade Exposure T on Labor Markets M

We now turn to estimating *Model II* of Table 1, where we consider the change in the log of total employment as outcome M_{it} . We emphasize that the effect of import exposure on labor markets has been examined in detail in Autor et al. (2013), and their results have since then been replicated for several other high-wage countries including Germany (Dauth et al., 2014). Again using standard IV, we estimate

$$M_{it} = \Gamma_T^M \cdot T_{it} + \Gamma_X^M \cdot X_{it} + \epsilon_{it}^M, \quad (5)$$

where the only difference to equation (3) is the changed outcome variable.³³ The results are displayed in table 3, which shares the same structure as table 2, each cell reporting on a different regression specification.

Import exposure has a significant negative effect on employment. In our preferred specification in column 5, a one-standard-deviation increase in T_{it} (€1,350 per worker) decreases total

³³We use the exact same set of controls in tables 2 and 3.

employment by about 3 percent ($e^{-0.024 \cdot 1.35} - 1 = -0.032$). We report corresponding OLS results in [Online Appendix D](#) (table 4). We do not study any individual-level labor market results.³⁴

Without the new methodological framework advanced in this paper, this is as far as we can go. We have a causally identified effect of import exposure on voting ($\hat{\Gamma}_T^Y = 0.089$), and a causally identified effect on total employment ($\hat{\Gamma}_T^M = -0.024$). We now introduce the mediation model that allows us to identify how much of the former is explained by the latter.

4 Identification: From *Model III* to *Model IV* of Table 1

In Section 3, we estimate *Model I* and *Model II* of Table 1 to evaluate two causal parameters: the total effect of trade exposure T on labor markets M and the total effect of T on voting behavior Y . We identify these effects using a standard IV model where the instrument Z is the trade exposure of countries other than Germany as in [Autor et al. \(2013\)](#). This evaluation does not identify the causal effect of M on Y , and therefore does not allow us to identify the effect of T on Y that runs through M . In this section, we develop a mediation model that enables us to identify this additional causal effect.

Some notation is needed to clarify ideas. We use $\text{supp}(T)$, $\text{supp}(M)$ for the support of variables T and M . We use $M(t)$ and $Y(t)$ for the potential outcomes of M, Y when T is fixed at value $t \in \text{supp}(T)$ and $Y(m)$ for the potential outcome of Y when M is fixed at $m \in \text{supp}(M)$. $Y(t, m)$ stands for the counterfactual outcome Y when T is fixed at value $t \in \text{supp}(T)$ and M is fixed at value $m \in \text{supp}(M)$. We use $\perp\!\!\!\perp$ for statistical independence and $\not\perp\!\!\!\perp$ for its negation. For sake of notational simplicity, we suppress conditioning variables X that we wish to control for. Our analysis can be understood as conditioned on those without loss of generality.

The identification of causal effects of T on M and on Y in Section 3 arises from three properties

³⁴We refer the reader to [Dauth et al. \(2014\)](#) who present such evidence for Germany based on the *Sample of Integrated Labour Market Biographies* (SIAB).

of the instrumental variable:³⁵

$$\text{Exclusion Restriction of Labor Market Variables: } Z \perp\!\!\!\perp M(t) \quad (6)$$

$$\text{Exclusion Restriction of Voting Outcomes: } Z \perp\!\!\!\perp Y(t) \quad (7)$$

$$\text{IV Relevance : } Z \not\perp\!\!\!\perp T. \quad (8)$$

The general causal model that generates the IV properties (6)–(8) is given by:³⁶

$$T = f_T(Z, V, \epsilon_T), \quad M = f_M(T, V, \epsilon_M), \quad Y = f_Y(T, V, \epsilon_Y), \quad (9)$$

$$\text{where: } Z, V, \epsilon_T, \epsilon_M, \epsilon_Y \text{ are mutually statistically independent.} \quad (10)$$

Model (9)–(10) consists of the four observed variables Z, T, M, Y , three exogenous error terms $\epsilon_T, \epsilon_M, \epsilon_Y$, and an *unobserved confounding variable* V that is the source of endogeneity. Equations (9) define the causal relation among variables: Z causes T , T causes M and Y , and the confounder V causes T, M, Y but not Z . There are no restrictions on the functional forms of f_T, f_M, f_Y in (9).

Model (9)–(10) makes no causal distinction between M and Y . Thus we can interpret Model (9)–(10) as two separated IV models described in *Model I* and *Model II* of Table 1. Our task is to merge these two IV models into a single *mediation model* that incorporates the information that M causes Y .³⁷

A mediation model enables to unpack the mechanism through which T causes Y . If the causal effect of M on Y were identified, then the total effect of T on Y could be expressed as the sum of the effect of T on Y that operates through the causal chain $T \rightarrow M \rightarrow Y$ (the *indirect effect*) and the

³⁵ The IV properties of exclusion restriction (e.g. (6)–(7)) and IV-relevance (e.g. (8)) are necessary but not sufficient to identify the causal effect of T on an outcome Y . An extensive literature exists on the additional assumptions that generate the identification of treatment effects. For example, if T and Z are continuous and we assume linearity, then causal effects can be evaluated by two-stage least squares. Imbens and Angrist (1994) study a binary T and assume a monotonicity criteria that identifies the Local Average Treatment Effect (*LATE*). Vytlačil (2006) studies categorical treatments T and evokes a separability condition of the choice function. Heckman and Pinto (2017) present a monotonicity condition that applies to unordered choice models with multiple treatments. Pinto (2015) investigates identifying assumptions generated by revealed preference analysis. Heckman and Vytlačil (2005) investigate the binary treatment, continuous instruments and assume that the treatment assignment is characterized by a threshold-crossing function. Lee and Salanié (2015) assume a generalized set of threshold-crossing rules. Imbens and Newey (2007); Blundell and Powell (2003, 2004); Altonji and Matzkin (2005) study control function methods characterised by functional form assumptions.

³⁶ See Heckman and Pinto (2015a) for a discussion on the causal relations of the standard IV model.

³⁷ See Online Appendix G for a concise introduction to mediation models. See Appendix Online Appendix H for a discussion on assumptions that are commonly evoked in the literature of mediation analysis.

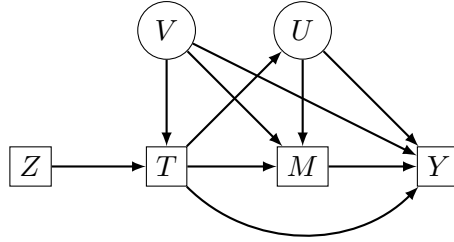
causal effect of T on Y that is not mediated by M (the *direct effect*):

$$\underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}} = \underbrace{\frac{\partial E(Y(t, m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}}. \quad (11)$$

A general mediation model should allow for two sources of endogenous effects: an *unobserved confounder* V that causes T, M, Y and an *unobserved mediator* U that is caused by T and causes M and Y . Panel A of Table 4 represents the general mediation model with instrumental variables as a DAG. Panel B describes the model equations.³⁸

Table 4: The General Mediation Model with IV

A. DAG Representation



B. Model Equations

Treatment variable: $T = f_T(Z, V, \epsilon_T)$,
Unobserved Mediator: $U = f_U(T, \epsilon_U)$,
Observed Mediator: $M = f_M(T, U, V, \epsilon_M)$,
Outcome: $Y = f_Y(T, M, U, V, \epsilon_Y)$,
where: $Z, V, \epsilon_T, \epsilon_M, \epsilon_Y, \epsilon_U$ are mutually statistically independent.

$$M(t) = f_M(t, U(t), V, \epsilon_M) = f_M(t, f_U(t, \epsilon_U), V, \epsilon_M), \quad (12)$$

$$Y(t) = f_Y(t, M(t), U(t), V, \epsilon_Y) = f_Y(t, M(t), f_U(t, V, \epsilon_U), V, \epsilon_Y), \quad (13)$$

$$Y(m) = f_Y(T, m, U, V, \epsilon_Y). \quad (14)$$

The counterfactuals $M(t), Y(t)$ of the general mediation model of Table 4 are given by Equations (12)–(13). The unobserved confounder V induces a correlation between T and these coun-

³⁸ All the model properties discussed in this section also hold if the unobserved confounder V additionally causes the unobserved mediator U .

terfactuals, thus the independence relation $T \perp\!\!\!\perp (Y(t), M(t))$ does not hold. Stated differently, T is endogenous due to V . Nevertheless, $V \perp\!\!\!\perp Z$, which implies that the exclusion restrictions $Z \perp\!\!\!\perp Y(t)$, and $Z \perp\!\!\!\perp M(t)$, still hold. Counterfactual $Y(m)$ in (14) refers to the causal effect of M on Y . Confounder V induces a correlation between M and $Y(m)$ thereby M is endogenous. Moreover, the IV relevance, $Z \not\perp\!\!\!\perp T$, implies that the exclusion restriction $Y(m) \perp\!\!\!\perp Z$ does not hold. In other words, instrument Z does not render the identification of the causal effect of M on Y .

We seek a general mediation model that enables the identification of counterfactual outcomes $Y(t), M(t), Y(m)$ and $Y(m, t)$. Moreover this mediation model should comply with the seven desired properties:

1. The model allows for confounders and unobserved mediators.
2. T and M are endogenous, that is, $T \not\perp\!\!\!\perp (M(t), Y(t))$ and $M \not\perp\!\!\!\perp Y(m, t)$.
3. Instrumental variables Z directly cause T .
4. Model does *not* require a dedicated instrument that directly causes M .
5. Instrument Z is suitable to identify three causal relations: $T \rightarrow Y$; $T \rightarrow M$; and $M \rightarrow Y$.³⁹
6. Additional causal assumptions must be motivated by the economic problem being examined.
7. Additional causal assumption to the general mediation model of Table 4 must be testable.

Properties 1 and 2 simply state that the model must account for potential sources of endogenous effects. Property 3 assures that an instrumental variable that causes T exists. Property 4 states that the model should not require additional instruments that target the mediation variable M . The existence of dedicated IV for M (i.e. an IV that has no impact on Y other than through M and additionally does not impact T) is unlikely in most empirical settings; see discussion on this in section 4.5. We therefore seek a solution for the identification of mediation effects for the prevalent IV model that has dedicated instruments only for T . Property 5 thus follows: a single set of instruments must enable the identification of the three causal effects of interest. Additional assumptions are necessary to obtain identification in this way. Property 6 requires these additional model assumptions to be plausible and have a clear interpretation. Lastly, Property 7 states that these additional assumptions should be testable.

³⁹The first two causal relations are identified in *Models I* and *II* in Table 1.

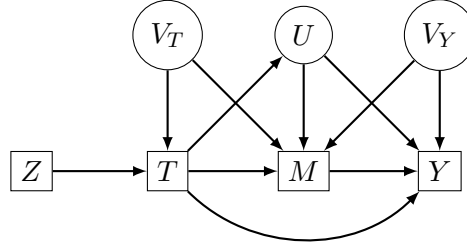
Regarding properties 5 and 6, Section 4.1 describes our model assumptions and examines the identification of these causal effects. We provide a detailed interpretation of our model assumptions in Section 4.2. Regarding property 7, Section 4.3 describes the estimation of causal parameters under linearity. We discuss a model specification test in Section 4.4.

4.1 The Restricted Mediation Model with Instrumental Variables

Table 5 describes the *Restricted Mediation Model* that complies with the seven desirable properties listed in Section 4. The restricted mediation model differs from the general mediation model of Table 4 by decomposing the confounder V into two unobserved variables: V_T that causes T, M and V_Y that causes Y, M .

Table 5: Restricted Mediation Model with IV

A. DAG Representation



B. Model Equations

$$\text{Treatment: } T = f_T(Z, V_T, \epsilon_T),$$

$$\text{Unobserved Mediator: } U = f_U(T, \epsilon_U),$$

$$\text{Observed Mediator: } M = f_M(T, U, V_T, V_Y, \epsilon_M),$$

$$\text{Outcome: } Y = f_Y(T, M, U, V_Y, \epsilon_Y),$$

$$\text{Independence: } V_T, V_Y, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y \text{ are statistically independent.}$$

Equations (15)–(18) list the counterfactual outcomes $M(t), Y(t), Y(m), Y(m, t)$ of the restricted model in Table 5:

$$M(t) = f_M(t, U(t), V_T, V_Y, \epsilon_M) = f_M(t, f_U(t, \epsilon_U), V_T, V_Y, \epsilon_M), \quad (15)$$

$$Y(t) = f_Y(t, M(t), U(t), V_Y, \epsilon_Y) = f_Y(t, f_M(t, U(t), V_T, V_Y, \epsilon_M), U(t), V_Y, \epsilon_Y), \quad (16)$$

$$Y(m) = f_Y(T, m, U, V_Y, \epsilon_Y) = f_Y(T, m, f_U(T, \epsilon_U), V_Y, \epsilon_Y), \quad (17)$$

$$Y(m, t) = f_Y(t, m, U(t), V_Y, \epsilon_Y) = f_Y(t, m, f_U(t, \epsilon_U), V_Y, \epsilon_Y), \quad (18)$$

where $U(t) = f_U(t, \epsilon_U)$.

Treatment T and Mediator M are endogenous variables in both the general and restricted mediation model. According to (15)–(16), confounder V_T induces a correlation between T and counterfactuals $M(t), Y(t)$, thus T is endogenous, i.e., $T \not\perp\!\!\!\perp (Y(t), M(t))$. Counterfactual $Y(m)$ in (17) stands for the outcome Y when the mediator M is fixed at a value $m \in \text{supp}(M)$, while $Y(m, t)$ in (18) denotes the counterfactual outcome Y when M is fixed at $m \in \text{supp}(M)$ and T is fixed at a value $t \in \text{supp}(T)$. $Y(m)$ is a function of random variable T and unobserved confounding variable V_Y , which also causes M . Therefore, $(T, M) \not\perp\!\!\!\perp Y(m)$. $Y(m, t)$ is a function of V_Y and thereby $M \not\perp\!\!\!\perp Y(m, t)$. $Y(m, t)$ is not a function of T , moreover, $V_Y \perp\!\!\!\perp T$ implies that $Y(m, t) \perp\!\!\!\perp T$. The restricted mediation model of Table 5 generates three exclusion restrictions described in Theorem T-1.

Theorem T-1 *The following statistical relations hold for the restricted mediation model of Table 5:*

	Targeted Causal Relation	IV Relevance		Exclusion Restrictions
Property 1	for $T \rightarrow Y$	$Z \not\perp\!\!\!\perp T$	and	$Z \perp\!\!\!\perp Y(t)$
Property 2	for $T \rightarrow M$	$Z \not\perp\!\!\!\perp T$	and	$Z \perp\!\!\!\perp M(t)$
Property 3	for $M \rightarrow Y$	$Z \not\perp\!\!\!\perp M T$	and	$Z \perp\!\!\!\perp Y(m) T$

Proof P-1 See Appendix D.

The exclusion restrictions of Properties 1 and 2 in T-1 are identical to the ones in (6)–(7), generated by the standard IV model described in Model I and Model II of Table 1. These exclusion restrictions arise from the statistical independence between the instrument Z and unobserved confounders V_T, V_Y . Property 3 however states an exclusion restriction, $Z \perp\!\!\!\perp Y(m)|T$, that does not arise from the standard IV model. This implies that Z can be used to evaluate the causal relation of M on Y if (and only if) conditioned on T . Indeed, while $Z \perp\!\!\!\perp Y(m)|T$ holds, $Z \perp\!\!\!\perp Y(m)$ does not.

Properties 1 and 2 of T-1 identify the counterfactual outcomes $M(t)$ and $Y(t)$, i.e. (15) and (16), by applying standard IV techniques. A novel feature of our model is the possibility to use Z to identify the counterfactual outcome $Y(m, t)$, i.e. (18). Property 3 of T-1 is useful to identify of the conditional counterfactual $(Y(m)|T = t)$. Corollary C-1 states that $(Y(m)|T = t)$ is equal in distribution to the counterfactual outcome $Y(m, t)$. Therefore Property 3 of T-1 can be used to identify $Y(m, t)$.

Corollary C-1 *In the restricted mediation model of Table 5, the counterfactual outcome $Y(m)$ conditioned on $T = t$ is equal in distribution to the counterfactual outcome $Y(m, t)$, i.e., $(Y(m)|T = t) \stackrel{d}{=} Y(m, t)$.*

Proof P-2 See [Appendix E](#).

4.2 Understanding the Restricted Mediation Model and its Identification

Property 3 of Theorem T-1 may come as a surprise as it states that Z becomes a valid instrument for M conditional on T , despite the fact that Z directly causes T instead of M . The restricted mediation model imposes a causal restriction on unobserved confounding variables: V_T directly causes T and M , while V_Y causes M and Y , but there is no confounder that directly causes T , M and Y jointly. This section clarifies the novel ideas of the restricted mediation model.

Our task is twofold. First we build the intuition on the causal relations of the restricted mediation model of Table 5: (i) we interpret the model in light of our empirical context; and (ii) we provide familiar examples in labor economics where the assumption fails. Second, we clarify how the restricted mediation model enables the identification of the causal effect of M on Y : (i) we build intuition based a well-known example that uses college distance as an IV to identify the causal effect of college attendance on income; (ii) we apply this intuition to examine the identification of causal relations among trade exposure, labor market variations and political polarization.

Model Interpretation

Understanding Confounder V_T : Confounder V_T stands for unobserved variables that affect a region's trade exposure and labor market variables. The most common concern here relates to domestic demand conditions, which may be either regional or industry-specific. The key for us is that it is indeed plausible that domestic demand conditions simultaneously impact trade and voters only to the extent that they manifest in local wages and employment.

Understanding Confounder V_Y : Confounder V_Y stands for unobserved variables that jointly affect local labor market outcomes and voting decisions. In German data, local politics is a prime candidate for V_Y because "Germany's constitution grants municipalities and districts a high degree of local autonomy, which leads to a highly decentralized industrial policy. Zoning laws allow German municipalities to create commercial districts integrated with regional zoning plans and a

separate law allows them to set local business tax rates” (Schmidt and Buehler, 2007). Another pertinent example may be that workers in some local labor market are impacted by structural decline or by faster automation, which also makes them tilt towards protectionist policies or politicians. The key for us is that such local labor market shocks are uncorrelated with import exposure. Autor, Dorn, and Hanson (2015) show convincing evidence to validate this assumption, which is adopted also by the extensive literature that follows Autor et al. (2013).

When the Identifying Assumption Fails: The model we present is general and can be potentially applied to a wide range of empirical questions. It is therefore important to also discuss research questions for which the identifying assumption of the restricted mediation model is unlikely to hold. Suppose that a researcher is interested in understanding the mechanisms generating earning gaps across college majors. Let earnings be the labor market value of human capital, which comprises three major components: (a) unobserved abilities such as cognition; (b) specific knowledge associated with each college major; and (c) academic performance such as the GPA. The distribution of the unobserved cognition (item (a)) may differ across majors and it is a source of selection bias. The *total effect* of a college major on earnings is decomposed into the *indirect effect* that is mediated by academic performance (item (c)) and the *direct effect* that pertains to the major specific knowledge (item (b)).

In our notation, T stands for the choice of college major, M for academic performance and Y for earnings. Unobserved ability plays the role of the unobserved confounder V . Unobserved ability is likely to directly affect the choice of major T , school performance M and earnings Y . It does not seem plausible that ability impacts earnings only through its effect on the GPA in college. In this case, V should not be split into V_T that causes T, M and V_Y that causes M, Y .

Identification and the Role of the Instrumental Variable

The novel property of the restricted mediation model is that instrument Z is a valid instrument to identify the causal effect of M on Y when conditioned on T , that is, $Z \perp\!\!\!\perp Y(m)|T$. To clarify the intuition of this result, we begin with a well-known example that uses college distance as an IV before applying the intuition to our empirical context.

Understanding the Exclusion Restriction $Z \perp\!\!\!\perp Y(m)|T$:

Suppose a researcher observes that the income of individuals who did varsity sports during

college is substantially higher than the income of their peers. The researcher hypothesizes that this may be because varsity sports contributes to establishing social networks that are useful in the job market. The researcher is interested in identifying the causal effect of college on income that is mediated by these athletic activities.

Consider the stylized model where T is a college indicator, $T = 1$ for college attendance and $T = 0$ otherwise. Outcome Y denotes income during adulthood. As an instrument, Z stands for the distance between the home of prospective students to college. For sake of simplicity, suppose Z takes two values: $Z = 1$ if college is near and $Z = 0$ if college is far. If $Z \perp\!\!\!\perp (Y(1), Y(0))$ holds and we assume that prospective students are more likely to go to college if it is near, then the Local Average Treatment Effect (LATE) of college attendance on income is identified (see [Imbens and Angrist 1994](#)).

Let the mediator M be an indicator that takes value $M = 1$ if a student enrolls in an athletic club and $M = 0$ otherwise. Let the unobserved variable V_T be athletic ability. Students with more athletic ability ($V_T = 1$) are more likely to attend college than students without athletic ability ($V_T = 0$), and upon entering college join an athletics team. In short, V_T causes T and M .

College distance Z and athletic ability V_T are statistically independent, that is, $Z \perp\!\!\!\perp V_T$. However, conditioned on going to college ($T = 1$), students that live far from college ($Z = 0$) are more likely to have athletic ability ($V_T = 1$). In other words, by conditioning on college attendance ($T = 1$), we induce a negative correlation between V_T and college distance Z . Notationally, we have that $Z \not\perp\!\!\!\perp V_T | T = 1$ even though $Z \perp\!\!\!\perp V_T$. Effectively it means that college distance Z becomes a proxy for athletic ability V_T , and therefore an instrument for M .⁴⁰

Application to the Context of Trade, Labor Markets and Voting: We now apply this intuition to the empirical context of import exposure, labor markets and voting. Regarding *Model I* and *Model II* in Table 1, the existing literature has argued that import exposure T is endogenous in a regression of M on T because regional domestic demand conditions (V_T) affect both local import exposure T and local manufacturing employment M . For instance, an industry-specific domestic demand shock will reduce both local import exposure (T) and local employment (M) in regions that are specialized in that industry. The existing literature has not articulated a particular

⁴⁰ At first glance, this result may seem counter-intuitive. Instrument Z directly causes only T and Z is unconditionally independent of V_T . However when T lacks variation (conditioning), then variation in Z induces variation in V_T (induced correlation), which consequently induce variations in M (because V_T causes M).

distinction in the endogeneity concerns applying to a regression of Y on T (Malgouyres, 2014; Feigenbaum and Hall, 2015; Autor et al., 2016). Suppose therefore that industry-specific domestic demand shocks impact voting (Y) only to the extent that they reduce employment. This endogeneity concern then is captured by V_T in our setup. The solution advanced in Autor et al. (2013) is to use other high-wage countries' imports as the basis of an instrument (Z) that is orthogonal to Germany-specific demand conditions, i.e. $Z \perp V_T$.

Consider the example where there is an unobserved negative shock in German demand for certain product categories (V_T). As a result there will be less employment in making these products domestically and also fewer imports of them. German regions that are specialized in those products will therefore see both lower import exposure (T) and lower local employment (M), confounding the causal relationship between T and M . Now consider Australian imports from China as the basis of the instrument (Z). Australian imports are independent of German domestic demand conditions, i.e. the instrument Z is statistically independent of V_T , $Z \perp V_T$, as in the previous college example. However, conditioned on German imports from China (T), high Australian imports in industry j will partly reflect low German demand in j . Again, we have that $Z \not\perp V_T | T = 1$ even though $Z \perp V_T$. In a regression, conditional on German imports (T), higher Australian imports from China (Z) in the same sector therefore 'cause' additional reductions in German employment, by virtue of proxying for negative German demand V_T . (Whether they indeed do so is a question of explanatory power, not econometric identification.)

4.3 The Restricted Mediation Model Under Linearity

We derive useful results under the assumption of linearity. Linearity is commonly invoked in the trade and local labor market literature we refer to, and Autor et al. (2013) and Acemoglu and Restrepo (2017) show that linear relationships arise naturally when aggregating up firm level models. We show that, under linearity, mediation effects can be evaluated using the well-known method of Two-stage Least Squares (2SLS). Let the equations f_T, f_U, f_M, f_Y of Panel B in Table 5 be defined as:

$$T = \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T, \quad (19)$$

$$U = \zeta_T \cdot T + \epsilon_U, \quad (20)$$

$$M = \varphi_T \cdot T + \varphi_U \cdot U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (21)$$

$$Y = \beta_T \cdot T + \beta_M \cdot M + \beta_U \cdot U + \beta_V \cdot V_Y + \epsilon_Y, \quad (22)$$

where $\xi_Z, \xi_V, \zeta_T, \varphi_T, \varphi_U, \delta_Y, \delta_T, \beta_T, \beta_M, \beta_U, \beta_V$ are scalar coefficients, $\epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y$ are unobserved error-terms, Z, T, M, Y are observed, V_T, V_Y, U are unobserved and variables $Z, V_T, V_M, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y$ are mutually independent variables. We also assume that all variables have zero mean and that U, V_T, V_Y have unit variance without loss of generality (see [Online Appendix I](#) for detailed description).

Under Model (19)–(22), the counterfactual variables $M(t), Y(t), Y(m, t)$ are given by:⁴¹

$$M(t) = \Lambda_T^M \cdot t + \epsilon_{M(t)} \quad \text{where } \Lambda_T^M = (\varphi_T + \varphi_U \zeta_T) \quad (23)$$

and $\epsilon_{M(t)} = \varphi_U \cdot \epsilon_U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M$.

$$Y(t) = \Lambda_T^Y \cdot t + \epsilon_{Y(t)} \quad \text{where } \Lambda_T^Y = (\beta_T + \beta_M(\varphi_T + \varphi_U \zeta_T) + \beta_U \zeta_T) \quad (24)$$

and $\epsilon_{Y(t)} = \beta_M \delta_T \cdot V_T + \beta_M \cdot \epsilon_M + \beta_U \cdot \epsilon_U + (\beta_V + \beta_M \delta_Y) \cdot V_Y + \epsilon_Y$.

$$Y(m, t) = \Pi_M^Y \cdot m + \Pi_T^Y \cdot t + \epsilon_{Y(m, t)} \quad \text{where } \Pi_T^Y = (\beta_T + \beta_U \zeta_T), \quad \Pi_M^Y = \beta_M \quad (25)$$

and $\epsilon_{Y(m, t)} = \beta_U \cdot \epsilon_U + \beta_V \cdot V_Y + \epsilon_Y$.

We are interested in identifying four causal parameters $\Lambda_T^Y, \Lambda_T^M, \Pi_T^Y$ and Π_M^Y . Parameter Λ_T^Y in (23) denotes the total effect of T on Y , while Λ_T^M in (24) denotes the effect of T on M . According to Equation (11), Parameter Π_T^Y in (25) stands for the *direct effect* of T on Y , while the *indirect effect* of T on Y is given by the product $\Pi_M^Y \cdot \Lambda_T^M$.⁴² We prove the identification of each causal parameter in [Online Appendix I.2](#).

Property 1 of [T-1](#) suggests that the 2SLS regression of Y on T using Z as instrument estimates Λ_T^Y .⁴³ Equation (26) describe the 2SLS while equation (27) corroborates the estimated parameter.

$$\text{First Stage: } T = \Gamma^T + \Gamma_Z^T \cdot Z + \epsilon^T; \quad \text{Second Stage: } Y = \Gamma^Y + \Gamma_T^Y \cdot T + \epsilon^Y. \quad (26)$$

$$\text{plim}(\hat{\Gamma}_T^Y) = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} = \frac{(\beta_T + \beta_M(\varphi_T + \varphi_U \zeta_T) + \beta_U \zeta_T) \xi_Z}{\xi_Z} = \Lambda_T^Y. \quad (27)$$

By the same token, Property 2 of [T-1](#) suggests that the 2SLS of M on T using Z as instrument

⁴¹See [Online Appendix I](#) for a detailed derivation.

⁴²See [Online Appendix I.1](#) for further discussion on these causal parameters.

⁴³See [Online Appendix J](#) for a detailed discussion on the estimation of the causal parameters of Model (19)–(22).

estimates Λ_T^M . Equation (28) describe this 2SLS and Equation (29) confirms the estimated parameter.^{44,45}

$$\text{First Stage: } T = \Gamma^T + \Gamma_Z^T \cdot Z + \epsilon^T; \quad \text{Second Stage: } M = \Gamma^M + \Gamma_T^M \cdot T + \epsilon^M. \quad (28)$$

$$\text{plim}(\hat{\Gamma}_T^M) = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} = \frac{(\varphi_T + \varphi_U \zeta_T) \cdot \xi_Z}{\xi_Z} = \Lambda_T^M. \quad (29)$$

Property 3 of **T-1** suggests that the causal effect of M on Y can be estimated by the 2SLS of Y on M conditioned on T that uses Z as instrument. Equations (30)–(31) specify this 2SLS regression. Equations (32)–(32) show that the causal effect of M on Y , that is, Π_M^Y , is estimated by the coefficient of variable M in the second stage (31). Equations (33)–(34) explain that the causal parameter Π_T^Y is estimated by the coefficient of variable T in the second stage (31).⁴⁶

$$\text{First Stage: } M = \Gamma^{M|T} + \Gamma_Z^{M|T} \cdot Z + \Gamma_T^{M|T} \cdot T + \epsilon^{M|T}; \quad (30)$$

$$\text{Second Stage: } Y = \Gamma^{Y|T} + \Gamma_M^{Y|T} \cdot M + \Gamma_T^{Y|T} \cdot T + \epsilon^{Y|T}. \quad (31)$$

$$\text{plim}(\hat{\Gamma}_M^{Y|T}) = \frac{\text{cov}(T, Z) \cdot \text{cov}(T, Y) - \text{cov}(T, T) \cdot \text{cov}(Z, Y)}{\text{cov}(M, T) \cdot \text{cov}(T, Z) - \text{cov}(T, T) \cdot \text{cov}(Z, M)} \quad (32)$$

$$= \frac{\text{cov}(V_T, V_T) \cdot \beta_M \cdot \delta_T \cdot \xi_Z}{\text{cov}(V_T, V_T) \cdot \delta_T \cdot \xi_Z} = \beta_M = \Pi_M^Y; \quad (33)$$

$$\text{plim}(\hat{\Gamma}_T^{Y|T}) = \frac{-(\text{cov}(M, Z) \cdot \text{cov}(T, Y) - \text{cov}(M, T) \cdot \text{cov}(Z, Y))}{\text{cov}(M, T) \cdot \text{cov}(T, Z) - \text{cov}(T, T) \cdot \text{cov}(Z, M)} \quad (34)$$

$$= \frac{\text{cov}(V_T, V_T) \cdot \delta_T \cdot \xi_Z \cdot (\beta_T + \beta_U \cdot \zeta_T)}{\text{cov}(V_T, V_T) \cdot \delta_T \cdot \xi_Z} = (\beta_T + \beta_U \cdot \zeta_T) = \Pi_T^Y. \quad (35)$$

It is worth pointing out that conditioning on T in the first stage is key to identification. In Section 4.1 we explain that the exclusion restriction $Z \perp\!\!\!\perp Y(m)$ does not hold for the mediation model of Table 5. Thus we should expect that the 2SLS of M on Y which uses Z as an instrument does not render a causal parameter. Equation (36) reports this 2SLS estimate, which does not have

⁴⁴Online Appendix J describes the estimation of this causal parameter in greater detail.

⁴⁵Equations (26) and (28) correspond to equations (3) and (4) in section 3.1 and (5) in section 3.2.

⁴⁶See Online Appendix I.2 for a detailed derivation of Equations (32)–(35). See Online Appendix J for a description of the 2SLS estimation.

a clear causal interpretation.⁴⁷

$$\frac{\text{cov}(Z, Y)}{\text{cov}(Z, M)} = \frac{(\varphi_T + \varphi_U \zeta_T) \cdot \xi_Z}{(\beta_T + \beta_M(\varphi_T + \varphi_U \zeta_T) + \beta_U \zeta_T) \cdot \xi_Z}, \quad (36)$$

4.4 A Model Specification Test

The restricted mediation model is build by imposing restrictions on the causal links of the unobserved confounder V of the general mediation model. In this section we discuss a simple method to test these causal restrictions.

Table 6 displays the restricted and general model as DAGs (Panel A) and the linear equations that subsume the causal relations of each model (Panel B). Our null hypothesis is that the data generating process conforms to the restricted mediation model. Our alternative hypothesis is that the general mediation model holds. Effectively, we test if the causal assumption that splits the confounder V of the general model into V_T and V_Y in the restricted model is empirically sound. To do so, we invoke the linearity assumption of Section 4.3.⁴⁸

Table 6: Restricted and General Mediation Model with one Instrumental Variable

A. DAG Representation



B. Linear Equations

$T = \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T$	$T = \xi_Z \cdot Z + \xi_V \cdot V + \epsilon_T$
$U = \zeta_T \cdot T + \epsilon_U$	$U = \zeta_T \cdot T + \epsilon_U$
$M = \varphi_T \cdot T + \varphi_U \cdot U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M$	$M = \varphi_T \cdot T + \varphi_U \cdot U + \delta \cdot V + \epsilon_M$
$Y = \beta_T \cdot T + \beta_M \cdot M + \beta_U \cdot U + \beta_V \cdot V_Y + \epsilon_Y$	$Y = \beta_T \cdot T + \beta_M \cdot M + \beta_U \cdot U + \beta_V \cdot V + \epsilon_Y$

Panel A presents the Directed Acyclic Graphs (DAG) of the restricted and general models. Panel B presents the equations that subsume the causal relations described in each model under the assumption of linearity.

⁴⁷This can be seen quite easily in *Model III* in Table 1: When Z is used as an instrument for M the exclusion restriction is quite obviously violated via Z 's effect on T which in turn impacts Y through channels other than M .

⁴⁸Our aim is not to test linearity itself, but rather to use linearity to do inference on the causal assumptions of the restricted model.

The identification of coefficients in linear models depends on the equations governing the covariance structure of observed data. Our test explores the differences in these equations between the restricted and the general model. [Online Appendix L](#) shows that Equalities (37)–(38) must hold for the restricted model.

$$\text{cov}(\mathbf{Z}, Y) - \text{cov}(\mathbf{Z}, M) \cdot \beta_M - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\beta}_T = 0 \quad (37)$$

$$\text{cov}(T, Y) - \text{cov}(T, M) \cdot \beta_M - \text{cov}(T, T) \cdot \tilde{\beta}_T = 0, \quad (38)$$

where $\tilde{\beta}_T = \varphi_T + \varphi_U \zeta_T$. Equalities (37)–(38) enable the identification of parameters $\tilde{\beta}_T$ and β_M . If the instrument Z consists of a single variable, then parameters $\tilde{\beta}_T$ and β_M are exactly identified and our model specification test does not apply (see [Online Appendix L.1](#)). If Z consists of two (or more) variables, then parameters $\tilde{\beta}_T$ and β_M are over-identified and Equation (38) constitute an over-identification restriction (see [Online Appendix L.2](#)).⁴⁹ The respective equalities for the case of the general mediation model are derived in [Online Appendix L.3](#) and are given by:

$$\text{cov}(\mathbf{Z}, Y) - \text{cov}(\mathbf{Z}, M) \cdot \beta_M - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\beta}_T = 0 \quad (39)$$

$$\text{cov}(T, Y) - \text{cov}(T, M) \cdot \beta_M - \text{cov}(T, T) \cdot \tilde{\beta}_T = \beta_V \cdot \xi_V. \quad (40)$$

Our inference explores the fact that the over-identification restriction (38) is equal to zero for restricted model while restriction (40) differs from zero in the case of the general model. It can be interpreted as a particular application of the SarganHansen test of overidentifying restrictions. Recall that model variables have mean zero w.l.o.g., thus we can express Equalities (37)–(38) as the following moment conditions:

$$E(\mathbf{Z} \cdot (Y - \beta_M \cdot M - \tilde{\beta}_T \cdot T)) = 0 \quad \text{for} \quad \text{cov}(\mathbf{Z}, Y) - \text{cov}(\mathbf{Z}, M) \cdot \beta_M - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\beta}_T = 0 \quad (41)$$

$$\text{and } E(T \cdot (Y - \beta_M \cdot M - \tilde{\beta}_T \cdot T)) = 0 \quad \text{for} \quad \text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = 0. \quad (42)$$

If multiple dedicated instruments for T exist, then parameters $\beta_M, \tilde{\beta}_T$ can be consistently estimated using Moment Condition (41). This condition is associated with Equations (37),(39) and thereby holds for both restricted and general models. On the other hand, Moment Condition (42) is valid only for the restricted model. Thus a model specification test consists on verifying if the model estimates comply with Moment Condition (42). Large absolute values of Moment Condition (42) constitute statistical evidence contrary to the restricted mediation model and the null

⁴⁹ Our test bears some similarities with the the Sargan-Hansen test that exploits model over-identifying restrictions to do inference on model coefficients.

hypothesis is rejected.

Moment Conditions (41)–(42) can be used to estimate parameters $\beta_M, \tilde{\beta}_T$ via the Generalized Method of Moments (GMM) of Hansen (1982). In practice, we evaluate two sets of GMM estimators for parameters $\beta_M, \tilde{\beta}_T$. The first estimator is based only on Moment Condition (41). The second estimator relies on both moment conditions (41) and (42). Large absolute differences between these estimates provide statistical evidence against the null hypothesis that data is generated by the restrictive model. Thus we then perform a Wald test on the null hypothesis that the two GMM estimates are equal. See Online Appendix L.4 for a description of the inference procedure.

We implement this test in Section 5.

4.5 Exploring Alternative Approaches and Related Literature

We investigate the mediation model in which the treatment variable T and the mediator variable M are endogenous. Our solution imposes causal relations among unobserved variables that enable the identification of three causal effects using only one dedicated instrument for T .

Our method contrasts to two broad alternative approaches to gaining identification in mediation analysis. One of these is to assume that the treatment T and the mediator M are exogenous given observed variables (Imai, Keele, and Yamamoto, 2010; Imai, Keele, Tingley, and Yamamoto, 2011b; Imai et al., 2011a).⁵⁰ In this case, treatment T is as good as randomly assigned and the resulting model is equivalent to assuming no confounding variables V and no unobserved mediators U in Model III of Table 1.⁵¹ Relatedly, Yamamoto (2014) studies the case of a binary treatment indicator and a single instrument, assuming that the instrument Z is independent of the counterfactual outcome $Y(m, t)$ and that the mediator variables is exogenous conditioned on treatment compliance.⁵²

A second class of models relies on additional instrumental variables dedicated to the mediator M . Powdthavee (2009); Burgess, Daniel, Butterworth, and Thompson (2015) and Jhun (2015)

⁵⁰Robins and Greenland (1992) and Geneletti (2007) consider instruments that perfectly correlate with the mediator variable such that the exogeneity condition still holds.

⁵¹If the treatment T were indeed randomly assigned, then one could use the interaction of the treatment with observed covariates as instruments to identify the causal effect of M on Y . Versions of this approach are examined in Ten Have, Joffe, Lynch, Brown, Maisto, and Beck (2007); Dunn and Bentall (2007); Small (2012); Gennetian, Bos, and Morris (2002).

⁵²In our notation, this means that $Y(m, t) \perp\!\!\!\perp Z$ and $Y(t, m) \perp\!\!\!\perp M(t) | (T, P = c)$, where T denotes treatment assignment and P stands for an indicator of treatment compliance. Neither assumption holds in Model III or Model IV of Table 1.

achieve identification using two instruments and parametric assumptions that shape the endogeneity of T and M . Two important contributions to this literature that use non-parametric identification are [Frolich and Huber \(2017\)](#) and [Jun et al. \(2016\)](#).⁵³ This second class of models does not assume away confounding effects; i.e. variables T, M, Y remain endogenous. It thus constitutes an alternative approach to our identification problem, which is to seek for another instrument that is dedicated to M .⁵⁴ Because of its natural appeal, we discuss this approach here and contrast its identification requirements explicitly to ours. A standard mediation model with two instrumental variables is described as follows:

$$\text{Treatment variable: } T = f_T(Z_T, V, \epsilon_T), \quad (43)$$

$$\text{Observed Mediator: } M = f_M(T, Z_M, V, \epsilon_M), \quad (44)$$

$$\text{Outcome: } Y = f_Y(T, M, V, \epsilon_Y), \quad (45)$$

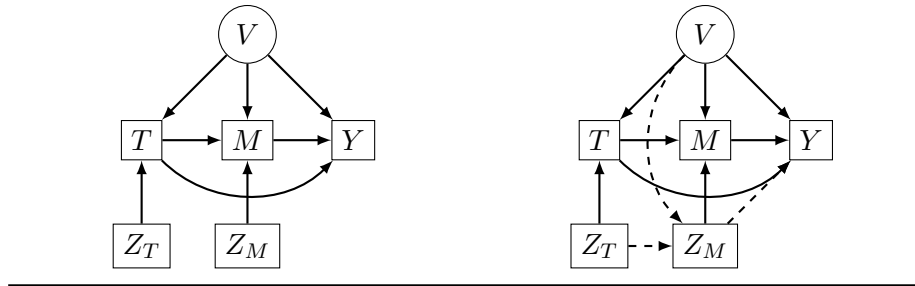
$$\text{where: } (Z_T, Z_M) \perp\!\!\!\perp V. \quad (46)$$

This model is presented as a DAG in Table 7. In this model, the exclusion restriction $Z_M \perp\!\!\!\perp Y(m)$ and also $Z_M \perp\!\!\!\perp Y(m)|T$ hold. Thereby Z_M can be used to evaluate the causal effects of M on Y .⁵⁵

Table 7: General Mediation Model and Violation of Exclusion Restrictions

A. Directed Acyclic Graph (DAG) Representation

General IV Model with Two Instruments Violations of the Exclusion Restriction



The left figure gives the directed acyclic graph (DAG) representation of the general IV Model with two dedicated instruments. The right figure gives the same DAG, but also depicts the identification concerns discussed in the body of the text.

⁵³Both papers examine the effect of a binary indicator for treatment T . [Frolich and Huber \(2017\)](#) relies on two dedicated instruments (for T and M) and a monotonicity restriction with respect to M . [Jun et al. \(2016\)](#) uses three dedicated instruments but does not require the monotonicity restriction.

⁵⁴Recently, [Frolich and Huber \(2017\)](#) provide an important contribution on the mediation model with two dedicated instruments.

⁵⁵If T were to cause Z_M , then only $Z_M \perp\!\!\!\perp Y(m)|T$ would hold.

The empirical challenge in evaluation Model (43)–(46) is to find a suitable candidate for Z_M . There are three potential concerns with any dedicated instrument for M : (i) Z_M may correlate with V ; (ii) Z_M may directly affect Y ; and (iii) Z_M may correlate with Z_T . Concerns (i) and (ii) define the usual requirements for any valid instrument to identify the effect of M on Y . The latter concern (iii) is specific to the mediation context. The three concerns are depicted as dashed errors in the right figure of Table 7.

A potential candidate for Z_M is automation. Automation, i.e. replacing workers with machines, robots and computer-assisted technologies, is usually viewed as the ‘other big shock’ that has hit high-wage labor markets in the last decades. For example, [Acemoglu and Restrepo \(2017\)](#) estimate that an additional robot per thousand workers has reduced employment in the U.S. by about 0.18–0.34 percentage points and wages by 0.25–0.50 percent. The effects of automation are not expected to abate. For example, [Frey and Osborne \(2017\)](#) predict that 47% of U.S. workers are at risk of automation over the next two decades. In brief, automation has had and will likely continue to have substantial effects on labor market outcomes M and therefore seems like a good candidate dedicated instrument Z_M .

We view concern (iii) as addressed in this context because [Autor et al. \(2015\)](#) have provided convincing evidence that automation and import exposure are largely orthogonal, making the two forces separable in the data at both the industry-level and the regional level. Concern (i) still is that firms may automate in response to other unobserved factors that could directly impact their labor demand. Indeed, firm-level technology upgrading does appear to respond to the China shock as shown by [Bloom, Draca, and Van Reenen \(2016\)](#). This violates the independence $Z_M \perp\!\!\!\perp V$ in (46) and thereby the exclusion restriction $Z_M \perp\!\!\!\perp Y(m)|T$ does not hold. However, this concern may again be largely addressed if we think of Z_M not as actually measured technology upgrading but as some more exogenous measure, e.g. *exposure to robot adoption* as in [Acemoglu and Restrepo \(2017\)](#) or employment-weighted occupational measures like *routine task intensity* ([Autor and Dorn, 2013](#)) or *automatability* ([Frey and Osborne, 2017](#)).

In our empirical context, concern (ii)—automation could impact voting behavior through channels other than M —is the most worrisome, and in fact clearly disqualifies automation as a dedicated instrument for M . While a German assembly-line worker will likely neither observe nor care about Australian imports of Chinese consumer electronics (i.e. Z_T), he/she will not only be

aware of the potential automatability of their assembly-line job (i.e. Z_M) but may indeed seek out a more protectionist political agenda in anticipation of automation's consequences, i.e. even before any detrimental effects in the labor market.

5 Model IV: Mediation Analysis

5.1 Applying the Identification Framework

We now apply *Model IV* of Table 1, i.e. the mediation model outlined in section 4. This allows us to estimate the *mediated effect* of import exposure on voting that runs through labor markets. The extent to which trade exposure polarized voters because it caused labor market adjustments is identified by a comparison of this *mediated* (or *indirect*) effect with the *total effect* of trade exposure on voting. The total effect of T on Y in *Model I* is estimated as $\hat{\Gamma}_T^Y = 0.089$, reported in table 2.⁵⁶ The mediated effect is the product of T on M in *Model II*, estimated as $\hat{\Gamma}_T^M = -0.024$, reported in table 2 times the effect of M on Y , estimated here.

As shown in section 4.3, we estimate $\hat{\Gamma}_M^{Y|T}$ in the following second stage equation:

$$Y_{it} = \Gamma_T^{Y|T} \cdot T_{it} + \Gamma_M^{Y|T} \cdot M_{it} + \Gamma_X^{Y|T} \cdot X_{it} + \epsilon_{it}^{Y|T}. \quad (47)$$

The First Stage now takes the form

$$M_{it} = \Gamma_{IM}^{M|T} \cdot Z_{it}^{IM} + \Gamma_{EX}^{M|T} \cdot Z_{it}^{EX} + \Gamma_T^{M|T} \cdot T_{it} + \Gamma_X^{M|T} \cdot X_{it} + \epsilon_{it}^{M|T}. \quad (48)$$

Equation (47) differs from the second stage equation (3) in that we are interested in the causal effect of the mediator M on outcome Y . Equation (48) differs from the first stage equation (4) in that trade exposure T is now included. It is worth noting that the coefficient $\Gamma_T^{Y|T}$, i.e. the *direct effect* of trade on voting that does not operate through M , is not identified as a causal parameter in (47). Instead, it is identified only as the residual of three identified coefficients, namely $\hat{\Gamma}_T^Y - \hat{\Gamma}_M^{Y|T} \times \hat{\Gamma}_T^M$.

Table 8 summarizes the results of applying our mediation model. The top panel reports on

⁵⁶The final outcome of focus is Y , the vote share of extreme-right parties, i.e. the only significant voting response to trade observed in the data.

estimating the first stage equation (48). The middle panel reports on estimating the second stage equation (47). The bottom panel reports p-values of the specification test proposed in section 4.4.

It is helpful to approach the table with some priors about expected signs. As discussed, the literature worries about domestic industry-specific demand conditions as a source of confounding bias V_T : German industries that experience negative domestic demand shocks will see fewer imports and less employment. German industries that experience positive domestic demand shocks will see fewer exports and less unemployment.⁵⁷

The discussion in section 4.2 implies that $Z|T$ can serve as a proxy for V_T . This suggests that, conditional on Germany's imports T , other countries' imports ($Z|T$) should "cause" additional adjustments in German labor markets. Indeed, this is what we find in the top-panel of table 8. Other countries' imports from China 'reduce' German employment ($\hat{\Gamma}_{IM}^{M|T} = -0.005$). By the same logic, other countries' exports to China 'increase' it ($\hat{\Gamma}_{EX}^{M|T} = 0.006$).⁵⁸

The middle panel of table 8 is the most important one, reporting estimates from the Second Stage equation (47). The causal effect of a trade-induced drop in employment on the extreme right's vote share is highly significant, with a p-value below 0.001. The point estimate $\hat{\Gamma}_M^{Y|T}$ indicates that a one-percent drop in employment raises the change in the extreme right's vote share by 0.06 percentage points (6.224/100). This is broadly consistent in magnitude with our finding in table 2 that a one-standard-deviation increase in T (€1,350 per worker) increases the extreme-right vote share by 0.12 percentage points. The product with $\hat{\Gamma}_T^M$, reported in table 3, is $-6.224 \times -0.024 = 0.149$, and is highly significant.

The percentage of the total populist effect that is explained by labor markets is $(\hat{\Gamma}_M^{Y|T} \cdot \hat{\Gamma}_T^M) / \hat{\Gamma}_T^Y$. It ranges from 141% to 175% across columns 1–5. This implies that the *mediated effect* of import exposure on voting that runs through labor markets and the residual *direct effect* of import exposure on voting have opposite signs, the direct effect being in the aggregate politically moderating.⁵⁹ In other words, if the only effect of import exposure was to decrease employment, the political

⁵⁷Consistent with this, the 2SLS estimates $\hat{\Gamma}_T^M$ in table 3 are larger than the OLS estimates.

⁵⁸As before, we use two separate instruments Z_{it}^{IM}, Z_{it}^{EX} which is necessary for the model specification test.

⁵⁹While we can only speculate, other channels that are potentially moderating are import exposure's effect on task-upgrading (Becker and Muendler, 2015), and switching production towards more differentiated and higher mark-up varieties (Holmes and Stevens, 2014). Other channels that could work in the opposite direction may involve anxiety about the future (Mughan and Lacy, 2002; Mughan et al., 2003) or a general aversion to changes in the status quo economic structure. Of course, one can think of some of these also as labor market channels broadly speaking. The more important point here is that they are not part of the readily available labor market aggregates usually studied.

Table 8: The Mediation Model in Practice

	(1)	(2)	(3)	(4)	(5)
<u>First Stage /Z:</u>					
$\widehat{\Gamma}_{IM}^{M T}$	-0.002 [0.353]	-0.004* [0.081]	-0.004* [0.045]	-0.005** [0.015]	-0.005** [0.015]
$\widehat{\Gamma}_{EX}^{M T}$	0.005** [0.012]	0.005** [0.012]	0.005*** [0.002]	0.006*** [0.000]	0.006*** [0.000]
$\widehat{\Gamma}_T^{M T}$	-0.007 [0.145]	-0.005 [0.275]	-0.002 [0.574]	0.000 [0.937]	0.000 [0.959]
R-Squared	0.357	0.442	0.457	0.531	0.573
<u>Second Stage /Z:</u>					
$\widehat{\Gamma}_M^{Y T}$	-7.156** [0.016]	-5.694** [0.026]	-6.099** [0.013]	-6.164*** [0.000]	-6.224*** [0.000]
$\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{Y T}$	0.166* [0.066]	0.135* [0.071]	0.153* [0.049]	0.151*** [0.010]	0.149** [0.013]
% Effect Mediated by M :	141	136	135	176	167
<u>Specification Test</u>					
test: $\widehat{\Gamma}_M^{Y T} = \widehat{\Gamma}_{M,Alt}^{Y T}$	[0.745]	[0.712]	[0.600]	[0.762]	[0.882]
test: $\widehat{\Gamma}_T^{Y T} = \widehat{\Gamma}_{T,Alt}^{Y T}$	[0.672]	[0.911]	[0.623]	[0.765]	[0.876]

Notes: (a) Columns 1–5 introduce controls in the same way as table 2. *p-values* are reported in square brackets, with standard errors are clustered at the level of 96 commuting zones. (b) The top panel presents results of the the first-stage equation (48). Other countries’ imports from China “reduce” German employment ($\widehat{\Gamma}_{IM}^{M|T} = -0.005$), and exports “increase” it ($\widehat{\Gamma}_{EX}^{M|T} = 0.006$). (c) The middle panel is the key one, reporting estimates from the second-stage equation (47). The causal effect of a trade-induced drop in employment on the extreme right’s vote share is highly significant, with a *p-value* below 0.001. The point estimate $\widehat{\Gamma}_M^{Y|T}$ indicates that a one-percent drop in employment raises the change in the extreme right’s vote share by 0.06 percentage points (6.224/100). The product with $\widehat{\Gamma}_T^M$, reported in table 3, is $-6.224 \times -0.024 = 0.149$, which is 167% of the *total effect* $\widehat{\Gamma}_T^Y$, which is reported as 0.089 in table 2. (d) The bottom panel reports on the specification test proposed in section 4.4. It reports the *p-value* of the hypothesis test that $\widehat{\Gamma}_M^{Y|T}$ and $\widehat{\Gamma}_T^{Y|T}$ are the same whether they are estimated under the *Restricted Mediation Model* or the alternative *General Model*. This specification test does not reject our model assumptions.

response would be even be stronger than the one observed in the data.

The bottom panel of table 8 reports on the specification test proposed in section 4.4. Our null hypothesis is that the causal assumptions of the *Restricted Mediation Model* hold. Under the null, the estimates from the *Restricted Mediation Model* and the alternative *General Model* are both consistent. Thereby we test if the difference in point estimates is statistically significant using a Wald test that is based on the χ^2 -statistic associated with this difference. Estimates are generated by a two-step GMM estimator whose errors are clustered by region. We use the same set of conditioning variables described in the main paper. In all specification tests we do not reject the null hypothesis that the causal assumptions of the restricted model hold.

5.2 Additional Observed Mechanisms

As an extension, we now turn to a wider range of observed labor market outcomes which have been found to be affected by trade exposure in the existing literature (Autor et al., 2013; Dauth et al., 2014). Some of these outcomes are strongly correlated with total employment but may have independent effects on the final voting outcome Y . In addition to (i) total employment, we consider the following additional mediators: (ii) manufacturing’s employment share, (iii) manufacturing wages, (iv) non-manufacturing wages, and (v) unemployment. In Online Appendix M we show that (ii) and (iii) are indeed significantly impacted by import competition, while the results for (iv) and (v) is weak. This is consistent with prior research that clearly shows that the labor market effects of import exposure are concentrated in manufacturing employment.

In turn, when we separately apply our mediation framework to (ii) and (iii), we estimate significant mediated effects. Indeed, the mediated effect that we estimate when we apply our framework to (ii) is practically identical to that reported in section 5.2 for (i). We thus have several observed mechanisms, that are estimated to have significant mediating effects, but that are highly correlated. How is one to go about aggregating these results?

Without additional dedicated instruments, our method can only identify the effect of as many mechanisms as there are treatments. We therefore need to aggregate the multiple mechanisms into a single index. A *principal component analysis* is attractive in this regard because it generates indices that are purely statistical measurements based on the total variation in labor market outcomes

and are orthogonal to one another by construction.⁶⁰ This approach is appealing as long as the mediating effects are sharply concentrated in a single principal component, and this principal component has a clear interpretation. We label the principal components as our ‘labor market components’ (LMC).

One can best interpret the LMCs through their relation to the labor market outcomes we observe, specifically through their factor loadings. Panel B of table 9 reports on the factor loadings of the LMCs. We follow the convention of reporting these for the LMCs with an eigenvalue larger than 1. In our data, LMC₁ and LMC₂ together explain about 80 percent of the variation in the labor market data ($0.541 + 0.256$).⁶¹

LMC₁ is somewhat ambiguous: it loads positively on changes in wages, but negatively on manufacturing employment and also positively on unemployment.⁶² By contrast, LMC₂’s interpretation is unambiguous: Its factor loadings are strongly positive for changes in manufacturing employment and total employment, and negative for changes in unemployment. LMC₂ is clearly associated with the labor market aspects that we know to be most affected by import exposure.

The top-half of Panel A of table 9 reports the results of re-estimating equation (5) for the LMCs one at a time, i.e. the equivalent of the results reported for total employment in table 3. Here, we do report results for all five specifications to illustrate the sharpness induced by this approach.⁶³ This sharpness is indeed striking: Comparing the results for LMC₂ to those of the other four, it is the only one significantly impacted by import competition in all specifications, and the p-value is below 0.001 in all specifications except the first. By contrast, none of the other four LMCs are at all impacted by import competition. If one agrees with our above interpretation of the LMCs, then these results resonate closely with Autor et al. (2015) who show that trade exposure has had large effects on (overall and manufacturing) employment whilst the polarization of work and the rise

⁶⁰By contrast, methods that take weighted averages (Christensen and Miguel, 2016; Kling, Liebman, and Katz, 2007) are usually applied to creating an outcome-index, but are unattractive for creating a mediating variable index precisely because they pre-impose weights. Similarly, *factor analysis* is more suitable when there are strong priors on how to group variables (Heckman, Pinto, and Savelyev, 2013).

⁶¹ The third LMC only explains about 10 percent. The remaining components are reported in Table 9 in Online Appendix M.

⁶²Our interpretation of LMC₁ is that it reflects the polarization of high-wage countries’ labor markets (Goos, Manning, and Salomons, 2009, 2014), associated with both higher wages and higher unemployment. A related view on LMC₁ is provided by the urban agglomeration literature, where Duranton and Puga (2005) point out that regional specialization has become “functional” as opposed to “sectoral” over the last decades, implying a tendency for headquarters and business services to cluster in large cities, a trend that appears to be clearly borne out in Germany (Bade, Laaser, and Soltwedel, 2003).

⁶³By construction, there are as many principal components as there are mediating variables.

Table 9: Mediation Analysis With Principal Components

Panel A. Estimating Second-Stage Equations (5) and (47)

		(1)	(2)	(3)	(4)	(5)
		IV	IV	IV	IV	IV
<i>Second Stage: Effect of T on M</i>						
$\widehat{\Gamma}_T^M$	M: LMC ₁	-0.033 [0.489]	0.017 [0.653]	0.019 [0.625]	0.028 [0.462]	0.035 [0.308]
	M: LMC ₂	-0.264*** [0.003]	-0.298*** [0.000]	-0.333*** [0.000]	-0.329*** [0.000]	-0.329*** [0.000]
	M: LMC ₃	-0.028 [0.517]	-0.036 [0.385]	-0.046 [0.297]	-0.048 [0.242]	-0.060* [0.077]
	M: LMC ₄	0.024 [0.409]	-0.003 [0.887]	-0.026 [0.406]	-0.027 [0.386]	-0.020 [0.505]
	M: LMC ₅	-0.016 [0.433]	-0.008 [0.711]	0.010 [0.629]	0.010 [0.636]	0.008 [0.718]
<i>Second Stage: M on Y, conditional on T</i>						
$\widehat{\Gamma}_M^{Y T}$	M: LMC ₁	-1.237 [0.016]	-2.428 [0.220]	-3.341 [0.297]	-2.904 [0.328]	-2.042 [0.238]
$\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{Y T}$		0.041 [0.506]	-0.042 [0.673]	-0.064 [0.658]	-0.081 [0.556]	-0.072 [0.440]
$\widehat{\Gamma}_M^{Y T}$	M: LMC ₂	-0.424** [0.041]	-0.532** [0.029]	-0.493** [0.014]	-0.460*** [0.002]	-0.447*** [0.003]
$\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{Y T}$		0.112* [0.091]	0.159* [0.061]	0.164** [0.039]	0.151** [0.016]	0.147** [0.019]
$\widehat{\Gamma}_M^{Y T}$	M: LMC ₃	-0.193 [0.573]	-0.308 [0.273]	-0.337 [0.229]	-0.109 [0.635]	-0.097 [0.687]
$\widehat{\Gamma}_M^{Y T}$	M: LMC ₄	2.079 [0.090]	4.919 [0.490]	2.535 [0.575]	0.349 [0.779]	-0.073 [0.963]
% Effect Mediated by LMC ₂ :		95	160	145	176	165

Panel B. The Two Main Principal Components' Eigenvector

	Eigen-value	Eigen-value: Proportion	$\Delta \ln$ (Total Empl.)	Δ Share Manuf. Empl.	$\Delta \ln(\text{Avg}$ Manuf. Wage)	$\Delta \ln(\text{Avg}$ Non-Man Wage)	Δ Share Unempl
LMC ₁	2.707	0.541	0.1711	-0.3632	0.5108	0.5486	0.5261
LMC ₂	1.281	0.256	0.7625	0.6004	0.2104	0.0607	-0.1012

Notes: (a) Panel A columns 1–5 introduce controls in the same way as table 2. *p-values* are reported in square brackets, standard errors are clustered at the level of 96 commuting zones. (b) The top half of Panel A reports the results of estimating equation (5) for the five principal components. There is a strikingly sharp division in import competition's impact, which is concentrated entirely on LMC₂. (c) The bottom half of Panel A reports the results of estimating equation (47) for the first four principal components (omitting the unimportant LMC₅ for brevity). There is again a strikingly sharp division between the significant causal effect of LMC₂ on voting, and the other LMCs. To conserve space, we omit $\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{Y|T}$ for the insignificant LMC₃–LMC₄. (c) Panel B reports on the factor loadings of the five labor market variables on LMC₁ and LMC₂. See discussion in text.

of service jobs (i.e. our LMC_1) were explained by other factors, primarily automation.

The bottom-half of Panel A of table 9 reports the results of estimating (47) for the first four principal components, i.e. the equivalent of the middle panel of table 8. We omit the fifth to conserve space. Unsurprisingly, given the sharpness of the patterns in the top panel, only LMC_2 is estimated to have a significant causal effect on voting. What is more, the mediated effect $\hat{\Gamma}_T^M \times \hat{\Gamma}_M^{Y|T}$ of LMC_2 ($-0.329 \times -0.447 = 0.147$) is practically identical to the 0.149 reported for total employment in table 8.

In summary, table 9 shows that the impact of import competition (T) on German labor markets was entirely focused on the LMC_2 , and that LMC_2 in turn captures *all* of the effect of T on extreme-right voting Y that works through *any* observed labor market channels. Overall, there is a consistent pattern whereby the *mediated effect* of import exposure on voting that runs through labor markets and the residual *direct effect* of import exposure on voting have opposite signs. As a result, if the only effect of import exposure on voting ran through measured labor adjustments, the political response would be even stronger than the one observed in the data.

6 Conclusion

A substantial body of recent evidence suggests that in high-wage manufacturing countries like Germany and the U.S., import exposure has had significant detrimental effects on the labor market outcomes of manufacturing workers. In this paper we show that import exposure has also induced voters to turn to a protectionist, populist, and nationalist policy agendas represented by Germany’s extreme right. The focus of our paper is to ask whether the effect of import exposure on voting for the extreme right is explained by (mediated by) import exposure’s effect on labor markets. There is good reason to believe it is: The aggregate effects coincide in the data and are mirrored in an individual-level analysis where those most prone to tilt towards the right are also those most vulnerable to the labor market consequences of import exposure.

In trying to answer this question, we face an empirical problem that is common to many research settings: Even though we can use standard IV methods to causally identify the effect of a treatment (import competition T_{it}) on a final outcome (voting Y_{it}) and we can causally identify the effect of T on a proposed mechanism outcome (total employment M_{it}), we *cannot* identify how

much of the former is explained by the latter. To make headway, we develop a new methodology that allows us to perform the required *mediation analysis* in an IV setting. Applying our method, we find that the effect of import exposure that is mediated by labor market adjustments is larger than the total effect of import exposure on extreme-right voting, which in turn implies that other channels that link import exposure to voting (the ‘residual effect’) are moderating in the aggregate.

Our findings provide a first causal estimate of the importance of labor market adjustments in explaining the effect of import exposure on voting. The novel methodology we develop for this purpose may be useful in a broad range of empirical applications studying causal mechanisms in IV settings.

References

- Acemoglu, D. and P. Restrepo (2017). Robots and jobs: Evidence from us labor markets. *MIT Unpublished Mimeo.*
- Altonji, J. G. and R. L. Matzkin (2005, July). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73(4), 1053–1102.
- Art, D. (2007). Reacting to the radical right lessons from germany and austria. *Party Politics* 13(3), 331–349.
- Arzheimer, K. (2009). Contextual Factors and the Extreme Right Vote in Western Europe, 1980–2002. *American Journal of Political Science* 53(2), 259–275.
- Autor, D. and D. Dorn (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review* 103(5), 1553–1597.
- Autor, D., D. Dorn, and G. Hanson (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review* 103(6), 2121–68.
- Autor, D., D. Dorn, G. Hanson, and K. Majlesi (2016). Importing political polarization? the electoral consequences of rising trade exposure. *NBER Working Paper*.
- Autor, D., D. Dorn, and G. H. Hanson (2015). Untangling trade and technology: Evidence from local labour markets. *The Economic Journal* 125(584), 621–646.
- Bade, F.-J., C.-F. Laaser, and R. Soltwedel (2003). Urban specialization in the internet age empirical findings for germany, processed. *Kiel Institute for World Economics*.
- Bagues, M. and B. Esteve-Volart (2014). Politicians’ Luck of the Draw: Evidence from the Spanish Christmas Lottery. *Accepted at Journal of Political Economy*.
- Baron, R. M. and D. A. Kenny (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51(6), 1173.

- Becker, S. O. and M.-A. Muendler (2015). Trade and tasks: an exploration over three decades in germany. *Economic Policy* 30(84), 589–641.
- Bender, S., A. Haas, and C. Klose (2000). Iab employment subsample 1975-1995 opportunities for analysis provided by the anonymised subsample. *IZA Discussion Paper* 117.
- Bloom, N., M. Draca, and J. Van Reenen (2016). Trade induced technical change? the impact of chinese imports on innovation, it and productivity. *The Review of Economic Studies* 83(1), 87–117.
- Blundell, R. and J. Powell (2003). Endogeneity in nonparametric and semiparametric regression models. In L. P. H. M. Dewatripont and S. J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Volume 2. Cambridge, UK: Cambridge University Press.
- Blundell, R. and J. Powell (2004, July). Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71(3), 655–679.
- Brunner, E., S. L. Ross, and E. Washington (2011). Economics and policy preferences: causal evidence of the impact of economic conditions on support for redistribution and other ballot proposals. *Review of Economics and Statistics* 93(3), 888–906.
- Burgess, S., R. M. Daniel, A. S. Butterworth, and S. G. Thompson (2015). Network mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *International Journal of Epidemiology* 44(2), 484–495.
- Charles, K. K. and M. J. Stephens (2013). Employment, wages, and voter turnout. *American Economic Journal: Applied Economics* 5(4), 111–143.
- Che, Y., Y. Lu, J. R. Pierce, P. K. Schott, and Z. Tao (2016). Does trade liberalization with china influence us elections? Technical report, National Bureau of Economic Research.
- Christensen, G. S. and E. Miguel (2016). Transparency, reproducibility, and the credibility of economics research. Technical report, National Bureau of Economic Research.
- Dauth, W., S. Findeisen, and J. Suedekum (2014). The Rise of the East and the Far East: German Labor Markets and Trade Integration. *Journal of European Economic Association* 12(6), 1643–1675.
- Dippel, C., R. Gold, and S. Heblich (2015). Globalization and its (dis-) content: Trade shocks and voting behavior. *NBER Working Paper* (w21812).
- Dunn, G. and R. Bentall (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine* 26(26), 4719–4745.
- Duranton, G. and D. Puga (2005). From sectoral to functional urban specialisation. *Journal of Urban Economics* 57(2), 343–370.
- Dustmann, C., B. Fitzenberger, U. Schönberg, and A. Spitz-Oener (2014). From sick man of europe to economic superstar: Germany’s resurgent economy. *The Journal of Economic Perspectives* 28(1), 167–188.
- Falck, O., R. Gold, and S. Heblich (2014). E-lections: Voting Behavior and the Internet. *American Economic Review* 104(7), 2238–65.

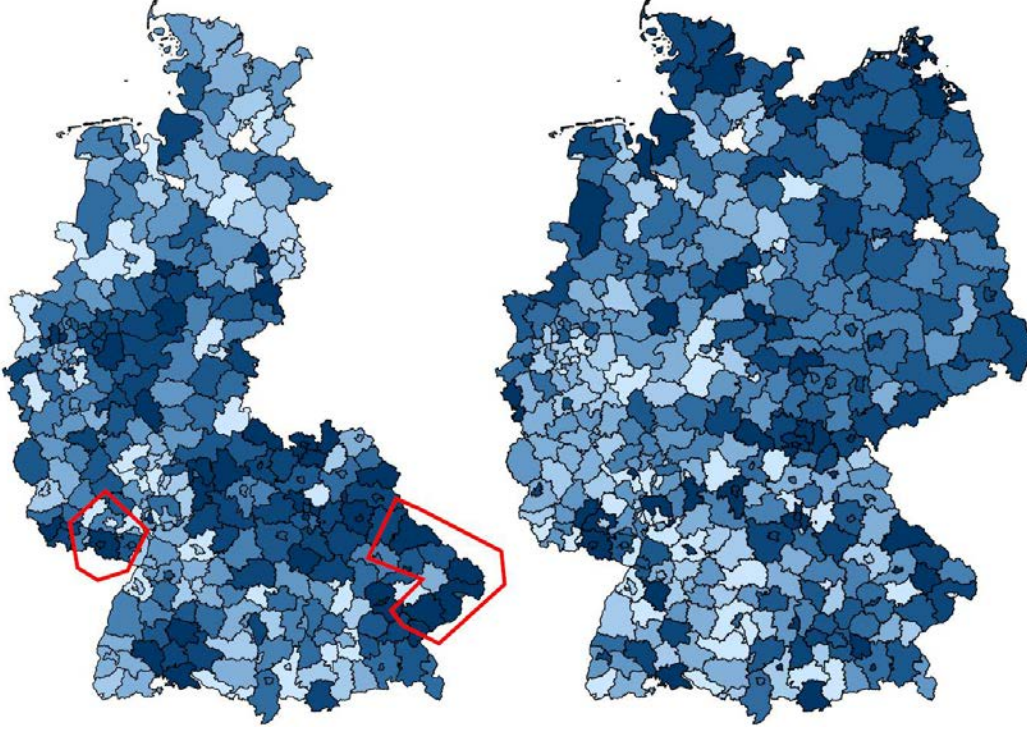
- Falck, O., S. Heblich, and A. Otto (2013). Agglomerationsvorteile in der wissenschaftsgesellschaft: Empirische evidenz für deutsche gemeinden. *ifo Schnelldienst* 66(3), 17–21.
- Falk, A., A. Kuhn, and J. Zweimüller (2011). Unemployment and Right-wing Extremist Crime. *The Scandinavian Journal of Economics* 113(2), 260–285.
- Feigenbaum, J. J. and A. B. Hall (2015). How legislators respond to localized economic shocks: Evidence from chinese import competition. *Journal of Politics* 77(4), 1012–30.
- Frank, T. (March 7th 2016). Millions of ordinary americans support donald trump. here’s why. *The Guardian*.
- Frey, C. B. and M. A. Osborne (2017). The future of employment: How susceptible are jobs to computerisation? *Oxford Martin School Unpublished Mimeo*..
- Frolich, M. and M. Huber (2017). Direct and indirect treatment effectscausal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, n/a–n/a.
- Geneletti, S. (2007). Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society B* 69(2), 199–215.
- Gennetian, L. A., J. Bos, and P. Morris (2002). Using instrumental variables analysis to learn more from social policy experiments. Mdrcc working papers on research methodology, MDRC (Manpower Demonstration Research Corporation).
- Giuliano, P. and A. Spilimbergo (2014). Growing up in a recession. *The Review of Economic Studies* 81(2), 787–817.
- Goos, M., A. Manning, and A. Salomons (2009). Job polarization in europe. *The American Economic Review* 99(2), 58–63.
- Goos, M., A. Manning, and A. Salomons (2014). Explaining job polarization: Routine-biased technological change and offshoring. *The American Economic Review* 104(8), 2509–2526.
- Grumke, T. (2012). *The Extreme Right in Europe*. Vandenhoeck & Ruprecht.
- GSOEP (2007). The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements. Technical Report 1.
- Hafeneger, B. and S. Schönfelder (2007). *Politische Strategien gegen die extreme Rechte in Parlamenten. Folgen für kommunale Politik und lokale Demokratie*. Friedrich-Ebert-Stiftung: Berlin.
- Hagan, J., H. Merckens, and K. Boehnke (1995). Delinquency and Disdain: Social Capital and the Control of Right-Wing Extremism Among East and West Berlin Youth. *American Journal of Sociology* 100(4), 1028–1052.
- Hansen, L. P. (1982, July). Large sample properties of generalized method of moments estimators. *Econometrica* 50(4), 1029–1054.
- Heckman, J. and R. Pinto (2017). Unordered monotonicity. *Forthcoming Econometrica*.
- Heckman, J. J. (2008). The principles underlying evaluation estimators with an application to matching. *Annales d’Economie et de Statistiques* 91–92, 9–73.

- Heckman, J. J. and R. Pinto (2015a). Causal analysis after Haavelmo. *Econometric Theory* 31(1), 115–151.
- Heckman, J. J. and R. Pinto (2015b). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews* 34(1-2), 6–31.
- Heckman, J. J., R. Pinto, and P. A. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 2052–2086.
- Heckman, J. J. and E. J. Vytlačil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73(3), 669–738.
- Hiscox, M. J. (2002). Commerce, coalitions, and factor mobility: Evidence from congressional votes on trade legislation. *American Political Science Review* 96(3), 593–608.
- Holmes, T. J. and J. J. Stevens (2014). An alternative theory of the plant size distribution, with geography and intra-and international trade. *Journal of Political Economy* 122(2), 369–421.
- Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological Methods* 15(4), 309–334.
- Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011a). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105(4), 765–789.
- Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011b). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105, 765–789.
- Imai, K., L. Keele, and T. Yamamoto (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* 25(1), 51–71.
- Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Imbens, G. W. and W. K. Newey (2007). Identification and estimation of triangular simultaneous equations models without additivity. Unpublished manuscript, Harvard University and MIT.
- Jensen, J. B., D. P. Quinn, and S. Weymouth (2016). Winners and losers in international trade: The effects on us presidential voting. Technical report, National Bureau of Economic Research.
- Jhun, M. A. (2015). *Epidemiologic approaches to understanding mechanisms of cardiovascular diseases: genes, environment, and DNA methylation*. Ph. D. thesis, University of Michigan, Ann Arbor.
- Jun, S. J., J. Pinkse, H. Xu, and N. Yildiz (2016). Multiple discrete endogenous variables in weakly-separable triangular models. *Econometrics* 4(1).
- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica* 75(1), 83–119.
- Krueger, A. B. and J.-S. Pischke (1997). A Statistical Analysis of Crime Against Foreigners in Unified Germany. *Journal of Human Resources* 32(1), 182–209.

- Krugman, P. R. (2008). Trade and Wages, Reconsidered. *Brookings Papers on Economic Activity* 2008(1), 103–154.
- Lee, S. and B. Salanié (2015). Identifying effects of multivalued treatments.
- Lubbers, M. and P. Scheepers (2001). *European Sociological Review* 17(4), 431–449.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Routledge.
- Malgouyres, C. (2014). The impact of exposure to low-wage country competition on votes for the far-right: Evidence from french presidential elections. *working paper*.
- Malgouyres, C. (2017). The impact of chinese import competition on the local structure of employment and wages: Evidence from france. *Journal of Regional Science* 57(3), 411–441.
- Mocan, N. H. and C. Raschke (2014). Economic Well-being and Anti-Semitic, Xenophobic, and Racist Attitudes in Germany. *National Bureau of Economic Research Working Paper* 20059.
- Mudde, C. (2000). *The Ideology of the Extreme Right*. Manchester University Press.
- Mughan, A., C. Bean, and I. McAllister (2003). Economic globalization, job insecurity and the populist reaction. *Electoral Studies* 22(4), 617–633.
- Mughan, A. and D. Lacy (2002). Economic Performance, Job Insecurity and Electoral Choice. *British Journal of Political Science* 32(3), 513–533.
- New York Times (2009). Ancient citys nazi past seeps out after stabbing. *February 11th*.
- Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods, Special Section: Naturally Occurring Section on Causation Topics in Psychological Methods* 19, 459–481.
- Petersen, M. L., S. E. Sinisi, and M. J. Van der Laan (2006). Estimation of direct causal effects. *Epidemiology* 17, 276–284.
- Pierce, J. R. and P. K. Schott (2016). The surprisingly swift decline of us manufacturing employment. *American Economic Review* 106(7), 1632–62.
- Pinto, R. (2015). Selection bias in a controlled experiment: The case of Moving to Opportunity. Unpublished Ph.D. Thesis, University of Chicago, Department of Economics.
- Powdthavee, N. (2009). Does education reduce blood pressure? estimating the biomarker effect of compulsory schooling in England. Discussion Paper 09/14, University of York, Department of Economics, York, UK.
- Robins, J. M. (2003). Semantics of causal dag models and the identification of direct and indirect effects. In N. L. P. J. Green, Hjort and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, MR2082403, pp. 70–81. Oxford: Oxford University Press.
- Robins, J. M. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3(2), 143–155.
- Rodrik, D. (1995). Political economy of trade policy. *Handbook of international economics* 3(3), 1457–1494.

- Rogowski, R. (1987). Political cleavages and changing exposure to trade. *American Political Science Review* 81(4), 1121–1137.
- Rosenbaum, P. R. and D. B. Rubin (1983, April). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55.
- Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes (with discussion). *Scandinavian Journal of Statistics* 31, 161–170.
- Scheve, K. F. and M. J. Slaughter (2001). What Determines Individual Trade-Policy Preferences? *Journal of International Economics* 54(2), 267–292.
- Schmidt, S. and R. Buehler (2007). The planning process in the us and germany: a comparative analysis. *International Planning Studies* 12(1), 55–75.
- Small, D. S. (2012). Mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables. *Journal of Statistical Research* 46(2), 91–103.
- Sommer, B. (2008). Anti-capitalism in the name of ethno-nationalism: ideological shifts on the german extreme right. *Patterns of Prejudice* 42(3), 305–316.
- Stöss, R. (2010). Rechtsextremismus im Wandel. Technical report, Friedrich Ebert Stiftung.
- Ten Have, T. R., M. M. Joffe, K. G. Lynch, G. K. Brown, S. A. Maisto, and A. T. Beck (2007, September). Causal mediation analyses with rank preserving models. *Biometrics* 63(3), 926–934.
- The Economist (July 30th 2016). The new political divide.
- Voigtländer, N. and H.-J. Voth (2015). Taught to Hate: Nazi Indoctrination and Anti-Semitic Beliefs in Germany. *Proceedings of the National Academy of Sciences* Forthcoming.
- Vytlacil, E. J. (2006, August). Ordered discrete-choice selection models and local average treatment effect assumptions: Equivalence, nonequivalence, and representation results. *Review of Economics and Statistics* 88(3), 578–581.
- Yamamoto, T. (2014, March). Identification and estimation of causal mediation effects with treatment noncompliance. Manuscript. Department of Political Science, Massachusetts Institute of Technology, Cambridge.

Figure A1: T_{it} in 1987–1998 (Left), and 1998–2009 (Right)



Notes: Trade Shocks mapped into 322 West German counties for 1987–1998 (left) and into 408 German counties for 1998–2009 (right). The two circles enclose the regions in Palatine (on the left) and Bavaria (on the right).

Appendix A Graphical Representation and Qualitative Evidence

Figure A1 shows the spatial dispersion of our key regressor T_{it} . While our empirical analysis will use fixed effects to wash out any secular trends, it is reassuring that even in the raw data there appears to be little correlation in T_{it} in space or time.⁶⁴ This reflects both Germany's diverse pattern of industrial production and the fact that the dominant driver of T_{it} changed from Eastern Europe in 1987–1998 to China in 1998–2009 (Dauth et al., 2014).

The enclosed region in the south-west of our map is Southwest-Palatine (*Südwestpfalz*), a region that traditionally produced shoe and leather manufacturing firms. In our data, Southwest-Palatine is in the top decile of negatively shocked districts in both periods. In 2005, the region's biggest city, Pirmasens, had an unemployment rate of 20 percent. At the same time, extreme-right parties increased their vote-share from 1.3 percent in 1987 to 3.45 percent in 2009. A study commissioned by the Friedrich Ebert Foundation conducted interviews with local politicians which suggested that

⁶⁴We use period-specific fixed effects for four broad regions. With East Germany as one of the regions we thus have a total of seven period-by-region fixed effects in our stacked panel.

the local *Republikaner* managed to mobilize enough voter support to enter Pirmasens' city parliament by explicitly linking import competition to social hardships (Hafeneger and Schönfelder, 2007).

The enclosed counties in the south eastern part of the map are Rottal-Inn, Passau, Freyung-Grafenau, Regen, and Cham. This is a traditional manufacturing region specialized in glass products and furniture making. This region too saw high import exposure, declining employment, and increasing support for the extreme right in the last decades, attracting international attention when neo-Nazis carried out a near-fatal attack on Passau's police chief in 2008 (*New York Times*, 2009). Additional descriptive statistics are reported in [Online Appendix C](#).

Appendix B Identification of T on Y with Gravity Residuals

An alternative approach to the IV approach pursued in the paper is to estimate gravity equations, which exploit essentially the same source of exogenous variation. The endogeneity concern with increasing imports ΔIM_{Gjt} is that they reflect not only increasing competitiveness of Chinese and Eastern European ('CE') industries⁶⁵, but also German industry-specific demand changes.

The gravity approach to solving this problem is to compare changes in German industries' exports to other countries O in relation to Chinese and Eastern European exports to O . This comparison reflects changes in Chinese and Eastern European comparative advantage over Germany, and allows constructing an exogenous measure ΔIM_{Gjt}^{grav} to replace ΔIM_{Gjt} .⁶⁶

[Online Appendix E](#) shows how to obtain the gravity-residuals ΔIM_{Gjt}^{grav} that replace ΔIM_{Gjt} the gravity-residuals ΔEX_{Gjt}^{grav} that replace ΔEX_{Gjt} . An exogenous measure for changes in German industries' trade exposure can be obtained from netting out both effects such that $\Delta Trade_{Gjt}^{grav} = \Delta IM_{Gjt}^{grav} - \Delta EX_{Gjt}^{grav}$. Substituting $\Delta Trade_{Gjt}$ in equation (1) with $\Delta Trade_{Gjt}^{grav}$ provides an exogenous measure of regional trade exposure based on the gravity approach as

$$T_{it}^{grav} = \sum_j \frac{L_{ijt}}{L_{jt}} \frac{\Delta Trade_{Gjt}^{grav}}{L_{it}} \quad (49)$$

We now substitute T_{it} from our baseline regression (3) with T_{it}^{grav} directly. Otherwise, we run

⁶⁵Competitiveness increases due to productivity increases, better market access, and decreasing relative trade cost.

⁶⁶As before, we chose Belgium, France, Greece, Italy, Luxembourg, the Netherlands, Portugal, Spain, and the UK as "other countries" O for our gravity regressions, to be comparable with [Dauth et al. \(2014\)](#).

exactly the same specifications when estimating

$$Y_{it} = \Gamma_T^Y \cdot T_{it}^{grav} + \Gamma_X^Y \cdot X_{it} + \epsilon_{it}^Y \quad (50)$$

Results are reported in Table 10. Again, each cell reports results from a different regression. Rows specify different outcome variables, and columns refer to different regression specifications. Results are consistent with our main specifications reported in table 2. The key observation is that the positive effect of trade exposure on extreme-right party votes is confirmed by this alternative identification strategy.

In addition, a few other effects that were insignificant in table 2 become sharper here: there is more evidence for a positive effect of trade exposure on turnout. Moreover, the positive effect on the vote share of the market-liberal party FDP turns significant in our preferred specification in column 5. Additionally, the negative effect on far-left parties is significant in the gravity regressions. Overall, these patterns are all internally consistent, and do not detract from our focus on extreme-right vote shares in the IV setting.

Appendix C Individual-Level Evidence

In this section, we test whether our regional-level results can be confirmed at the individual level. For each party, we aggregate individuals' self-reported voting intentions into a decadal cumulative share of years in which a respondent answered in the affirmative. Based on this, we calculate Y_{wt}^P as the ratio of the number of years that w states a preference for party P , divided by the number of years that w answered the question in the SOEP. It is better to measure the outcome as a cumulative share for the whole period instead of using a first difference approach because the latter relies only on individuals' answer at the beginning and the end of the period. Moreover, respondents do not answer all questions in every year, which increases the number of missing observations in a first difference specification.⁶⁷ For each party P , the dependent variable is a share

⁶⁷ As a result, we obtain about three times as many 'person-decade' observations, and correspondingly more precise estimates, using the share measure than with the first-difference measure.

between 0 and 1 for individual w in time period t and we separately estimate

$$Y_{wt}^P = \gamma_{Y-1}^Y \cdot Y_{wt-1}^P + \gamma_T^Y \cdot T_{it} + \gamma_X^Y \cdot X_{it-1} + \epsilon_{wt}. \quad (51)$$

for each party outcome. With a slight abuse of notation, Y_{wt-1}^P controls for w 's survey response to the same question in the base year. X_{it} refers to the same set of regional controls for the base-year as in table 2. Our focus is on estimating γ_T^Y , the effect of region i 's trade exposure T_{it} on a resident worker w 's reported party support.

Table 11 reports the results. Across rows it mimics closely our main table 2, except that there is no turnout measure in the SOEP. Every coefficient in table 11 reports the estimate of γ_T^Y from a separate regression. T_{it} is always instrumented as before, we do not report the first stage regressions again. Column 1 includes the two period fixed effects and four region fixed effects as well as the regional economic controls from table 2. We also add region i 's base-year socio-economic and voting controls X_{it-1} from table 2 for each period. To better gauge magnitudes, column 2 reports the same specification with standardized outcomes.

A county's import exposure shifts individuals' preferences to the extreme right, though the effect is weaker in the SOEP's stated preferences data than it was in the actual voting data in table 2, with a t -statistic of only 1.62. By contrast there is stronger evidence of a reduction in preference for the established left party, the SPD. No other party across the entire spectrum shows a response that is close to being significant.

Once we dig deeper into what types of workers are driving the observed patterns we find distinctive results. In columns 5–7 we split the sample by skill as well as by whether an individual works in manufacturing, i.e. whether their employment sector is more heavily exposed to trade competition.⁶⁸ Both the extreme right effect and the SPD effect are entirely driven by low-skill workers, while high-skill workers do not respond at all.⁶⁹ Splitting the low-skill sample into manufacturing and non-manufacturing employment, we see that the extreme-right response is entirely driven by low-skill workers in manufacturing sectors. This implies that those who are

⁶⁸In an earlier working paper, we focused on comparing the effect of individuals' trade exposure due to their industry of employment relative to their regions' trade exposure (Dippel, Gold, and Heblich, 2015). However, we have come to the conclusion that individuals' industry of employment is measured too coarsely in the SOEP to draw strong conclusions about the relative importance of these two types of trade exposure.

⁶⁹The SOEP reports skills as educational attainment according to the 'ISCED-1997' classification, where 'high' means some college.

most likely to experience adverse labor market effects from trade are the ones most likely to turn towards the extreme right because of increasing trade exposure in their region. By contrast, the reduction in the change in the SPD's vote share is entirely driven by low-skill *non*-manufacturing workers. A possible, though speculative, explanation is that low-skill workers in the service sector are affected by competition from laid-off manufacturing workers, or that laid-off manufacturing workers had to accept unattractive jobs in the service sector. In either case, they might blame the SPD-induced labor market reforms, such that trade exposure would only indirectly affect their changing party support.

Appendix D Proof of Theorem T-1

Proof P-1 *Our model stem from seven exogenous and statistically independent random variables: the unobserved five error terms $\epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y$, the observed instrument Z and the unobserved variables V_T , and V_Y . All remaining variables of the model can be expressed in term of these seven variables.*

We first show that non-independence relations $Z \not\perp\!\!\!\perp T$ and $Z \not\perp\!\!\!\perp M|T$ hold. To do so, it is useful to express the treatment T and the mediator variable M in terms of the model exogenous variables:

$$T = f_T(Z, V_T, \epsilon_T), \quad (52)$$

$$\text{and } M = f_M(T, U, V_T, V_Y, \epsilon_M),$$

$$\text{but } U = f_U(T, \epsilon_U),$$

$$\text{thus } = f_M(T, f_U(T, \epsilon_U), V_T, V_Y, \epsilon_M). \quad (53)$$

Equation (52) implies $Z \not\perp\!\!\!\perp T$. It remains to prove that $Z \not\perp\!\!\!\perp M|T$. To do so, it is useful to show that $Z \not\perp\!\!\!\perp V_T|T$. According to equation (52), conditioning on $T = t$ is equivalent to conditioning on the values of V_T, Z, ϵ_T such that $f_T(Z, V_T, \epsilon_T) = t$. This induces a correlation between Z and V_T and thereby $Z \not\perp\!\!\!\perp V_T|T$. While $Z \perp\!\!\!\perp (V_T, \epsilon_T)$ holds because (Z, V_T, ϵ_T) are statistically independent, $Z \perp\!\!\!\perp (V_T, \epsilon_T)|(T = t)$ does not.

The arguments of mediator M in (53) can be split into into two sets of variables: (T, V_T) and $(V_Y, \epsilon_U, \epsilon_M)$. Note that independence relation among exogenous variables $(Z, V_T, \epsilon_T) \perp\!\!\!\perp (V_Y, \epsilon_U, \epsilon_M)$ holds. But, according to (52), T is a function of (Z, V_T, ϵ_T) . Thereby $T \perp\!\!\!\perp (V_Y, \epsilon_U, \epsilon_M)$ also holds. Nevertheless, the remaining arguments of M are T, V_T and $Z \not\perp\!\!\!\perp V_T|T$ implies that $Z \not\perp\!\!\!\perp g(T, V_T, V_Y, \epsilon_U, \epsilon_M)|T$ for any

non-degenerate function $g(\cdot)$ of V_T and, in particular, $Z \not\perp\!\!\!\perp M|T$.

It remains to prove the three exclusion restrictions: (1) $Z \perp\!\!\!\perp Y(t)$; (2) $Z \perp\!\!\!\perp M(t)$; and (3) $Z \perp\!\!\!\perp Y(m)|T$. To so do, it is useful to express the counterfactuals $Y(t)$, $M(t)$ and $Y(m)$ in terms of the seven exogenous variables of the model:

$$U(t) = f_U(t, \epsilon_U), \quad (54)$$

$$\begin{aligned} M(t) &= f_M(t, U(t), V_T, V_Y, \epsilon_M) \\ &= f_M(t, f_U(t, \epsilon_U), V_T, V_Y, \epsilon_M) \end{aligned} \quad (55)$$

$$\begin{aligned} Y(t) &= f_Y(t, M(t), U(t), V_Y, \epsilon_Y) \\ &= f_Y(t, f_M(t, f_U(t, \epsilon_U), V_T, V_Y, \epsilon_M), f_U(t, \epsilon_U), V_Y, \epsilon_Y) \end{aligned} \quad (56)$$

$$\begin{aligned} Y(m) &= f_Y(T, m, U, V_Y, \epsilon_Y) \\ &= f_Y(T, m, f_U(T, \epsilon_U), V_Y, \epsilon_Y) \end{aligned} \quad (57)$$

According to equation (56), $Y(t)$ is a function of exogenous variables $\epsilon_U, V_T, V_Y, \epsilon_M, \epsilon_Y$, thus:

$$(\epsilon_U, V_T, V_Y, \epsilon_M, \epsilon_Y) \perp\!\!\!\perp Z \Rightarrow Y(t) \perp\!\!\!\perp Z.$$

According to (55), $M(t)$ is a function of exogenous variables $\epsilon_U, V_T, V_Y, \epsilon_M$, thus:

$$(\epsilon_U, V_T, V_Y, \epsilon_M) \perp\!\!\!\perp Z \Rightarrow M(t) \perp\!\!\!\perp Z.$$

Exogenous variables $(Z, V_T, \epsilon_T, \epsilon_U, V_Y, \epsilon_Y)$ are mutually statistically independent, and thereby $(Z, V_T, \epsilon_T) \perp\!\!\!\perp (\epsilon_U, V_Y, \epsilon_Y)$ holds. But according to (52), T is a function of exogenous variables (Z, V_T, ϵ_T) , thus $(T, Z) \perp\!\!\!\perp (\epsilon_U, V_Y, \epsilon_Y)$, which also implies that $Z \perp\!\!\!\perp (\epsilon_U, V_Y, \epsilon_Y)|T$. Moreover, according to (57), $Y(m)$ is a function of exogenous variables $(\epsilon_U, V_Y, \epsilon_Y)$, thus:

$$Z \perp\!\!\!\perp (\epsilon_U, V_Y, \epsilon_Y)|(T = t) \Rightarrow Z \perp\!\!\!\perp f_Y(t, m, f_U(t, \epsilon_U), V_Y, \epsilon_Y)|(T = t) \Rightarrow Z \perp\!\!\!\perp Y(m)|(T = t).$$

Appendix E Proof of Corollary C-1

Proof P-2 We need to prove that $P(Y(m) \leq y|T = t) = P(Y(m, t) \leq y)$ for $y \in \text{supp}(Y)$. It is useful to express counterfactual $Y(m, t)$ as a function of exogenous variables:

$$\begin{aligned} Y(m, t) &= f_Y(t, m, U(t), V_Y, \epsilon_Y) \\ &= f_Y(t, m, f_U(t, \epsilon_U), V_Y, \epsilon_Y) \end{aligned} \quad (58)$$

Moreover, exogenous variables $\epsilon_U, V_Y, \epsilon_Y, Z, V_T, \epsilon_T$ are mutually statistically independent, and, in particular:

$$(\epsilon_U, V_Y, \epsilon_Y) \perp\!\!\!\perp (Z, V_T, \epsilon_T). \quad (59)$$

According to Equations (57), we have that:

$$\begin{aligned} P(Y(m) \leq y|T = t) &\equiv P(f_Y(t, m, U, V_Y, \epsilon_Y) \leq y|T = t), \\ &\equiv P(f_Y(t, m, f_U(t, \epsilon_U), V_Y, \epsilon_Y) \leq y|f_T(Z, V_T, \epsilon_T) = t), \\ &= P(f_Y(t, m, f_U(t, \epsilon_U), V_Y, \epsilon_Y) \leq y), \\ &\equiv P(Y(t, m) \leq y), \end{aligned}$$

where the third equality comes from the independence relation (59).

Table 10: Gravity Results for Effect of T_{it} on Voting

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline Gravity	+ Structure Gravity	+ Industry Gravity	+ Voting Gravity	+Socio Gravity	Standard. Gravity
Δ Turnout	0.000** (2.143)	0.000* (1.774)	0.000** (1.980)	0.000** (1.966)	0.000* (1.706)	0.002* (1.706)
<i><u>Established Parties:</u></i>						
Δ Vote Share CDU/CSU	0.006 (0.873)	0.003 (0.392)	0.003 (0.342)	0.001 (0.186)	0.001 (0.089)	0.000 (0.089)
Δ Vote Share SPD	-0.008 (-1.197)	-0.004 (-0.674)	-0.005 (-0.720)	0.004 (0.732)	0.004 (0.668)	0.000 (0.668)
Δ Vote Share FDP	0.001 (0.204)	0.006 (1.365)	0.004 (1.064)	0.007* (1.840)	0.007* (1.933)	0.001* (1.933)
Δ Vote Share Green Party	0.006** (2.021)	0.000 (0.071)	0.001 (0.408)	0.000 (0.047)	0.000 (0.012)	0.000 (0.012)
<i><u>Non-established Parties</u></i>						
Δ Vote Share Extreme-Right Parties	0.004* (1.855)	0.006** (2.430)	0.006** (2.276)	0.003 (1.575)	0.003* (1.779)	0.002* (1.779)
Δ Vote Share Far-Left Parties	-0.011* (-1.814)	-0.012* (-1.884)	-0.012* (-1.894)	-0.014** (-2.382)	-0.013** (-2.177)	-0.003** (-2.177)
Δ Vote Share Other Small Parties	0.002 (0.755)	0.002 (0.696)	0.003 (1.141)	-0.001 (-0.710)	-0.001 (-0.766)	-0.001 (-0.766)
Period-by-region F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730	730

Notes: (a) Each cell reports results from a separate regression. The data is a stacked panel of first-differences at the *Landkreis* level. Each regression has 730 observations, i.e. 322 *Landkreise* in West Germany, observed in 1987–1998 and 1998–2009, and 86 *Landkreise* in East Germany, observed only in 1998–2009. We drop three city-states (Hamburg, Bremen, and Berlin in the East). (b) All specifications include region-by-period fixed effects. Column 1 controls only for start-of-period manufacturing. Column 2 adds controls for the structure of the workforce (share female, foreign, and high-skilled). Column 3 adds controls for dominant industries (employment share of the largest industry, in automobiles, and chemicals). Column 4 adds start-of-period voting controls. Column 5 adds socioeconomic controls at the start of the period (population share of unemployed individuals, and individuals aged 65+). This is our preferred specification. Finally, Column 6 presents our preferred specification with standardized outcome variables to facilitate comparison. (c) All standard errors are clustered at the level of 96 commuting zones. All specifications include region-by-period fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: Individual-Level Analysis

	(1)	(2)	(3)	(4)	(5)
	All Controls	Standardized	High-Skill	Low-Skill & Manuf.	Low-Skill & Not Manuf.
<i><u>Established Parties:</u></i>					
Would Vote CDU/CSU	0.001 (0.292)	0.003 (0.292)	-0.007 (0.278)	-0.013 (0.794)	0.008 (0.827)
Would Vote SPD	-0.008* (1.901)	-0.016* (1.901)	-0.013 (0.460)	-0.011 (0.400)	-0.017* (1.930)
Would Vote FDP	0.001 (0.459)	0.005 (0.459)	-0.018 (0.420)	0.011 (0.664)	0.007 (0.568)
Would Vote Green Party	0.003 (1.000)	0.012 (1.000)	0.070 (1.474)	0.025 (0.909)	0.002 (0.152)
<i><u>Non-Established Parties:</u></i>					
Would Vote Extreme-Right Parties	0.003 (1.619)	0.023 (1.619)	0.010 (0.875)	0.083** (2.206)	0.006 (0.475)
Would Vote Far-Left Parties	-0.001 (1.059)	-0.007 (1.059)	-0.051 (1.358)	0.019 (1.356)	-0.009 (1.055)
Would Vote Other Small Parties	-0.001 (0.642)	-0.007 (0.642)	0.005 (0.182)	-0.026 (1.053)	-0.003 (0.190)
Period-by-region F.E.	Yes	Yes	Yes	Yes	Yes
Observations	9,669	9,669	1,348	2,199	6,122

Notes: (a) Each cell in this table reports on a separate regression. An observation is an individual w over a period t , where we consider 1990–1998, and 1998–2009, closely mirroring the local labor market results. Each row reports on stated preferences for a different party. The outcome in each row is the ratio of the number of years that w states a preference for party P , divided by the number of years that w answered the question in the SOEP. The reported coefficient in all cells is the IV coefficient of regional trade exposure T_{it} . (b) Column 1 is the baseline specification which includes period and four region fixed effects as well as all the regional economic, voting and demographic controls from table 2, and individuals' base-year stated political preferences. This is the full set of controls included in all columns. To better gauge magnitudes, column 2–5 standardize all outcomes by their mean. In columns 3–5, we break the sample by individuals' skill as well as by whether they are employed in the manufacturing sector ($1,348 + 2,199 + 6,122 = 9,669$). High-skill workers (column 3) do not change their political support at all in response to trade exposure. Column 4 shows that it is the population most affected by trade exposure – low-skill manufacturing workers – that drives the effects on the far right. (c) t -statistics are reported in round brackets, standard errors are clustered at the region level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Online Appendix

to

**“Instrumental Variables and Causal Mechanisms:
Unpacking the Effect of Trade on Workers and Voters”**

Online Appendix A Data Sources

Online Appendix A.1 Election Data

We focus on federal elections (*Bundestagswahlen*) because the timing of state elections (*Landtagswahlen*) and local elections (*Kommunalwahlen*) varies widely across German regions. Federal elections took place in 1987, in December 1990 after the reunification on October 3, and in 1994, 1998, 2002, 2005, and 2009. We define the first-period outcomes as changes in the vote-share from 1987 to 1998, and second-period outcomes as changes from 1998 to 2009. Election outcomes are observed at the level of 412 districts (*Landkreis*) in Period 2 and 322 West German districts in Period 1.

The average vote share of extreme-right parties is persistently below 5 percent in both periods. This presented a major challenge for our data collection, since official election statistics do not report all votes shares below the 5 percent minimum threshold separately by party. To extract information on extreme-right parties from this residual category, we had to contact the statistical offices of the German states and digitize some results from hard copies. By doing so, we have generated a unique data set that provides detailed insight into Germany's political constellation and allows us to create a precise measure of spatial variation in preferences also for fringe parties. This measure eventually allows us to extend existing studies on spatial variation of extreme-right activities and partisanship that were typically bound to the state level (Falk, Kuhn, and Zweimüller (2011), Lubbers and Scheepers (2001)) or limited in their time horizon (Krueger and Pischke (1997)) to a new level of detail.

Online Appendix A.2 Trade Data

Our trade data stem from the U.N. Commodity Trade Statistics Database (Comtrade). The database provides information on trade flows between country pairs, detailed by commodity type. As in Dauth et al. (2014), we express all trade flows in thousands and convert them to 2005 Euros. To merge four-digit SITC2 product codes with our three-digit industry codes, we use a crosswalk provided by Dauth et al. (2014), who themselves employ a crosswalk provided by the U.N. Statistics Division to link product categories to NACE industries. In 92 percent of the cases, commodities map unambiguously into industries. For ambiguous cases, we use national employment shares from 1978 to partition them to industries. In this way, we end up with 157 manufacturing industries (excluding fuel products), classified according to the WZ73 industry classification.

Online Appendix A.3 Labor Market Data

We obtain information on local labor markets from two different sources. Information on employment, education, and the share of foreigners stems from the Social Security records in Germany.⁷⁰ Based on the Social Security records, we calculate the trade exposure measures for local labor markets, the share of high-skilled workers (with a tertiary degree), foreign workers, workers in the automobile or chemical industry, and wages. For the years before 1999, social security data are recorded at the place of work only. After 1999, place-of-work and place-of-residence information is available.

The remaining variables are provided by the German Federal Statistical Office. These variables include the overall population, the female population share, the population share of individuals

⁷⁰See Bender et al. (2000) for a detailed description of the data from the Institute for Employment Research (IAB). For an additional description of the regional distribution of wages across German municipalities, see Falck, Heblich, and Otto (2013)

of working age (aged 18 to 65), the population share of individuals older than 65, and the unemployment rate, which is calculated by dividing the number of unemployed individuals by the working-age population.

Online Appendix B Background on German Politics 1987 to 2009

Online Appendix B.1 The German Election System

Since the end of WWII, Germany has had a multiparty party system, with the two largest parties—the *Christian Democratic Union* (CDU) and the *Social Democratic Party of Germany* (SPD)—forming coalitions with either the *Free Democratic Party* (FDP) or the Greens (*Bündnis 90/Die Grünen*) during our observation period (1987 to 2009).⁷¹ German elections are based on the principle of proportionality. The main vote, called the “second vote” (*Zweitstimme*), is being cast for parties but not for individual candidates.⁷² We will exclusively focus on this party vote. The overall number of parliamentary seats is determined in proportion to a party’s share of the second vote. Parties further have to surpass a 5 percent minimum threshold to be represented in federal parliament. However, this does not mean that small parties do not capture any votes. Small parties that failed to pass the 5 percent threshold still captured about 11 percent of the total votes in our election data.

Online Appendix B.2 The Political Party Spectrum in Germany

We always classify the CDU, the SPD, the FDP, and the Greens as established parties. The conservative CDU and the social-democratic SPD are the dominant parties in Germany, in terms of both membership and votes obtained. For our period of analysis, one of those two parties was always in power. The liberal FDP participated in governments led by the CDU. The Greens are, for ideological reasons, usually the SPD’s preferred coalition partner. On the extreme right of the political spectrum, three parties have regularly run in federal elections. The National Democratic Party of Germany (NPD - *Nationaldemokratische Partei Deutschlands*), founded in 1964, the Republicans (REP - *Die Republikaner*), founded in 1983, and the German People’s Union (DVU - *Deutsche Volksunion*), founded in 1987 (and merged with the NPD in 2011).⁷³ They all follow neo-Nazi ideologies, are anti-democratic, polemicize against globalization, and agitate against immigrants and foreigners. All three have been monitored by the German Federal Office for the Protection of the Constitution (*Verfassungsschutz*). None of these extreme-right parties has ever passed the 5 percent hurdle required to enter Germany’s national parliament, and it is unthinkable that any mainstream party would ever form a coalition with them (see Art (2007)). On the far left of the political spectrum, there are around 10 parties and factions that are often related with each other. Besides the left party (*Die Linke*) and its predecessors, the *Party of Democratic Socialism* (PDS) and *Labour and Social Justice The Electoral Alternative* (WASG), three branches have been dominant: Successors to the Communist Party of Germany, which had been outlawed in 1956, e.g., the *German Communist Party* (DKP) and the *Communist Party of Germany* (KPD); Leninist, Stalinist, and Maoist

⁷¹In this paper, we will always report votes for the CDU and its Bavarian subsection *Christian Social Union* (CSU) as combined CDU votes and refer to it as the CDU.

⁷²Voters can additionally elect individual candidates on a first-past-the-post basis. Ironically, this second ballot is called the “primary vote” (*Erststimme*). In every election district, the candidate who wins the majority of primary votes is directly elected to parliament. However, electoral law ensures that this has no significant effect on the overall distribution of seats, which is determined by the second vote.

⁷³In [Online Appendix B.4](#), we provide a history of these three parties. See also comprehensive work by Stöss (2010) or Mudde (2000).

organizations like the Marxist-Leninist Party of Germany (MLPD); and Trotskyist organizations such as the Party for Social Justice (PSG). Like the parties on the extreme right, these far-left parties are regularly monitored by either the Federal Office for the Protection of the Constitution or its state-level equivalents. We classify other parties that ran for elections but do not fit the above categories as small parties (see [Falck et al. 2014](#)).

Online Appendix B.3 Stance on Trade and Globalization

Both the large parties CDU and SPD have market-liberal as well as protectionist factions. In comparison, the CDU tends to be more market-friendly. Still, it was a government led by the SPD that implemented substantial labor market reforms in 2003-2005, amongst others decreasing employment protection, unemployment benefits, and establishing a low wage sector in Germany. The smaller FDP explicitly follows a market-liberal agenda, while the Green party focusses on environmental issues. More generally, the political left has traditionally been seen as opposing globalization and capturing the anti-globalization vote.⁷⁴ However, this is no unambiguous relationship, as the *The Economist* (2016) observes when headlining “Farewell, left versus right. The contest that matters now is open against closed.” Throughout Europe, the political left has found it difficult to take a coherent position against globalization in the last two decades, often hampered by internal intellectual conflicts ([Sommer 2008](#), [Arzheimer 2009](#)). In contrast, the right and far right successfully attended an anti-globalization agenda ([Mughan et al., 2003](#)). For the case of Germany, [Sommer \(2008, p. 312\)](#) argues that “in opposing globalization, the left-wing usually criticizes an unjust and profit-oriented economic world order. [It] does not reject globalization per se but rather espouses a different sort of globalization. In contrast, the solutions proposed by the extreme right keep strictly to a national framework. The extreme right’s claim, therefore, that it is the only political force that opposes globalization fundamentally [...] rings true.” The following excerpt from the extreme-right NPD’s ‘candidate manual’ illustrates how Germany’s far right rolls protectionist anti-globalization themes into its broader nationalistic, anti-Semitic agenda: “Globalization is a planetary spread of the capitalist economic system under the leadership of the Great Money. This has, despite by its very nature being Jewish-nomadic and homeless, its politically and military protected location mainly on the East Coast of the United States” ([Grumke, 2012, p. 328](#)).⁷⁵

Online Appendix B.4 The Extreme-Right in West Germany

There is a strong sense of historical cultural roots and their time-persistence when it comes to explaining votes for far-right parties in Germany today. [Mocan and Raschke \(2014\)](#) use state-level survey aggregates from the ALLBUS, a general population survey for Germany, to show that people who live in states that had provided above-median support of the Nazi party in the 1928 elections have stronger anti-semitic feelings today. [Voigtländer and Voth \(2015\)](#) use the same data to show that the effects of historical antisemitic attitudes on today’s political attitudes was amplified for the cohorts that grew up during Nazi Germany’s indoctrination programs in 1933–1945.

Having said that, there is substantial time-variation in the popularity of the far-right in Germany. The NPD, the oldest of the three major right-wing parties we consider, was founded in 1964

⁷⁴To some extent this may still be the case. [Che et al. \(2016\)](#) for example argue that trade liberalization with China has turned American voters towards the Democrats, though it seems as if this might have not been true for the 2016 presidential elections.

⁷⁵The bundling of protectionist anti-globalization themes with xenophobic content has also been noted in the 2016 U.S. presidential election, see for example *The Guardian* (2016).

as the successor to the German Reich Party (DRP). Its goal was to unite a number of fragmented far-right parties under one umbrella. Between 1966 and 1968, the NPD was elected into seven state parliaments, and in the 1969 federal election it missed the 5 percent minimum threshold by just 0.7 percentage points. Afterwards, support for the NPD declined and it took the NPD more than 25 years to re-enter state parliaments in Saxony (2004) and Mecklenburg-Western Pomerania (2006). In both states, the party got reelected in the subsequent elections, in 2009 and 2011, respectively. In 2001, the federal parliament brought in a claim to the German Constitutional Court to forbid the NPD due to its anti-constitutional program. The claim was turned down in 2003 because the NPD's leadership was infiltrated by domestic intelligence services agents, which caused legal problems. A second claim to forbid the party, filed on December 7th 2015, was denied by the constitutional judges on January 17th 2017.

The DVU was founded by publisher Gerhard Frey as an informal association in 1971. Frey published far-right newspapers such as the German National Newspaper (DNZ) and a number of books with the goal of mitigating Germany's role in WWII. His reputation as a publisher of far-right material helped Frey to become an influential player in the German postwar extreme right scene (Mudde (2000)). In 1986, Frey took it one step further starting his own far-right party German List (*Deutsche Liste*). After some name changes, the party became known as German People's Union (DVU) from 1987 on. Since its foundation, the DVU got parliamentary seats in the state assemblies of Brandenburg (1999, 2004), Bremen (1991, 1999, 2003, 2007), Schleswig-Holstein (1992), and Saxony-Anhalt (1998). In 2010, the DVU merged with the NPD.

The Republicans (Die Republikaner) were founded in 1983 as an ultraconservative breakaway from the Christian Democratic Union (CDU) and the Christian Social Union of Bavaria (CSU). Under their leader, Franz Schönhuber (who also ran as a candidate for the DVU and NPD in his later political career), the party moved further to the extreme right by propagating a xenophobic view on immigrants, and particularly asylum seekers. Compared to the NPD and DVU, the Republicans were considered to be less openly extreme right which helped it secure votes from the ultraconservative clientele. The REP got parliamentary seats in Berlin's senate (1989) and the state parliament of Baden-Wuerttemberg (1992, 1996).

Online Appendix B.5 The Extreme-Right in East Germany after the Reunification

In the first decade after reunification, only the two mainstream parties, CDU and SPD, were able to establish themselves regionwide in East Germany next to the Party of Democratic Socialism (PDS), the successor of the Socialist Unity Party (SED), which had been ruling the German Democratic Republic till its collapse.

During this time smaller parties were struggling to put a party infrastructure into place in East Germany. Accordingly, while all three extreme-right parties tried to establish themselves in East Germany after reunification, they did not gain major political attention until the late 1990s (Hagan, Merkens, and Boehnke, 1995). At the same time, we saw some of the worst excesses of far-right crime in East Germany in the early 1990s, when migrants' and asylum seekers' residences were set on fire and a mob of people from the neighborhood applauded. Research by Krueger and Pischke (1997) suggests that neither unemployment nor wages can explain these incidences of extreme-right-driven crime from 1991 to 1993. It is more likely that the sudden increase in the number of immigrants and asylum seekers caused these xenophobic excesses in the early 1990s.

In the mid-1990s, the initial euphoria of reunification passed and East German labor markets experienced stronger exposure to international competition. East Germany now faced almost twice as much unemployment as West Germany, and this economic malaise caused feelings of

deprivation that often transformed into violent crime against immigrants. Militant right-wing groups declared “nationally liberated zones” in East Germany where foreigners were undesired. In line with that, [Lubbers and Scheepers \(2001\)](#) find that unemployed people have been more likely to support extreme right parties in Germany, and [Falk et al. \(2011\)](#) find a significant relationship between extreme-right crimes and regional unemployment levels over the years 1996–1999.⁷⁶ The story goes that the political heritage of the GDR may have preserved ethnic chauvinism, which, in combination with subsequent economic hardship, provided a fertile ground for extreme-right parties.

Online Appendix C Descriptive Statistics

Table 1 provides descriptive statistics for our main variables. The table is organized in the following way: Each row presents the distribution of one variable, sliced into its 25th percentile, median, and 75th percentile. Columns 1–3 do this for Period 1 from 1987–1998, and columns 4–6 for Period 2 from 1998–2009. T_{it} is defined in units of 1,000 € per worker in constant 2005 prices.

A comparison of columns 1–3 and 4–6 shows that trade exposure was relatively balanced between import competition and export access in Period 1, with an average T_{it} of just 68 € per worker. In Period 2, trade exposure was more export-heavy, with changes in export access exceeding changes in import competition by on average 663 € per worker.⁷⁷

Looking at the labor market outcomes, we find evidence of economic stagnation in Period 1. Most importantly, we see a decline in the share of manufacturing employment across all regions concurrent with increasing unemployment. Indeed, Germany was considered “the sick man of Europe” during the 1990s. The period of stagnation was followed by an equally prolonged export and productivity boom. Following Gerhard Schröder’s electoral victory in 1998, Germany’s inflexible labor market institutions underwent substantial reforms. In the course of these reforms, we observe important changes in the behavior of trade unions and employers’ associations. Most importantly, firms and local labor union chapters were now allowed to deviate from collective bargaining agreements to flexibly adopt to local labor market conditions; see ([Dustmann, Fitzenberger, Schönberg, and Spitz-Oener, 2014](#)).⁷⁸ As a result of these reforms, the decline in manufacturing employment slowed down significantly during Period 2.

Finally, the table shows substantial variation in political trends across the two periods. From 1987 to 1998, established parties saw an average 4.7 percentage point reduction in their share of the popular vote, while small parties and the extreme right saw an increasing vote share. From 1998 to 2009, the main parties CDU and SPD as well as the extreme-right parties lost electoral support.⁷⁹ In summary, period 1 (1987–1998) saw changes in import competition and export access that roughly balanced out, economic stagnation and an increase in support for the extreme right. This was followed by increased export access, economic stabilization, and political moderation in period 2.

⁷⁶Note that [Falk et al.’s \(2011\)](#) findings do not necessarily contradict [Krueger and Pischke \(1997\)](#) who find no relationship between unemployment and extreme-right-driven crimes. It may very well be that the motivation for crimes changed over the 1990s.

⁷⁷[Dauth et al. \(2014\)](#) explore this finding in detail

⁷⁸A perusal of the *OECD Labour Market Policies and Institutions Indicators Database* nicely illustrates this regulatory change. On the core ‘strictness of employment protection’ index, Germany stayed in a tight band between 3.13–3.25 throughout Period 1, but this measure then dropped rapidly to an average of 1.46 during Period 2. See www.oecd.org/employment/emp/employmentdatabase-labourmarketpoliciesandinstitutions.htm

⁷⁹The large decrease in SPD vote share reflects the party breaking with its left wing, which subsequently merged with the socialist party PDS to form the new party *Die Linke*. In our data, *Die Linke* is classified as far left. See section [Online Appendix B](#) for more details.

Online Appendix Table 1: The Core Variables in 1987–1998 and in 1998–2009

percentile:		(1)	(2)	(3)	(4)	(5)	(6)
		Period 1 (1987-1998),			Period 2 (1998-2009),		
		25th	median	75th	25th	median	75th
T_{it}		-0.264	0.068	0.521	-1.222	-0.663	-0.144
\widehat{T}_{it}	(instrumented with Z_{it})	-0.068	0.143	0.402	-1.150	-0.574	-0.113
Y_{it} :	Δ Turnout	-0.034	-0.020	-0.012	-0.167	-0.128	-0.095
	Δ Vote Share CDU/CSU	-9.234	-7.659	-5.730	-4.493	-2.258	0.620
	Δ Vote Share SPD	4.120	6.472	8.248	-19.904	-17.936	-16.079
	Δ Vote Share FDP	-2.933	-2.188	-1.467	6.942	8.459	9.820
	Δ Vote Share Green	-1.779	-1.282	-0.616	2.513	3.673	4.770
	Δ Share Extreme-Right	1.520	2.086	3.099	-1.525	-1.021	-0.478
	Δ Share Far-Left	0.677	0.908	1.165	5.688	7.078	8.373
	Δ Share Small Parties	1.211	1.487	1.796	0.716	1.514	2.525
M_{it} :	$\Delta \log(\text{Total Employment})$	-0.067	0.001	0.081	-0.110	-0.044	0.021

Notes: Period one (1987–1998) is for West German labor markets only, $N = 322$. Period two (1998–2009) is for West plus East German labor markets, $N = 408$. The numbers for 1998–2009 do not change substantively if we drop the East. The table displays the 25th percentile, median, and 75th percentile of T_{it} , the voting outcomes Y_{it} , and manufacturing's share of employment M_{it} .

Online Appendix D Robustness and Further Results

Online Appendix D.1 Additional Results on the Core Table 2

Online Appendix D table 3 presents the OLS results corresponding to the paper's table 2. Online Appendix D table 2 reports the coefficients on all controls in our core table 2. The initial share of manufacturing is significantly associated with increases in the extreme-right vote-share over time. In line with that, unreported specifications show that omitting the initial manufacturing share considerably increases the estimated effect of T_{it} on extreme-right voting. While not our focus, this relationship suggests that general structural decline and economic depression provide fertile grounds for extreme-right parties (Arzheimer, 2009). Regions with more educated workers and higher female labor force participation are less prone to shift right. Older demographics appear more prone to vote right, a finding that corroborates qualitative evidence (Stöss, 2010). Finally, high initial vote shares for extreme-right parties imply a reversion in the data, perhaps indicating cyclicity, where past swing voters to the right tend to swing back toward the mainstream.

Online Appendix D.2 Labor Market Outcomes

Like Table 3 did for the voting regressions, Online Appendix D table 4 presents the OLS results corresponding to the regressions in Online Appendix D table 3.

Online Appendix E Constructing Gravity Residuals

Gravity-residuals can be obtained from the residuals of the regression

$$\log(EX_{djt}^{CE-O}) - \log(EX_{djt}^{G-O}) = \alpha_d + \alpha_j + \epsilon_{djt}^{IM}, \quad (60)$$

where $\log(EX_{djt}^{CE-O})$ are industry j 's log export values from China and Eastern Europe to destination market d , $\log(EX_{djt}^{G-O})$ are German industries' exports to the same countries, α_d are destination-market and α_j are industry-fixed effects.⁸⁰ ϵ_{djt}^{IM} thus captures CE 's competitive advantage over Germany at time t in destination market d and industry j . Averaging residuals ϵ_{djt}^{IM} over destination markets d and taking first differences provides a measure for overall changes in CE 's comparative advantage over time. Exponentiating this term and multiplying it with Germany's start-of-period imports from CE gives rise to $\Delta IM_{Gjt}^{grav} = IM_{Gjt-1} \times \exp(\bar{\epsilon}_{jt}^{IM} - \bar{\epsilon}_{jt-1}^{IM})$, which is a counterfactual measure of changes in German industries' import exposure that is solely driven by CE 's increasing comparative advantage.

Conversely, ΔEX_{Gjt} increases due to better access to the CE markets and to German-specific supply conditions. While German-specific supply conditions will affect German exports in general, the relative attractiveness of CE markets over other export destinations should be independent of German-specific effects. Thus, changes in German industries' exports to China and Eastern Europe in relation to German industries' exports to other countries O provides an exogenous measure ΔEX_{Gjt}^{grav} for ΔEX_{Gjt} . It can be obtained from the residuals of the regression

$$\log(EX_{djt}^{G-CE}) - \log(EX_{djt}^{G-O}) = \alpha_d + \alpha_j + \epsilon_{djt}^{EX}, \quad (61)$$

⁸⁰Since many CE countries did not report trade data in the late 1980s and early 1990s, we use imports from CE and Germany reported by other countries O to measure Germany's and CE 's exports to O .

Online Appendix Table 2: Coefficients on Controls in Table 2

	(1) Turnout	(2) CDU/CSU	(3) SPD	(4) FDP	(5) Green Party	(6) Right	(7) Left	(8) Small
T_{it}	0.002 (1.223)	-0.066 (-0.501)	-0.009 (-0.073)	0.119 (1.583)	-0.018 (-0.413)	0.089** (2.055)	-0.092 (-0.859)	-0.024 (-0.564)
<i>Controls Specification 1:</i>								
Empl-share manufacturing $_{-1}$	-0.000 (-1.303)	0.023 (1.092)	0.002 (0.122)	0.009 (0.793)	-0.024*** (-2.932)	0.017** (2.407)	-0.010 (-0.776)	-0.017** (-2.430)
<i>Controls Specification 2:</i>								
Pop-share college-educated $_{-1}$	0.004*** (2.920)	-0.041 (-0.811)	0.131** (2.538)	-0.055 (-1.341)	0.156*** (3.530)	-0.093*** (-5.032)	-0.146** (-2.197)	0.049 (1.544)
Pop-share foreign-born $_{-1}$	0.001 (0.358)	-0.205*** (-3.020)	-0.154* (-1.820)	0.156*** (3.820)	-0.008 (-0.185)	0.094*** (3.708)	0.095 (1.228)	0.021 (0.672)
Pop-share female $_{-1}$	0.011*** (3.104)	0.353** (2.146)	-0.012 (-0.064)	0.056 (0.534)	0.160 (1.475)	-0.262*** (-3.083)	-0.325*** (-2.602)	0.029 (0.408)
Employment-share in automotive $_{-1}$	-0.000 (-0.045)	0.019 (0.629)	-0.038** (-2.047)	-0.001 (-0.091)	0.030* (1.827)	-0.004 (-0.353)	-0.002 (-0.157)	-0.004 (-0.475)
Employment-share in chemistry $_{-1}$	-0.000 (-0.955)	0.036 (1.214)	-0.050*** (-3.196)	-0.013 (-0.889)	0.017 (0.915)	0.014 (0.821)	-0.004 (-0.199)	-0.002 (-0.189)
Employment in largest industry $_{-1}$	0.024 (0.810)	-1.807 (-1.090)	2.668** (1.982)	-1.159 (-1.212)	-1.649* (-1.704)	0.077 (0.084)	0.739 (0.569)	1.132** (2.071)
<i>Controls Specification 3:</i>								
Unemployment-share $_{-1}$	-0.003** (-2.539)	0.061 (0.819)	-0.034 (-0.431)	-0.145*** (-2.897)	-0.112*** (-2.966)	-0.051 (-1.467)	0.347*** (3.576)	-0.066*** (-2.652)
Pop-share above age 65 $_{-1}$	-0.005*** (-3.461)	-0.113* (-1.661)	-0.077 (-1.137)	-0.013 (-0.314)	-0.002 (-0.053)	0.079*** (2.598)	0.142*** (3.215)	-0.017 (-0.563)
Voter Turnout $_{-1}$	-0.000 (-0.535)	0.073*** (2.939)	0.115*** (4.710)	-0.036* (-1.740)	-0.023 (-1.562)	-0.016 (-1.638)	-0.061* (-1.934)	-0.052*** (-3.519)
CDU/CSU Voteshare $_{-1}$	-0.025*** (-4.059)	-0.255 (-0.987)	-0.111 (-0.506)	0.222 (1.217)	0.004 (0.033)	-0.635*** (-4.891)	0.612*** (3.227)	0.163 (1.177)
SPD Voteshare $_{-1}$	-0.010** (-2.327)	-0.119 (-0.502)	-0.366* (-1.946)	-0.079 (-0.521)	0.142 (1.310)	-0.084 (-0.993)	0.064 (0.284)	0.441*** (3.561)
FDP Voteshare $_{-1}$	-0.010** (-2.411)	-0.293 (-1.334)	-0.392** (-2.143)	-0.024 (-0.161)	0.022 (0.218)	-0.089 (-1.041)	0.373** (1.977)	0.403*** (3.328)
Green Party Voteshare $_{-1}$	-0.010** (-2.381)	-0.081 (-0.373)	-0.628*** (-3.599)	-0.089 (-0.588)	0.026 (0.262)	-0.076 (-0.903)	0.440** (2.426)	0.409*** (3.387)
Far-Right Voteshare $_{-1}$	-0.012*** (-2.897)	0.120 (0.528)	-0.488*** (-2.643)	-0.225 (-1.491)	0.007 (0.072)	-0.098 (-1.165)	0.359* (1.667)	0.324*** (2.702)
Far-Left Voteshare $_{-1}$	-0.014*** (-3.060)	-0.349 (-1.572)	-0.321 (-1.625)	-0.127 (-0.791)	0.059 (0.468)	-0.091 (-1.021)	0.468** (2.338)	0.359*** (2.885)
Period-by-region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730	730	730	730

Notes: T-statistics reported, standard errors are clustered at the level of 96 commuting zones, *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix Table 3: OLS Version of Table 2

	(1) Baseline OLS	(2) + Structure OLS	(3) + Industry OLS	(4) + Voting OLS	(5) +Socio OLS	(6) Standard. OLS
Δ Turnout	0.004*** (2.932)	0.003*** (2.669)	0.004*** (3.059)	0.003** (2.337)	0.003** (2.430)	0.040** (2.430)
<i><u>Established Parties:</u></i>						
Δ Vote Share CDU/CSU	-0.081 (-1.015)	-0.093 (-1.204)	-0.113 (-1.423)	-0.062 (-0.963)	-0.067 (-1.020)	-0.016 (-1.020)
Δ Vote Share SPD	-0.037 (-0.416)	-0.035 (-0.399)	-0.044 (-0.471)	0.061 (0.884)	0.062 (0.929)	0.005 (0.929)
Δ Vote Share FDP	0.094** (1.971)	0.114*** (2.672)	0.105** (2.398)	0.081* (1.805)	0.088** (2.097)	0.016** (2.097)
Δ Vote Share Green Party	0.046 (1.221)	0.034 (1.016)	0.063* (1.755)	0.062* (1.835)	0.068** (2.042)	0.024** (2.042)
<i><u>Non-established Parties</u></i>						
Δ Vote Share Extreme-Right Parties	0.038* (1.703)	0.042** (1.963)	0.036 (1.522)	-0.009 (-0.483)	-0.004 (-0.240)	-0.002 (-0.240)
Δ Vote Share Far-Left Parties	-0.108* (-1.669)	-0.105 (-1.565)	-0.109 (-1.597)	-0.138** (-2.159)	-0.153** (-2.491)	-0.039** (-2.491)
Δ Vote Share Other Small Parties	0.048 (1.586)	0.042 (1.439)	0.062** (2.186)	0.003 (0.138)	0.007 (0.259)	0.005 (0.259)
Period-by-region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730	730

Notes: T-statistics reported, standard errors are clustered at the level of 96 commuting zones, *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix Table 4: OLS Version of Table 3

	(1) Baseline OLS	(2) + Structure OLS	(3) + Industry OLS	(4) + Voting OLS	(5) +Socio OLS
Δ log(Total Employment)	-0.013*** (3.138)	-0.012*** (3.066)	-0.011** (2.514)	-0.009** (2.070)	-0.009* (1.919)
Period-by-region FE	Yes	Yes	Yes	Yes	Yes
Observations	730	730	730	730	730

Notes: T-statistics reported, standard errors are clustered at the level of 96 commuting zones, *** p<0.01, ** p<0.05, * p<0.1.

where $\log(EX_{djt}^{G-CE})$ are industry j 's log export values from Germany to China and Eastern Europe, $\log(EX_{djt}^{G-O})$ are German industries' exports to other countries, and α_d and α_j are again destination-country and industry-fixed effects. ϵ_{djt}^{EX} now captures CE 's relative attractiveness over other sales markets at time t in destination market d and industry j . Averaging residuals ϵ_{djt}^{EX} over destination markets d and taking first differences provides a measure for overall changes in the attractiveness of Chinese and Eastern European sales markets over time. Exponentiating this term and multiplying it with Germany's start-of-period exports to CE gives rise to $\Delta EX_{Gjt}^{grav} = EX_{Gjt-1} \times \exp^{\bar{\epsilon}_{jt}^{EX} - \bar{\epsilon}_{jt-1}^{EX}}$, which is a counterfactual measure of changes in German industries' export exposure that is solely driven by CE 's increasing attractiveness as sales market.

Online Appendix F Subsample Results for the Effect of Trade T on Labor M and Voting Y

Column 1 of [Online Appendix F table 5](#) reports on total employment (as in the paper’s table 3), estimated separately for Period 1 (1987–1998), and Period 2 (1998–2009), as well as for West Germany only in Period 2. Columns 2–9 of [Online Appendix F table 5](#) similarly decompose the same eight political outcomes as reported in table 2. The sample sizes are 322, 408, and 322 respectively.

The discussion in section ?? suggests that the effect of trade shocks on labor markets should be more pronounced in the second period, when companies were more flexible to react. We found some evidence for this pattern in the individual results in table 11. This motivates us to decompose the effect of trade exposure on local labor markets by period in this section.

Table 5 reports on manufacturing’s share in employment (column 1) and on the eight voting outcomes, with each results estimated separately for Period 1 (1987–1998), and Period 2 (1998–2009), as well as for West Germany only in Period 2. The sample sizes are respectively 322, 408, and again 322.

Comparing the three sub-panels shows evidence for increasing flexibility in labor markets between Period 1 and Period 2. This is nicely reflected by the core result for manufacturing employment in column 1. The observed contrast between periods is not driven by the inclusion of East German regions in period 2. In fact, the contrast between the two periods is more pronounced once we focus on West Germany. Columns 2–9 show that voting responses to trade were strongest when labor markets were least regulated. Combining the evidence, table 5 suggests that trade exposure had the biggest effect on both voting and labor markets in the second period in West Germany, i.e. when labor markets were most deregulated and subject to market forces.

This evidence suggests doing the same breakdowns in the SOEP data. In table 6 we also split the individual-level SOEP results from the paper’s table 11 by period. We therefore report the results separately by period in columns 1 and 2. It turns out that both the extreme right and SPD results are driven entirely by period 2, i.e. after Germany’s labor markets were de-regulated. The individual-level results are thus also in the period-breakdowns consistent with the regional results.

We interpret this symmetry as “reduced form evidence” for the important role of labor markets as mediators in the transmission from trade shocks to voting responses. However, without additional econometric structure, it is not possible to infer on the causality of the labor market mechanisms.

Online Appendix Table 5: Decomposing the Results by Period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	log(Total Empl.)	Turnout	CDU / CSU	SPD	FDP	Greens	Right	Left	Small
<i>Period 1</i>									
T_{it}	-0.027 (1.600)	0.000 (0.013)	-0.298 (1.159)	0.320 (1.558)	0.013 (0.150)	-0.003 (0.030)	-0.025 (0.243)	-0.001 (0.041)	-0.007 (0.105)
<i>Period 2</i>									
T_{it}	-0.011 (1.311)	0.000 (0.080)	-0.115 (0.704)	-0.173 (1.072)	0.076 (0.821)	0.081 (1.142)	0.071* (1.696)	0.058 (0.360)	0.003 (0.044)
<i>Period 2, West only</i>									
T_{it}	-0.019** (1.990)	0.002 (0.514)	-0.095 (0.542)	-0.161 (0.987)	0.083 (0.886)	0.110 (1.342)	0.084** (2.078)	-0.023 (0.187)	0.001 (0.018)

Notes: The table reports subsample estimations. Column 1 reports on the log of total employment, as in table 3. Columns 2–9 report on the same eight political outcomes as in table 2. Every result reported in table 5 is from a 2SLS estimation that breaks treatment into separate import competition and export access effects, instrumented with Z_{it}^{IM} and Z_{it}^{EX} , defined in (2). Every panel additionally reports the results for three separate sub-samples: period 1 (1987–1998) and period 2 (1998–2009), and period 2 without the 86 East German districts. The sample sizes are 322, 408, and 322 respectively. All specifications include region fixed effects. Standard errors are clustered at the level of 96 commuting zones. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

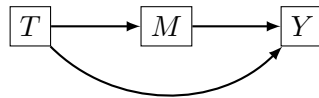
Online Appendix G Mediation Model without Confounding Variables

The simplest mediation model consists of three observed variables T, M, Y and three statistically independent error terms $\epsilon_T, \epsilon_M, \epsilon_Y$. Causal relations are defined by the following equations:

$$T = f_T(\epsilon_T), M = f_M(T, \epsilon_M), Y = f_Y(T, M, \epsilon_Y). \quad (62)$$

Input variables of functions f_T, f_M, f_Y are said to cause their respective output variables. Thus T causes M and Y while M causes Y . Neither functions f_T, f_M, f_Y nor error terms $\epsilon_T, \epsilon_M, \epsilon_Y$ are observed. We use the notations $\text{supp}(T), \text{supp}(M), \text{supp}(Y)$ for the support of T, M, Y respectively. Figure 1 displays Model (62) as a *Directed Acyclic Graph* (DAG).

Online Appendix Figure 1: Mediation Model without Confounding Variables



Fixing is a key concept in causal analysis. Fixing is defined as the causal operation that assigns a value to an argument of a structural equation. It is used to generate counterfactual variables. For instance, Model (62) renders three counterfactual variables. The counterfactual mediator $M(t)$ is generated by fixing the argument T of function f_M to a value $t \in \text{supp}(T)$, that is, $M(t) = f_M(t, \epsilon_M)$. The counterfactual outcome $Y(t)$ for T fixed at t is given by $Y(t) = f_Y(t, M(t), \epsilon_Y)$ and the counterfactual outcome Y when T is fixed at t and M is fixed at m is given by $Y(t, m) =$

Online Appendix Table 6: Individual-Level Analysis

	(1)	(2)	(3)	(4)
	1990-1998	1998-2009	Low-Skill & Manuf., 1998-2009	Low-Skill & Not Manuf., 1998-2009
<u><i>Established Parties:</i></u>				
Would Vote CDU/CSU	-0.025 (-0.743)	0.002 (0.227)	-0.006 (-0.350)	0.003 (0.257)
Would Vote SPD	0.027 (0.761)	-0.019** (-2.217)	0.001 (0.031)	-0.022** (-2.352)
Would Vote FDP	-0.038 (-0.720)	0.015 (1.177)	0.002 (0.116)	0.021 (1.431)
Would Vote Green Party	0.019 (0.409)	0.016 (1.295)	0.007 (0.363)	0.007 (0.565)
<u><i>Non-Established Parties:</i></u>				
Would Vote Extreme-Right Parties	0.029 (0.735)	0.028* (1.802)	0.088** (2.013)	0.016 (1.035)
Would Vote Far-Left Parties	-0.008 (-0.670)	-0.005 (-0.751)	0.026 (1.579)	-0.010 (-1.043)
Would Vote Other Small Parties	0.018 (0.340)	-0.012 (-1.072)	-0.048* (-1.674)	-0.001 (-0.042)
Period-by-region F.E.	Yes	Yes	Yes	Yes
Observations	3,694	5,975	1,168	3,817

Notes: (a) Columns 1–2 split the the paper’s table 11 by period (3,694 + 5,975= 9,669). The results are driven entirely by period 2, i.e. after Germany’s labor markets were de-regulated. No part of the political spectrum responds in period 1. In period 2, SPD support is reduced in response to trade exposure and support for the extreme right goes up. In columns 3–4, we focus on the second period, which sharpens the results of the worker-type breakdowns in the paper’s table 11. (b) Standard errors are clustered at the region level. *** p<0.01, ** p<0.05, * p<0.1.

$f_Y(t, m, \epsilon_Y)$. We refer to [Heckman and Pinto \(2015a\)](#) for a detailed discussion on the fixing operator, counterfactual outcomes and causal models.

We are interested in identifying the effect of T on outcome Y , but most importantly, we are interested in identifying the mechanism M through which T causes Y . This task is often referred to as mediation analysis and it requires the identification of all three counterfactual variables $Y(t)$, $M(t)$, $Y(m, t)$. [Robins and Greenland \(1992\)](#) examine the case of a binary treatment $\text{supp}(T) = \{t_0, t_1\}$ and define three primary causal parameters in mediation analysis: the *total*, *direct* and *indirect* effects.

$$\begin{aligned} \text{Total Eff. : } \quad ATE &= E(Y(t_1) - Y(t_0)) && \equiv E(Y(t_1, M(t_1)) - Y_i(t_0, M(t_0))), \\ \text{Direct Eff. : } \quad ADE(t) &= E(Y(t_1, M(t)) - Y(t_0, M(t))) && \equiv \int E(Y(t_1, m) - Y(t_0, m)) dF_{M(t)}(m), \\ \text{Indirect Eff. : } \quad AIE(t) &= E(Y(t, M(t_1)) - Y(t, M(t_0))) && \equiv \int E(Y(t, m)) [dF_{M(t_1)}(m) - dF_{M(t_0)}(m)], \end{aligned}$$

where $F_{M(t)}(m)$ stand for the cumulative probability distribution of counterfactual mediator $M(t)$. ATE is the average causal effect of T on Y . $ADE(t)$ is the causal effect of T on Y when we hold the distribution of M fixed at $M(t)$. $AIE(t)$ is the causal effect of T on Y induced by the change in the distribution of the mediator M . The total effect ATE can be expressed as the sum of direct and indirect effect as:

$$\begin{aligned} ATE &= E(Y(t_1, M(t_1)) - Y_i(t_0, M(t_0))) \\ &= \left(E(Y(t_1, M(t_1))) - E(Y(t_0, M(t_1))) \right) + \left(E(Y(t_0, M(t_1)) - Y_i(t_0, M(t_0))) \right) = ADE(t_1) + AIE(t_0) \\ &= \left(E(Y(t_1, M(t_1))) - E(Y(t_1, M(t_0))) \right) + \left(E(Y(t_1, M(t_0)) - Y_i(t_0, M(t_0))) \right) = AIE(t_1) + ADE(t_0). \end{aligned}$$

Model (62) has no confounding variables. That is to say that model (62) assumes no unobserved variable that jointly causes T , M and Y . This implies that variables T , M are independent of counterfactual outcomes, that is, $T \perp\!\!\!\perp (Y(t), M(t))$ and $M \perp\!\!\!\perp Y(t, m)$. Indeed, $T = f_T(\epsilon_T)$ depends only on ϵ_T , $M(t) = f_M(t, \epsilon_M)$ and $Y(t) = f_Y(t, M(t), \epsilon_Y)$ only depend on ϵ_M, ϵ_Y . But ϵ_T is independent of ϵ_M, ϵ_Y . Thus we can write:

$$(\epsilon_Y, \epsilon_M) \perp\!\!\!\perp \epsilon_T \Rightarrow (f_Y(t, f_M(t, \epsilon_M), \epsilon_Y), f_M(t, \epsilon_M)) \perp\!\!\!\perp f_T(\epsilon_T) \Rightarrow (Y(t), M(t)) \perp\!\!\!\perp T$$

On the other hand, $Y(t, m) = f_Y(t, m, \epsilon_Y)$ only depends on ϵ_Y . Thus we can write:

$$\epsilon_Y \perp\!\!\!\perp (\epsilon_M, \epsilon_T) \Rightarrow f_Y(t, m, \epsilon_Y) \perp\!\!\!\perp f_M(f_T(\epsilon_T), \epsilon_M) \Rightarrow Y(t, m) \perp\!\!\!\perp M.$$

A substantial literature on mediation analysis assumes no confounding variables. This literature often evokes the Sequential Ignorability Assumption of [Imai et al. \(2010\)](#). [Online Appendix H](#) shows that Model (62) also implies Sequential Ignorability.

If the independence relations $T \perp\!\!\!\perp (Y(t), M(t))$ and $M \perp\!\!\!\perp Y(t, m)$ hold, then we are able to express average counterfactual outcomes in terms of conditioned expectations from observed data. We illustrate this fact for the counterfactual outcome $Y(t)$. The observed outcome Y can be expressed as:

$$Y = \sum_{t \in \text{supp}(T)} Y(t) \cdot \mathbf{1}[T = t],$$

where $\mathbf{1}[T = t]$ is an indicator function that takes value one if $T = t$ and zero otherwise. If

$T \perp\!\!\!\perp Y(t)$ holds then $E(Y(t)) = E(Y(t)|T = t)$ also holds and we can express $E(Y(t))$ as:

$$E(Y(t)) = E(Y(t)|T = t) = E\left(\sum_{t \in \text{supp}(T)} Y(t) \cdot \mathbf{1}[T = t] | T = t\right) = E(Y|T = t).$$

The expectation $E(Y|T = t)$ can be evaluated from observed data and thereby $E(Y(t))$ is said to be identified.

Online Appendix H The Sequential Ignorability Assumption

A large literature on mediation analysis relies on the Sequential Ignorability Assumption **A-1** of Imai et al. (2010) to identify mediation effects.

Assumption A-1 *Sequential Ignorability* (Imai et al., 2010):

$$(Y(t', m), M(t)) \perp\!\!\!\perp T|X \quad (63)$$

$$Y(t', m) \perp\!\!\!\perp M(t)|(T, X), \quad (64)$$

where X denotes pre-intervention variables that are not caused by T, M and Y such that $0 < P(T = t|X) < 1$ and $0 < P(M(t) = m|T = t, X) < 1$ holds for all $x \in \text{supp}(X)$ and $m \in \text{supp}(M)$.

Under Sequential Ignorability **A-1**, it is easy to show that the distributions of counterfactual variables are identified by $P(Y(t, m)|X) = P(Y|X, T = t, M = m)$ and $P(M(t)|X) = P(M|X, T = t)$ and thereby the mediating causal effects can be expressed as:

$$ADE(t) = \int \left(E(Y|T = t_1, M = m, X = x) - E(Y|T = t_0, M = m, X = x, X = x) \right) dF_{M|T=t, X=x}(m) dF_X(x) \quad (65)$$

$$AIE(t) = \int \left(E(Y|T = t, M = m, X = x) \left[dF_{M|T=t_1, X=x}(m) - dF_{M|T=t_0, X=x}(m) \right] \right) dF_X(x). \quad (66)$$

Imai, Tingley, Keele and Yamamoto offer a substantial line of research that explores the identifying properties of Sequential Ignorability Assumption **A-1**. See Imai et al. (2011b) for a comprehensive discussion of the benefits and limitations of the sequential ignorability assumption.

The main critics of Sequential Ignorability **A-1** is that it does not hold under the presence of either *Confounders* or *Unobserved Mediators* (Heckman and Pinto, 2015b).

The independence relation (63) assumes that T is exogenous conditioned on X . There exists no unobserved variable that causes T and Y or T and M . For instance, the Sequential Ignorability **A-1** holds for the model defined in (62) because:

$$(\epsilon_Y, \epsilon_M) \perp\!\!\!\perp \epsilon_T \Rightarrow (f_Y(t', m, \epsilon_Y), f_M(t, \epsilon_M)) \perp\!\!\!\perp f_T(\epsilon_T) \Rightarrow (Y(t', m), M(t)) \perp\!\!\!\perp T. \quad (67)$$

$$\epsilon_Y \perp\!\!\!\perp \epsilon_M | \epsilon_T \Rightarrow f_Y(t', m, \epsilon_Y) \perp\!\!\!\perp f_M(t, \epsilon_M) | f_T(\epsilon_T) \Rightarrow Y(t', m) \perp\!\!\!\perp M(t) | T, \quad (68)$$

where the initial independence relation in (67) and (68) comes from the independence of error terms.

This assumption is expected to hold in experimental data when treatment T is randomly assigned. The independence relation (64) assumes that M is exogenous conditioned on X and T . It assumes that no confounding variable causing M and Y . Sequential Ignorability **A-1** is an extension of the Ignorability Assumption of Rosenbaum and Rubin (1983) that also assumes that a treatment T is exogenous when conditioned on pre-treatment variables. Robins (2003); Petersen, Sinisi, and Van der Laan (2006); Rubin (2004) state similar identifying criteria that assume no confounding variables. Those assumptions are not testable.

Figure 7 in the paper reveals that Sequential Ignorability **A-1** assumes that: (1) the confounding variable V is observed, that is, the pre-treatment variables X ; and (2) that there is no unobserved mediator U . This assumption is unappealing for many because it solves the identification problem generated by confounding variables by assuming that those do not exist (Heckman, 2008).

Consider a change in the treatment variable T denoted by $\Delta(t) = t_1 - t_0$. The Direct and indirect effects can be expressed by:

$$\begin{aligned} ADE(t') &= \left(\lambda_{YT} \cdot t_1 + \lambda_{YM} \cdot E(M(t')) \right) - \left(\lambda_{YT} \cdot t_0 + \lambda_{YM} \cdot E(M(t')) \right) \\ \therefore ADE &= \lambda_{YT} \cdot \Delta(t) \end{aligned} \tag{69}$$

$$\begin{aligned} \text{and } AIE(t') &= \left(\lambda_{YT} \cdot t' + \lambda_{YM} \cdot E(M(t_1)) \right) - \left(\lambda_{YT} \cdot t' + \lambda_{YM} \cdot E(M(t_0)) \right) \\ &= \left(\lambda_{YT} \cdot t' + \lambda_{YM} \lambda_M \cdot t_1 \right) - \left(\lambda_{YT} \cdot t' + \lambda_{YM} \lambda_M \cdot t_0 \right) \\ \therefore AIE &= \lambda_{YM} \cdot \lambda_M \cdot \Delta(t) \end{aligned} \tag{70}$$

Online Appendix I Identification of Causal Parameters

The linear mediation model we investigate can be fully described by the following equations:

$$\text{Instrumental Variable } Z = \epsilon_Z, \quad (71)$$

$$\text{Treatment } T = \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T, \quad (72)$$

$$\text{Unobserved Mediator } U = \zeta_T \cdot T + \epsilon_U, \quad (73)$$

$$\text{Observed Mediator } M = \varphi_T \cdot T + \varphi_U \cdot U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (74)$$

$$\text{Outcome } Y = \beta_T \cdot T + \beta_M \cdot M + \beta_U \cdot U + \beta_V \cdot V_Y + \epsilon_Y, \quad (75)$$

$$\text{Exogenous Variables } Z, V_T, V_M, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y \text{ are statistically independent variables,} \quad (76)$$

$$\text{Scalar Coefficients } \xi_Z, \xi_V, \zeta_T, \varphi_T, \varphi_U, \delta_Y, \delta_T, \beta_T, \beta_M, \beta_U, \beta_V \quad (77)$$

$$\text{Unobserved Variables } V_T, V_M, U, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y. \quad (78)$$

We assume that all variables have mean zero. This assumption does not incur in loss of generality, but simplify notation as intercepts can be suppressed.

We first eliminate the unobserved mediator U from Equations (74)–(75) by iterated substitution. Equations (75)–(75) are then expressed as:

$$M = (\varphi_T + \varphi_U \zeta_T) \cdot T + \varphi_U \cdot \epsilon_U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (79)$$

$$Y = (\beta_T + \beta_U \zeta_T) \cdot T + \beta_M \cdot M + \beta_U \cdot \epsilon_U + \beta_V \cdot V_Y + \epsilon_Y. \quad (80)$$

We use the following transformation of parameters to save on notation:

$$\tilde{\varphi}_T = \varphi_T + \varphi_U \zeta_T, \quad (81)$$

$$\tilde{\beta}_T = \beta_T + \beta_U \zeta_T, \quad (82)$$

$$\tilde{U} = \epsilon_U. \quad (83)$$

We use equations (79)–(83) to simplify Model (71)–(75) into the following equations:

$$\text{Instrumental Variable } Z = \epsilon_Z, \quad (84)$$

$$\text{Treatment } T = \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T, \quad (85)$$

$$\text{Observed Mediator } M = \tilde{\varphi}_T \cdot T + \varphi_U \cdot \tilde{U} + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (86)$$

$$\text{Outcome } Y = \tilde{\beta}_T \cdot T + \beta_M \cdot M + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \quad (87)$$

In this linear model, the counterfactual outcomes $M(t), Y(t), Y(m), Y(m, t)$ are given by:

$$M(t) = \tilde{\varphi}_T \cdot t + \varphi_U \cdot \tilde{U} + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (88)$$

$$Y(m) = \tilde{\beta}_T \cdot T + \beta_M \cdot m + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \quad (89)$$

$$Y(t, m) = \tilde{\beta}_T \cdot t + \beta_M \cdot m + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \quad (90)$$

$$\begin{aligned} Y(t) &= \tilde{\beta}_T \cdot t + \beta_M \cdot M(t) + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \\ &= (\tilde{\beta}_T + \beta_M \tilde{\varphi}_T) \cdot t + (\beta_U + \beta_M \varphi_U) \cdot \tilde{U} + (\beta_V + \beta_M \delta_Y) \cdot V_Y + \beta_M \delta_T \cdot V_T + \beta_M \cdot \epsilon_M + \epsilon_Y. \end{aligned} \quad (91)$$

We claim that the coefficients associated with unobserved variables V_T, \tilde{U}, V_Y may only be identified up a linear transformation. Consider the coefficients δ_T, β_V that multiply the unobserved variable V_T in Equations (85) and (86) respectively. Suppose a linear transformation that multiplies V_T by a constant $\kappa \neq 0$. The model would remain the same if coefficients δ_T, β_V were divided by the same constant κ . This is a typical fact in the literature of linear factor models. We solve this non-identification problem by impose that each unobserved variable V_T, \tilde{U}, V_Y has unit variance:

$$\text{var}(V_T) = \text{var}(\tilde{U}) = \text{var}(V_Y) = 1. \quad (92)$$

Assumption (92) is typically termed as *anchoring* of unobserved factors in the literature of factor analysis. This assumption does not incur in any loss of generality for the identification of direct, indirect or total causal effects of T (and M) on Y as expressed in the following section.

Online Appendix I.1 Defining Causal Parameters

The literature of mediation analysis term relevant causal parameters as:

- Total Effect of T on Y , that is, $\frac{dE(Y(t))}{dt}$.
- Direct Effect of T on Y , that is $\frac{\partial E(Y(t, m))}{\partial t}$.
- Effect of M on Y , that is, $\frac{dE(Y(m))}{dm}$.
- Effect of T on M , that is, $\frac{dE(M(t))}{dt}$.
- Indirect Effect of T on Y , that is $\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}$.

According to the counterfactual variables in (88)–(91), these causal effects are given by:

$$\text{Total Effect of } T \text{ on } Y : \frac{dE(Y(t))}{dt} = \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T. \quad (93)$$

$$\text{Direct Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial t} = \tilde{\beta}_T. \quad (94)$$

$$\text{Effect of } M \text{ on } Y : \frac{dE(Y(m))}{dm} = \beta_M. \quad (95)$$

$$\text{Effect of } T \text{ on } M : \frac{dE(M(t))}{dt} = \tilde{\varphi}_T. \quad (96)$$

$$\text{Indirect Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt} = \beta_M \cdot \tilde{\varphi}_T. \quad (97)$$

Online Appendix I.2 Identifying Equations

Model (84)–(87) can be conveniently expressed in matrix notation. In Equation (98) we define $\mathbf{X} = [Z, T, M, Y]'$ as the vector of observed variables, $\mathbf{V} = [V_T, V_Y, \tilde{U}]'$ as the vector of unobserved confounding variables, and $\boldsymbol{\epsilon} = [\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y]'$ as the vector of exogenous error terms. According to (76), the random vectors \mathbf{V} and $\boldsymbol{\epsilon}$ are independent, that is, $\mathbf{V} \perp\!\!\!\perp \boldsymbol{\epsilon}$. We use \mathbf{K} in (98) for the matrix of parameters that multiply \mathbf{X} and \mathbf{A} for the matrix of parameters that multiply \mathbf{V} .

$$\mathbf{X} = \begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} V_T \\ V_Y \\ \tilde{U} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}, \quad \mathbf{K} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \tilde{\varphi}_T & 0 & 0 \\ 0 & \tilde{\beta}_T & \beta_M & 0 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ \xi_V & 0 & 0 \\ \delta_Y & \delta_Y & \varphi_U \\ 0 & \beta_V & \beta_U \end{bmatrix}. \quad (98)$$

Using the notation in (98), we can express the linear system (84)–(87) as following:

$$\underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \tilde{\varphi}_T & 0 & 0 \\ 0 & \tilde{\beta}_T & \beta_M & 0 \end{bmatrix}}_{\mathbf{K}} \cdot \underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ \xi_V & 0 & 0 \\ \delta_T & \delta_Y & \varphi_U \\ 0 & \beta_V & \beta_U \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{pmatrix} V_T \\ V_Y \\ \tilde{U} \end{pmatrix}}_{\mathbf{V}} + \underbrace{\begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}}_{\boldsymbol{\epsilon}}, \quad (99)$$

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon}. \quad (100)$$

The coefficients in matrices \mathbf{K} , \mathbf{A} are identified through the covariance matrices of observed variables. We use $\Sigma_{\mathbf{X}} = \text{cov}(\mathbf{X}, \mathbf{X})$ for the covariance matrix of observed variables \mathbf{X} , and $\Sigma_{\boldsymbol{\epsilon}} = \text{cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon})$ for the vector of error terms $\boldsymbol{\epsilon}$. $\Sigma_{\boldsymbol{\epsilon}}$ is a diagonal matrix due to statistical independence of error terms. We also use $\Sigma_{\mathbf{V}} = \text{cov}(\mathbf{V}, \mathbf{V})$ for the covariance of unobserved variables \mathbf{V} . The unobserved variables in \mathbf{V} are statistically independent and have unit variance (92), thus $\Sigma_{\mathbf{V}} = \mathbf{I}$ where \mathbf{I} is the identity matrix. Moreover, $\mathbf{V} \perp\!\!\!\perp \boldsymbol{\epsilon}$ implies that $\text{cov}(\mathbf{V}, \boldsymbol{\epsilon}) = \mathbf{0}$, where $\mathbf{0}$ is a matrix of elements zero.

Equation (103) determines the relation between the covariance matrices of observed and unobserved variables:

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon} \Rightarrow (\mathbf{K} - \mathbf{I}) \mathbf{X} = \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon}, \quad (101)$$

$$\Rightarrow (\mathbf{K} - \mathbf{I}) \Sigma_{\mathbf{X}} (\mathbf{K} - \mathbf{I})' = \mathbf{A} \Sigma_{\mathbf{V}} \mathbf{A}' + \Sigma_{\boldsymbol{\epsilon}}, \quad (102)$$

$$\Rightarrow (\mathbf{K} - \mathbf{I}) \Sigma_{\mathbf{X}} (\mathbf{K} - \mathbf{I})' = \mathbf{A} \mathbf{A}' + \Sigma_{\boldsymbol{\epsilon}}, \quad (103)$$

where the second equation is due to $\mathbf{V} \perp\!\!\!\perp \boldsymbol{\epsilon}$ and the third equations comes from $\Sigma_{\mathbf{V}} = \mathbf{I}$.

Equation (103) generates ten equalities. Four equalities are due to the diagonal of the covariance matrices $(\mathbf{K} - \mathbf{I}) \Sigma_{\mathbf{X}} (\mathbf{K} - \mathbf{I})'$ and $\mathbf{A} \mathbf{A}' + \Sigma_{\boldsymbol{\epsilon}}$ in (103). The remaining six equalities from the off-diagonal relation of the covariance matrices in (103).

The diagonal elements of $\Sigma_{\boldsymbol{\epsilon}}$ are the variances of the error terms $\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y$. Thereby each diagonal equation generated by (103) adds one unobserved term to the system of quadratic equations. The point-identification of the model coefficients arises from the six off-diagonal equations

generated by (103). Those are listed below:

$$\text{cov}(Z, T) - \text{cov}(Z, Z) \cdot \xi_Z = 0 \quad (104)$$

$$\text{cov}(Z, M) - \text{cov}(Z, T) \cdot \tilde{\varphi}_T = 0 \quad (105)$$

$$\text{cov}(Z, Y) - \text{cov}(Z, M) \cdot \beta_M - \text{cov}(Z, T) \cdot \tilde{\beta}_T = 0 \quad (106)$$

$$\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = 0 \quad (107)$$

$$\text{cov}(M, Y) - \text{cov}(T, M) \cdot \tilde{\beta}_T - \text{cov}(M, M) \cdot \beta_M = \beta_U \cdot \varphi_U + \beta_V \cdot \delta_Y \quad (108)$$

$$\text{cov}(T, M) - \text{cov}(T, T) \cdot \tilde{\varphi}_T = \delta_T \cdot \xi_V \quad (109)$$

Simple manipulation of Equations (104)–(109) generate the identification of the following parameters:

$$\xi_Z = \frac{\text{cov}(Z, T)}{\text{cov}(Z, Z)} \quad \text{from Eq.(104)} \quad (110)$$

$$\tilde{\varphi}_T = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} \quad \text{from Eq.(105)} \quad (111)$$

$$\beta_M = \frac{\text{cov}(Z, T) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)} \quad \text{from Eqs.(106)–(107)} \quad (112)$$

$$\tilde{\beta}_T = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \quad \text{from Eqs.(106)–(107)} \quad (113)$$

$$\beta_U \cdot \varphi_U + \beta_V \cdot \delta_Y = \text{cov}(M, Y) - \text{cov}(M, M) \cdot \beta_M - \text{cov}(T, M) \cdot \tilde{\beta}_T \quad \text{from Eq.(108)} \quad (114)$$

$$\delta_T \cdot \xi_V = \frac{\text{cov}(T, M) \text{cov}(Z, M) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(Z, M)} \quad \text{from Eq.(109)} \quad (115)$$

Moreover, if we divide Equation (106) by $\text{cov}(Z, T)$ we obtain:

$$\frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} \cdot \beta_M - \frac{\text{cov}(Z, T)}{\text{cov}(Z, T)} \cdot \tilde{\beta}_T = 0 \quad (116)$$

$$\Rightarrow \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \tilde{\varphi}_T \cdot \beta_M - \tilde{\beta}_T = 0 \quad (117)$$

$$\Rightarrow \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}. \quad (118)$$

The four causal of interest parameters defined in (93)–(96) are respectively identified by Equations (111), (112), (113) and (118):

$$\frac{dE(M(t))}{dt} = \tilde{\varphi}_T = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)}, \quad (119)$$

$$\frac{dE(Y(m))}{dm} = \beta_M = \frac{\text{cov}(Z, Y) \text{cov}(T, T) - \text{cov}(Y, T) \text{cov}(Z, T)}{\text{cov}(Z, M) \text{cov}(T, T) - \text{cov}(M, T) \text{cov}(Z, T)}, \quad (120)$$

$$\frac{\partial E(Y(t, m))}{\partial t} = \tilde{\beta}_T = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}, \quad (121)$$

$$\frac{dE(Y(t))}{dt} = \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}. \quad (122)$$

Next section explains that each causal effect (119)–(122) can be evaluated by standard Two-stage Least Squares regressions.

Online Appendix J Estimation of Causal Parameters

Our goal is to show that the four causal parameters listed in Equations (119)–(122) can be estimated using the standard Two-stage Least Square (2SLS) estimator. We revise the standard equations of the 2SLS estimators for sake of completeness.

Equations (123)–(124) present the first and stages of a generic 2SLS regression in which T stands for the endogenous variable, Z is the instrumental variable and Y is the targeted outcome.

$$\text{First Stage: } T = \kappa_1 + \beta_1 \cdot Z + \epsilon_1, \quad (123)$$

$$\text{Second Stage: } Y = \kappa_2 + \beta_2 \cdot T + \epsilon_2. \quad (124)$$

The 2SLS estimator relies on the assumptions that the instrument Z is statistically independent of the term ϵ_2 while T is not. It is well-known that the 2SLS estimator $\hat{\beta}_2$ is given by the ratio of the sample covariances $\text{cov}(Z, Y)$ and $\text{cov}(Z, T)$. Moreover $\hat{\beta}_2$ is a consistent estimator of parameter β_2 :

$$\text{plim}(\hat{\beta}_2) = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} = \beta_2. \quad (125)$$

Consider the inclusion of additional covariates X in both stages of the 2SLS method. Variables X in (126)–(127) play the role of control covariates in the first stage and second stages of the 2SLS estimator. Control covariates X directly causes Y in (127) while the instrument Z only causes Y though it impact on T .

$$\text{First Stage: } T = \kappa_1 + \beta_1 \cdot Z + \psi_1 \cdot X + \epsilon_1, \quad (126)$$

$$\text{Second Stage: } Y = \kappa_2 + \beta_2 \cdot T + \psi_2 \cdot X + \epsilon_2. \quad (127)$$

The 2SLS model (126)–(127) relies on the assumption that the instrument Z and control covariates X are independent of error term ϵ_2 , that is, $(Z, X) \perp\!\!\!\perp \epsilon_2$. The 2SLS estimator $\hat{\beta}_2$ for parameter β_2 is expressed by Equation (128) and it is a consistent estimator under model assumptions.

$$\text{plim}(\hat{\beta}_2) = \frac{\text{cov}(Z, Y) \text{cov}(X, X) - \text{cov}(Y, X) \text{cov}(Z, X)}{\text{cov}(Z, T) \text{cov}(X, X) - \text{cov}(T, X) \text{cov}(Z, X)} = \beta_2. \quad (128)$$

The 2SLS estimator $\hat{\psi}_2$ for parameter ψ_2 is expressed by Equation (129) and it is a consistent estimator under model assumptions.

$$\text{plim}(\hat{\psi}_2) = -\frac{\text{cov}(Z, Y) \text{cov}(T, X) - \text{cov}(Y, X) \text{cov}(Z, T)}{\text{cov}(Z, T) \text{cov}(X, X) - \text{cov}(T, X) \text{cov}(Z, X)} = \psi_2. \quad (129)$$

Each of the identification formulas for the causal effects in (119)–(122) describes a ratio of covariances that corresponds to one of the three 2SLS formulas (125), (128) or (128).

The effect of choice T on mediator M is given by:

$$\frac{dE(M(t))}{dt} = \tilde{\varphi}_T = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)}.$$

According to Equation (125), this effect can be estimated by the 2SLS regression (123)–(124) in which Z is the instrument, T is the endogenous variable and M is the outcome.

The total effect of T on outcome Y is given by:

$$\frac{dE(Y(t))}{dt} = \tilde{\varphi}_T \cdot \beta_M + \tilde{\beta}_T = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}.$$

According to Equation (125), this effect can be estimated by the 2SLS regression (123)–(124) in which Z is the instrument, T is the endogenous variable and Y is the outcome.

The causal effect of mediator M on outcome Y is given by:

$$\frac{dE(Y(m))}{dm} = \beta_M = \frac{\text{cov}(Z, Y) \text{cov}(T, T) - \text{cov}(Y, T) \text{cov}(Z, T)}{\text{cov}(Z, M) \text{cov}(T, T) - \text{cov}(M, T) \text{cov}(Z, T)},$$

which can be estimated by the 2SLS regression (123)–(124) where Z is the instrument, T is the endogenous variable and M is the outcome.

The causal effect of mediator M on outcome M is given by:

$$\frac{dE(Y(m))}{dm} = \beta_M = \frac{\text{cov}(Z, Y) \text{cov}(T, T) - \text{cov}(Y, T) \text{cov}(Z, T)}{\text{cov}(Z, M) \text{cov}(T, T) - \text{cov}(M, T) \text{cov}(Z, T)}.$$

According to the 2SLS estimator in (128), this causal effect can be estimated by $\hat{\beta}_2$ in the 2SLS regression (126)–(127) in which Z plays the role of the instrument, M is the endogenous variable, T is the control covariate and Y is the outcome.

The Indirect Effect of choice T on outcome Y is given by:

$$\frac{\partial E(Y(t, m))}{\partial m} = \tilde{\beta}_T = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}.$$

According to the 2SLS estimator in (129), this causal effect can be estimated by $\hat{\psi}_2$ in the 2SLS regression (126)–(127) in which Z plays the role of the instrument, M is the endogenous variable, T is the control covariate and Y is the outcome.

Online Appendix K Total, Indirect and Direct Effects under One Instrument

Online Appendix I.2 describes a linear mediation model whose primary causal effects are identified by the following equations:

$$\text{Total Effect of } T \text{ on } Y : \frac{dE(Y(t))}{dt} = \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}. \quad (130)$$

$$\text{Direct Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial t} = \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}. \quad (131)$$

$$\text{Effect of } M \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial m} = \frac{\text{cov}(Z, T) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)} \quad (132)$$

$$\text{Effect of } T \text{ on } M : \frac{dE(M(t))}{dt} = \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} \quad (133)$$

$$\text{Indirect Effect of } T \text{ on } Y : \frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}. \quad (134)$$

The literature of mediation analysis typically expresses the total effect of T on Y as the sum of its direct and indirect effects. In our notation, this decomposition is stated as following:

$$\underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}} = \underbrace{\frac{\partial E(Y(t, m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}}. \quad (135)$$

We show that the decomposition described in (135) is exact in the case of a single instrument. That is to say that the covariance ratio that identifies the total effect of T on Y in equation (130) is equal to the covariance ratio that identifies the direct effect in Equations (131) plus the multiplication of the covariance ratios that identify the effect of T on M in (133) and the effect of M on Y

described in Equation (132). We thank David Slichter for pointing this fact.

$$\begin{aligned}
& \underbrace{\frac{\partial E(Y(t, m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t, m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}} \\
&= \underbrace{\frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}}_{\frac{\partial E(Y(t, m))}{\partial t}} + \underbrace{\frac{\text{cov}(Z, T) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)}}_{\frac{\partial E(Y(t, m))}{\partial m}} \cdot \underbrace{\frac{\text{cov}(Z, M)}{\text{cov}(Z, T)}}_{\frac{dE(M(t))}{dt}} \\
&= \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} + \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(T, T) \text{cov}(Z, Y)}{\text{cov}(T, M) \text{cov}(Z, T) - \text{cov}(T, T) \text{cov}(Z, M)} \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} \\
&= \frac{\text{cov}(Z, M) \text{cov}(T, Y) - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} + \frac{\text{cov}(T, T) \text{cov}(Z, Y) \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} - \text{cov}(Z, M) \text{cov}(T, Y)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, Y) \frac{\text{cov}(Z, M)}{\text{cov}(Z, T)} - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \text{cov}(Z, Y) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \text{cov}(Z, Y) \text{cov}(T, M) \frac{\text{cov}(Z, T)}{\text{cov}(Z, T)}}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \frac{\text{cov}(T, T) \text{cov}(Z, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} - \text{cov}(Z, T) \text{cov}(T, M) \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \\
&= \left(\frac{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)}{\text{cov}(T, T) \text{cov}(Z, M) - \text{cov}(Z, T) \text{cov}(T, M)} \right) \cdot \left(\frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} \right) \\
&= \frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)} = \underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}}.
\end{aligned}$$

The first equality expresses the total effect of T on Y in terms of its direct and indirect effects. The second equality substitutes the direct and indirect effects by their identification formulas described in (131), (132) and (130). The third equation isolates and eliminates the common term $\text{cov}(Z, M)$ in the denominator of $\frac{dE(Y(m))}{dm}$. The fourth equation flips the sign of the terms in the last covariance ratio. Now the overall sum has the same denominator. The fifth equation eliminates the common term in the sum of the numerators of both ratios. The sixth equation exchange the covariances $\text{cov}(Z, M)$ and $\text{cov}(Z, Y)$ of the first term of the numerator. The seventh equation includes the term $\frac{\text{cov}(Z, T)}{\text{cov}(Z, T)}$ which is equal to one. The eighth equation exchange the covariances $\text{cov}(Z, Y)$ and $\text{cov}(Z, T)$ of the second term of the numerator. The ninth equation isolates the common denominator of the expression. The tenth equation eliminates the common first term of both numerator and denominator. The resulting formula is the covariate ratio $\frac{\text{cov}(Z, Y)}{\text{cov}(Z, T)}$ which, according to (130), is equal to the total effect of choice T on outcome Y .

Online Appendix L Model Specification Test

The nonparametric version of the restricted model is given by the following equations:

$$T = f_T(Z, V_T, \epsilon_T), \quad (136)$$

$$U = f_U(T, \epsilon_U), \quad (137)$$

$$M = f_M(T, U, V_T, \epsilon_M), \quad (138)$$

$$Y = f_Y(T, M, U, V_Y, \epsilon_Y), \quad (139)$$

$$Z \perp\!\!\!\perp V_Y \perp\!\!\!\perp V_T \perp\!\!\!\perp \epsilon_Y \perp\!\!\!\perp \epsilon_M \perp\!\!\!\perp \epsilon_U \perp\!\!\!\perp \epsilon_T. \quad (140)$$

The identification of causal effects in our mediation model exploits three exclusion restrictions:

$$Z \perp\!\!\!\perp Y(t), \quad Z \perp\!\!\!\perp M(t), \quad \text{and} \quad Z \perp\!\!\!\perp Y(m)|T. \quad (141)$$

Exclusion Restrictions (141) do not depend on the linearity and hold for the nonparametric model defined by (136)–(140). Exclusion restrictions alone do not guarantee identification. The literature on instrumental variables offers a range of additional assumptions that enable identification of causal effects. Examples of such additional assumptions are: monotonicity (Imbens and Angrist, 1994), separability of (Heckman and Vytlacil, 2005), control function (Blundell and Powell, 2004) or revealed preference analysis (Pinto, 2015).

The three exclusion restrictions in (141) arise from the restrictions imposed on the causal relations of the unobserved variables V_T and V_Y . Specifically the Mediation Model (136)–(140) assumes that V_T jointly causes T and M while V_Y causes M and Y jointly, but neither V_T or V_Y causes T , M and Y simultaneously. A more general model allows for a common unobserved variable V causes T , M and Y simultaneously, as described by Equations (142)–(146).

$$T = f_T(Z, V, \epsilon_T), \quad (142)$$

$$U = f_U(T, \epsilon_U), \quad (143)$$

$$M = f_M(T, U, V, \epsilon_M), \quad (144)$$

$$Y = f_Y(T, M, U, V, \epsilon_Y), \quad (145)$$

$$Z \perp\!\!\!\perp V \perp\!\!\!\perp \epsilon_Y \perp\!\!\!\perp \epsilon_M \perp\!\!\!\perp \epsilon_U \perp\!\!\!\perp \epsilon_T. \quad (146)$$

Exclusion restriction $Z \perp\!\!\!\perp Y(m)|T$ does not hold in the general model described by Equations (142)–(146). Exclusion restrictions $Z \perp\!\!\!\perp Y(t)$ and $Z \perp\!\!\!\perp M(t)$ hold for both the Restricted Model (136)–(140) and the General Model (142)–(146).

For sake of clarity, we term the mediation model described by (136)–(140) as the Restricted Model. In contrast, we term the mediation model described by (142)–(146) as the General model.

Our goal is to test whether the Restricted Model (136)–(140) holds instead of the General model (142)–(146). That is to say, we want to test whether an unobserved random variable V jointly causes the treatment T , Mediator M and outcome Y . To do so, we evoke the linearity assumption of Online Appendix I that is used in the empirical estimation of our mediation model. Our aim is not to test those linear assumptions, instead we assume linearity to test the causal relations in the Restricted Model (136)–(140) against the ones in the General Model (142)–(146).

We show that an instrumental variable consisting of a single variable does not generate a test that infers if the Restricted Model (136)–(140) is rejected in favor of the General Model (142)–(146). nevertheless we show that a model specification test can be generated if we have two instrumental variables. By two instruments we mean two random variables that cause the treatment variable

T but are not caused by the unobserved confounding variables. As mentioned, we maintain the assumption of linearity for both General and Restricted models. Our test bares some similarities with the the Sargan-Hansen test that exploits model over-identifying restrictions to do inference on model coefficients.

Online Appendix L.1 The Case of a General Model with a Single Instrumental Variable

In this section we compare the Restricted Mediation Model that is assumed to hold in our estimations with the General Mediation Model that allows for an unobserved variable V to cause the three main variables we examine: treatment T , Mediator M and outcome Y .

Figure 7 summarizes key properties and differences of these two models. Panel A of Figure 7 presents the Directed Acyclic Graphs (DAG) of the restricted model examined in the paper and the General Mediation model that does not assume the restriction on the causal restriction on confounding variables. Panel B presents the structural equations associated with each model. Panel C presents the linear equations that subsume the causal relations described in each model. Panel D displays the equalities generated by the covariance structure arising from the linear equations. Those are used to identify model coefficients.

Let the General Linear mediation Model with one Instrumental Variable be described by the following equations:

$$\text{Instrumental Variable } Z = \epsilon_Z, \quad (147)$$

$$\text{Treatment } T = \xi_Z \cdot Z + \xi_V \cdot V + \epsilon_T, \quad (148)$$

$$\text{Unobserved Mediator } U = \zeta_T \cdot T + \epsilon_U, \quad (149)$$

$$\text{Observed Mediator } M = \varphi_T \cdot T + \varphi_U \cdot U + \delta \cdot V + \epsilon_M, \quad (150)$$

$$\text{Outcome } Y = \beta_T \cdot T + \beta_M \cdot M + \beta_U \cdot U + \beta_V \cdot V + \epsilon_Y, \quad (151)$$

$$\text{Exogenous Variables } Z, V, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y \text{ are statistically independent variables,} \quad (152)$$

$$\text{Scalar Coefficients } \xi_Z, \xi_V, \zeta_T, \varphi_T, \varphi_U, \delta, \beta_T, \beta_M, \beta_U, \beta_V \quad (153)$$

$$\text{Unobserved Variables } V, U, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y. \quad (154)$$

We assume that all variables have mean zero. This assumption does not incur in less of generality, but simplify notation as intercepts can be suppressed.

The main difference between the Restricted Model (Equations (71)–(78) of [Online Appendix I](#)) and the (Equations (147)–(154) above) resides on the equations that define the data generating process of the mediator M . The Restricted Model evokes two terms $\delta_Y \cdot V_Y$ and $\delta_T \cdot V_T$ in Equation (150) while these terms are subsumed by the term $\delta \cdot V$ in Equation (150) of the General Model. The rest of the coefficients share the same notation of coefficients in both models. The unobserved variable V_Y that causes outcome Y in the Restricted Model (75) is replaced by the unobserved variable V in the General Model (151).

We eliminate the unobserved mediator U from Equations (150)–(151) in the same fashion that U is eliminated from Equations (74)–(75) of [Online Appendix I](#). The new equations are:

$$M = (\varphi_T + \varphi_U \zeta_T) \cdot T + \varphi_U \cdot \epsilon_U + \delta \cdot V + \epsilon_M, \quad (155)$$

$$Y = (\beta_T + \beta_U \zeta_T) \cdot T + \beta_M \cdot M + \beta_U \cdot \epsilon_U + \beta_V \cdot V + \epsilon_Y. \quad (156)$$

We use the same change in notation as performed in the restricted model: $\tilde{\varphi}_T = \varphi_T + \varphi_U \zeta_T$,

$\tilde{\beta}_T = \beta_T + \beta_U \zeta_T$, and $\tilde{U} = \epsilon_U$. Equation (147)–(151) are therefore simplified to:

$$\text{Instrumental Variable } Z = \epsilon_Z, \quad (157)$$

$$\text{Treatment } T = \xi_Z \cdot Z + \xi_V \cdot V + \epsilon_T, \quad (158)$$

$$\text{Observed Mediator } M = \tilde{\varphi}_T \cdot T + \varphi_U \cdot \tilde{U} + \delta \cdot V + \epsilon_M, \quad (159)$$

$$\text{Outcome } Y = \tilde{\beta}_T \cdot T + \beta_M \cdot M + \beta_U \cdot \tilde{U} + \beta_V \cdot V + \epsilon_Y. \quad (160)$$

Model (157)–(160) can be conveniently expressed in matrix notation: Equation (161) of the General Model is the counterpart of Equation (98) of the restricted model in [Online Appendix I.2](#).

Following previous notation, we use $\mathbf{X} = [Z, T, M, Y]'$ for the vector of observed variables, and $\varepsilon = [\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y]'$ for the vector of exogenous error terms. The vector the vector of unobserved variables that generate endogenous variables is defined as $\mathbf{V}_G = [V, \tilde{U}]'$. According to (152), the random vectors \mathbf{V}_G and ε are independent, that is, $\mathbf{V}_G \perp\!\!\!\perp \varepsilon$. We use \mathbf{K} for the matrix of parameters that multiply \mathbf{X} and \mathbf{A}_G for the matrix of parameters that multiply \mathbf{V}_G .

$$\mathbf{X} = \begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}, \quad \mathbf{K} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \tilde{\varphi}_T & 0 & 0 \\ 0 & \tilde{\beta}_T & \beta_M & 0 \end{bmatrix}, \quad \mathbf{A}_G = \begin{bmatrix} 0 & 0 \\ \xi_V & 0 \\ \delta & \varphi_U \\ \beta_V & \beta_U \end{bmatrix}, \quad \mathbf{V}_G = \begin{pmatrix} V \\ \tilde{U} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}. \quad (161)$$

Using the notation defined in (98), we can express the linear system (84)–(87) as following:

$$\underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \tilde{\varphi}_T & 0 & 0 \\ 0 & \tilde{\beta}_T & \beta_M & 0 \end{bmatrix}}_{\mathbf{K}} \cdot \underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} 0 & 0 \\ \xi_V & 0 \\ \delta & \varphi_U \\ \beta_V & \beta_U \end{bmatrix}}_{\mathbf{A}_G} \cdot \underbrace{\begin{pmatrix} V \\ \tilde{U} \end{pmatrix}}_{\mathbf{V}_G} + \underbrace{\begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}}_{\varepsilon}, \quad (162)$$

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A}_G \cdot \mathbf{V}_G + \varepsilon. \quad (163)$$

The identification of coefficients in \mathbf{K} , \mathbf{A}_G depends on the covariance matrix of the observed data. We follow the notation of [Online Appendix I.2](#) closely. We use $\Sigma_{\mathbf{X}} = \text{cov}(\mathbf{X}, \mathbf{X})$ for the covariance matrix of observed variables \mathbf{X} , and $\Sigma_{\varepsilon} = \text{cov}(\varepsilon, \varepsilon)$ for the vector of error terms ε . Σ_{ε} is a diagonal matrix due to statistical independence of error terms. We also use $\Sigma_{\mathbf{V}_G} = \text{cov}(\mathbf{V}_G, \mathbf{V}_G)$ for the covariance of unobserved variables \mathbf{V}_G . The unobserved variables in \mathbf{V}_G are statistically independent and have variance one, thus we have that $\Sigma_{\mathbf{V}_G} = \mathbf{I}$, where \mathbf{I} is the identity matrix of dimension 2. Moreover, $\mathbf{V}_G \perp\!\!\!\perp \varepsilon$ implies that $\text{cov}(\mathbf{V}_G, \varepsilon) = \mathbf{0}$, where $\mathbf{0}$ is a matrix of zero elements.

Equation (164) determines the relation between the covariance matrices of observed and unobserved variables:

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A}_G \cdot \mathbf{V}_G + \varepsilon \Rightarrow (\mathbf{K} - \mathbf{I}) \Sigma_{\mathbf{X}} (\mathbf{K} - \mathbf{I})' = \mathbf{A}_G \mathbf{A}_G' + \Sigma_{\varepsilon}, \quad (164)$$

where the second equation is due to $\mathbf{V}_G \perp\!\!\!\perp \varepsilon$ and the third equations comes from $\Sigma_{\mathbf{V}_G} = \mathbf{I}$. Equation (164) generates ten equalities: four equalities are generated by the equality of the diagonal of the covariance matrices and six equations from the off-diagonal relation of the covariance matrices.

The identification analysis of the coefficients of the General Model arises from the six off-

diagonal equations generated by (164). Those are listed below:

$$\text{cov}(Z, T) - \text{cov}(Z, Z) \cdot \xi_Z = 0 \quad (165)$$

$$\text{cov}(Z, M) - \text{cov}(Z, T) \cdot \tilde{\varphi}_T = 0 \quad (166)$$

$$\text{cov}(Z, Y) - \text{cov}(Z, M) \cdot \beta_M - \text{cov}(Z, T) \cdot \tilde{\beta}_T = 0 \quad (167)$$

$$\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = \beta_V \cdot \xi_V \quad (168)$$

$$\text{cov}(M, Y) - \text{cov}(T, M) \cdot \tilde{\beta}_T - \text{cov}(M, M) \cdot \beta_M = \beta_U \cdot \varphi_U + \beta_V \cdot (\delta + \xi_V \cdot \tilde{\varphi}_T) \quad (169)$$

$$\text{cov}(T, M) - \text{cov}(T, T) \cdot \tilde{\varphi}_T = \delta \cdot \xi_V \quad (170)$$

Equations (165)–(170) of the General Mediation Model are the symmetric to Equations (104)–(109) of the Restricted Mediation Model.

Equations (165)–(166) are identical to Equations (104)–(105) of the restricted model. Therefore ξ_Z and $\tilde{\varphi}_T$ are identified by the same covariance ratios presented in Equations (110)–(111).

Even though Equation (167) is identical to Equation (106) of the restricted model, Equation (168) differs from Equation (107). Equation (107) of the Restricted Model states that

$$\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = 0.$$

The counterpart of (168) in the General model states that

$$\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = \beta_V \cdot \xi_V.$$

As a consequence, $\tilde{\beta}_T$ and β_M are not identified in the General Mediation Model. Moreover, the comparison of the covariance structure of both models does not allow to distinguish one model from the other. We conclude that a single instrument is insufficient to generate overidentifying restrictions that enable us to verify if observed data arises from the restricted or the general model.

Online Appendix L.2 Restricted Model with Multiple Instruments

In this section we show that two (or more) instrumental variables generate over-identifying restrictions that enables to perform a model specification test. The equations presented in this section follow closely the ones presented in Online Appendix I.2. The restricted model with two instrumental variables is described by the following equations:

$$\text{Vector of Instrumental Variables } \mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad (171)$$

$$\text{Treatment } T = \boldsymbol{\xi}_Z' \cdot \mathbf{Z} + \xi_V \cdot V_T + \epsilon_T, \text{ where } \boldsymbol{\xi}_Z = \begin{bmatrix} \xi_{Z,1} \\ \xi_{Z,2} \end{bmatrix}, \quad (172)$$

$$\text{Observed Mediator } M = \tilde{\varphi}_T \cdot T + \varphi_U \cdot \tilde{U} + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \quad (173)$$

$$\text{Outcome } Y = \tilde{\beta}_T \cdot T + \beta_M \cdot M + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \quad (174)$$

Model (171)–(174) can be conveniently expressed in matrix notation in the same fashion as Equations (98) of Online Appendix I.2.

The covariance equation (103) of Online Appendix I.2 also holds. The identification of model coefficients relies on the equations governing the covariance matrix of observed variables. These

identifying equations for the model (104)–(109) are given by:

$$\text{cov}(\mathbf{Z}, T) - \text{cov}(\mathbf{Z}, \mathbf{Z}) \cdot \boldsymbol{\xi}_Z = 0 \quad (175)$$

$$\text{cov}(\mathbf{Z}, M) - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\varphi}_T = 0 \quad (176)$$

$$\text{cov}(\mathbf{Z}, Y) - \text{cov}(\mathbf{Z}, M) \cdot \beta_M - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\beta}_T = 0 \quad (177)$$

$$\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = 0 \quad (178)$$

$$\text{cov}(M, Y) - \text{cov}(T, M) \cdot \tilde{\beta}_T - \text{cov}(M, M) \cdot \beta_M = \beta_U \cdot \varphi_U + \beta_V \cdot \delta_Y \quad (179)$$

$$\text{cov}(T, M) - \text{cov}(T, T) \cdot \tilde{\varphi}_T = \delta_T \cdot \xi_V \quad (180)$$

Equation (175) represent a system of two linear equations associated with each instrumental variable in $\mathbf{Z} = [Z_1, Z_2]'$. Equation (175) enables the identification of the vector of coefficients $\boldsymbol{\xi}_Z = [\xi_{Z,1}, \xi_{Z,2}]'$.

Equation (176) also represents a system of two linear equations that allows for the identification of the coefficient $\tilde{\varphi}_T$. Parameter $\tilde{\varphi}_T$ is overidentified as there are two linear equations that allow for the identification of the parameter.

Equation (177) represents a system of two linear equations that enable us to identify two coefficients: β_M and $\tilde{\beta}_T$. Equation (178) constitute an overidentification restriction for parameters β_M and $\tilde{\beta}_T$. This result differs from the identification using a single instrumental variable. We explain in Online Appendix I.2 that parameters β_M and $\tilde{\beta}_T$ required the use of two equations, that is (106) and (107), which are the counterpart of equations (177) and (178) above. In summary, Equation (178) constitute an overidentified restriction in the case of two instrumental variables while it is a necessary equation for the identification of $\beta_M, \tilde{\beta}_T$ when the dimension of the instrumental variable is equal to one.

Equations (178)–(180) are identical to (107)–(109) in Online Appendix I.2. Equation (179) enables the identification of the sum $\beta_U \cdot \varphi_U + \beta_V \cdot \delta_Y$ and Equation (180) identifies $\delta_T \cdot \xi_V$.

We conclude that the advent of more than one instrument does not render the identification of additional parameters. Instead it changes the identification status of parameters $\tilde{\varphi}_T, \beta_M$ and $\tilde{\beta}_T$ from just-identified to overidentified.

Online Appendix L.3 General Model with Multiple Instruments

The general model allows for an unobserved variable V to cause T, M and Y jointly. The equations presented in this section follow the ones presented in Online Appendix L.1. The general model with two instrumental variables stems from Equations (157)–(160) of Online Appendix L.1. The equations of the general model with two instrumental variables are displayed below:

$$\text{Vector of Instrumental Variables } \mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad (181)$$

$$\text{Treatment } T = \boldsymbol{\xi}'_Z \cdot \mathbf{Z} + \xi_V \cdot V + \epsilon_T, \text{ where } \boldsymbol{\xi}_Z = \begin{bmatrix} \xi_{Z,1} \\ \xi_{Z,2} \end{bmatrix}, \quad (182)$$

$$\text{Observed Mediator } M = \tilde{\varphi}_T \cdot T + \varphi_U \cdot \tilde{U} + \delta \cdot V + \epsilon_M, \quad (183)$$

$$\text{Outcome } Y = \tilde{\beta}_T \cdot T + \beta_M \cdot M + \beta_U \cdot \tilde{U} + \beta_V \cdot V + \epsilon_Y. \quad (184)$$

We can retrace the same steps described in [Online Appendix L.1](#) to generate the following identifying equations:

$$\text{cov}(\mathbf{Z}, T) - \text{cov}(\mathbf{Z}, \mathbf{Z}) \cdot \xi_Z = 0 \quad (185)$$

$$\text{cov}(\mathbf{Z}, M) - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\varphi}_T = 0 \quad (186)$$

$$\text{cov}(\mathbf{Z}, Y) - \text{cov}(\mathbf{Z}, M) \cdot \beta_M - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\beta}_T = 0 \quad (187)$$

$$\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = \beta_V \cdot \xi_V \quad (188)$$

$$\text{cov}(M, Y) - \text{cov}(T, M) \cdot \tilde{\beta}_T - \text{cov}(M, M) \cdot \beta_M = \beta_U \cdot \varphi_U + \beta_V \cdot (\delta + \xi_V \cdot \tilde{\varphi}_T) \quad (189)$$

$$\text{cov}(T, M) - \text{cov}(T, T) \cdot \tilde{\varphi}_T = \delta \cdot \xi_V \quad (190)$$

Equations (185)–(187) of the general model with two instrumental variables are identical to Equations (175)–(177) of the restricted model. Those equations enable the identification of the parameters $\xi_Z, \tilde{\varphi}_T, \beta_M, \tilde{\beta}_T$. Equations (189)–(190) of the general model are also identical to Equations (179)–(180) of the restricted model.

The key difference between the two models arise from Equation (178) of the Restricted Model and Equation (188) of the General Model. While (178) states that $\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = 0$, Equation (188) states that $\text{cov}(T, Y) - \text{cov}(T, T) \cdot \tilde{\beta}_T - \text{cov}(T, M) \cdot \beta_M = \beta_V \cdot \xi_V$.

Online Appendix L.4 Inference on General versus the Restricted Models

A simple model specification test can be performed by exploring Equations (177)–(178) of the Restricted model. Those two equations can be interpreted as a consequence of two moment conditions presented as following:

$$E(\mathbf{Z} \cdot (Y - \tilde{\beta}_T \cdot T - \beta_M \cdot M)) = 0 \Rightarrow \text{cov}(\mathbf{Z}, Y) = \text{cov}(\mathbf{Z}, M) \cdot \beta_M + \text{cov}(\mathbf{Z}, T) \cdot \tilde{\beta}_T \quad (191)$$

$$E(T \cdot (Y - \tilde{\beta}_T \cdot T - \beta_M \cdot M)) = 0 \Rightarrow \text{cov}(T, Y) = \text{cov}(T, T) \cdot \tilde{\beta}_T + \text{cov}(T, M) \cdot \beta_M \quad (192)$$

Moment Conditions (191)–(192) can be combined into a single equality using the matrix notation below:

$$E \left(\begin{bmatrix} \mathbf{Z} \\ T \end{bmatrix} \cdot \left(Y - [M, T] \cdot \begin{bmatrix} \beta_M \\ \tilde{\beta}_T \end{bmatrix} \right) \right) = 0 \Rightarrow \begin{bmatrix} \text{cov}(\mathbf{Z}, Y) \\ \text{cov}(T, Y) \end{bmatrix} = \begin{bmatrix} \text{cov}(\mathbf{Z}, M) & \text{cov}(\mathbf{Z}, T) \\ \text{cov}(T, M) & \text{cov}(T, T) \end{bmatrix} \cdot \begin{bmatrix} \beta_M \\ \tilde{\beta}_T \end{bmatrix} \quad (193)$$

We can apply the Generalized Method of Moments (GMM) of [Hansen \(1982\)](#) to the Moment Condition (193). The GMM estimator generated by moment (193) is given by:

$$\begin{bmatrix} \widehat{\beta_{M,1}} \\ \widehat{\tilde{\beta}_{T,1}} \end{bmatrix} = ([M, T]' \cdot \mathbf{P}_{Z,T} [M, T])^{-1} \cdot ([M, T]' \cdot \mathbf{P}_{Z,T} \cdot Y), \quad (194)$$

$$\text{such that } \mathbf{P}_{Z,T} = [Z, T] \cdot ([Z, T]' \cdot [Z, T])^{-1} \cdot [Z, T]', \quad (195)$$

where T, M, Y are $N \times 1$ data vectors associated respectively with observed variables T, M, Y ; Z is a $N \times K$ matrix of data associated with K instrumental variables; N denotes sample size and $\mathbf{P}_{Z,T}$ stands for the orthogonal projection on the space generated by the columns of $[Z, T]$.

The GMM estimator in (194) can be interpreted as a Two Stage Least Square regression (126)–(127), in which \mathbf{Z} plays the role of instrumental variables, M is the endogenous variable, T is a conditioning variable in both first and second stages and Y is the outcome.

The General Mediation Model (181)–(184) differs from the restricted model as Equality (192) does not hold. Nevertheless, the GMM method can be applied to Moment Condition (191), that

is, $E(\mathbf{Z} \cdot (Y - \tilde{\beta}_T \cdot T - \beta_M \cdot M)) = 0$, which still hold in the general model. GMM estimator (194) for the the general model that is based only on Moment Condition (191) is given by:

$$\begin{bmatrix} \widehat{\beta}_{M,2} \\ \widetilde{\beta}_{T,2} \end{bmatrix} = ([\mathbf{M}, \mathbf{T}]' \cdot \mathbf{P}_Z [\mathbf{M}, \mathbf{T}])^{-1} \cdot ([\mathbf{M}, \mathbf{T}]' \cdot \mathbf{P}_Z \cdot \mathbf{Y}), \quad (196)$$

$$\text{such that } \mathbf{P}_Z = \mathbf{Z} \cdot (\mathbf{Z}' \cdot \mathbf{Z})^{-1} \cdot \mathbf{Z}'. \quad (197)$$

The GMM Estimator (196) can be interpreted as the Two Stage Least Square regression (123) (124) in which \mathbf{Z} plays the role of instrumental variables, Y is the outcome and both M and T are endogenous variables.

If the causal assumptions of the restricted model hold, then both GMM estimators (194) and (196) provide consistent estimates of β_M and $\tilde{\beta}_T$. If the causal assumptions of the restricted model do not hold, then (194) does not generates a consistent estimate of $\widehat{\beta}_M, \widetilde{\beta}_T$. Thus large differences between the estimates $\widehat{\beta}_{M,1}, \widetilde{\beta}_{T,1}$ of (194) versus $\widehat{\beta}_{M,2}, \widetilde{\beta}_{T,2}$ of (196) provide statistical evidence against the null hypothesis that the causal assumptions of the Restrictive Model holds.

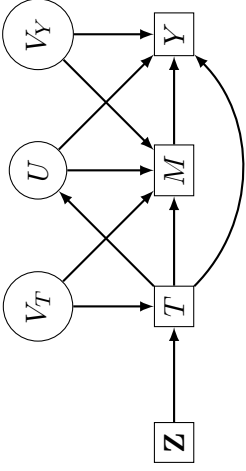
Online Appendix M Additional Observed Mechanisms

In this section, we apply our analysis to the other four outcomes without worrying about aggregation. To this end, the top-panel of table 8 reports the results of estimating equation (5), i.e. the causal effect of import competition on labor markets, for these other outcomes. It is essentially an extension to the results reported in table 3 for (i) total employment. Both (ii) manufacturing's employment share and (iii) manufacturing wages are consistently significantly impacted by import competition. There is also some, albeit weaker, evidence that import competition impacted (iv) non-manufacturing wages (in columns 1–2), and (v) unemployment (in column 5).

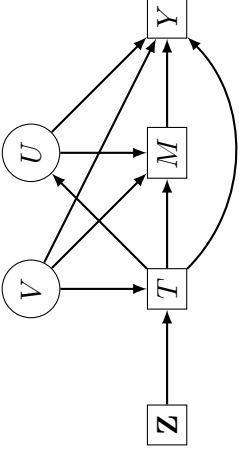
The bottom-panel of table 8 reports the results of applying our mediation framework to the other four outcomes. For brevity, we only report on the second-stage results, extending the results reported in the middle panel of table 8 for (i) total employment. Both (ii) manufacturing's employment share and (iii) manufacturing wages are estimated to have significant causal effects on voting, with p-values in columns 2–5 ranging between 0.028 and 0.113 for (ii) and between 0.042 and 0.099 for (iii). In the case of (ii) manufacturing's employment share the effect is also large: The estimated mediated effect $\hat{\Gamma}_M^{Y|T} \times \hat{\Gamma}_T^M = 0.146$ is practically identical to the 0.149 reported for (i) total employment in table 8.

A. DAG Representation

Restricted Model



General Model



B. Structural Equations

$$\begin{aligned}
 T &= f_T(\mathbf{Z}, V_T, \epsilon_T) \\
 U &= f_U(T, \epsilon_U) \\
 M &= f_M(T, U, V_T, \epsilon_M) \\
 Y &= f_Y(T, M, U, V_Y, \epsilon_Y) \\
 \mathbf{Z} &\perp\!\!\!\perp V_T \perp\!\!\!\perp V_Y \perp\!\!\!\perp \epsilon_Y \perp\!\!\!\perp \epsilon_M \perp\!\!\!\perp \epsilon_U \perp\!\!\!\perp \epsilon_T
 \end{aligned}$$

C. Linear Equations

$$\begin{aligned}
 T &= \boldsymbol{\xi}_Z \cdot \mathbf{Z} + \xi_V \cdot V_T + \epsilon_T \\
 M &= \tilde{\varphi}_T \cdot T + \varphi_U \cdot \tilde{U} + \delta_Y \cdot V_Y + \epsilon_M \\
 Y &= \tilde{\beta}_T \cdot T + \beta_M \cdot M + \beta_U \cdot \tilde{U} + \beta_V \cdot V_Y + \epsilon_Y
 \end{aligned}$$

D. Identifying Equations

$$\begin{aligned}
 \text{cov}(\mathbf{Z}, T) - \text{cov}(\mathbf{Z}, \mathbf{Z}) \cdot \boldsymbol{\xi}_Z &= 0 \\
 \text{cov}(\mathbf{Z}, M) - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\varphi}_T &= 0 \\
 \text{cov}(\mathbf{Z}, Y) - \text{cov}(\mathbf{Z}, M) \cdot \beta_M - \text{cov}(\mathbf{Z}, T) \cdot \tilde{\beta}_T &= 0 \\
 \text{cov}(T, Y) - \text{cov}(T, M) \cdot \beta_M - \text{cov}(T, T) \cdot \tilde{\beta}_T &= 0 \\
 \text{cov}(M, Y) - \text{cov}(M, M) \cdot \beta_M - \text{cov}(T, M) \cdot \tilde{\beta}_T &= \beta_U \cdot \varphi_U + \beta_V \cdot \delta_Y \\
 \text{cov}(T, M) - \text{cov}(T, T) \cdot \tilde{\varphi}_T &= \delta_T \cdot \xi_V
 \end{aligned}$$

Panel A presents the Directed Acyclic Graphs (DAG) of the restricted model examined in the paper and the General Mediation model that does not assume the restriction on the causal restriction on confounding variables. Panel B presents the structural equations associated with each model. Panel C presents the linear equations that subsume the causal relations described in each model. Panel D displays the equalities generated by the covariance structure arising from the linear equations.

Online Appendix Table 8: Additional Observed Mechanisms

		(1) IV	(2) IV	(3) IV	(4) IV	(5) IV
<i>Second Stage: Effect of T on M</i>						
$\widehat{\Gamma}_T^M$	M: Δ Share Manufacturing Employ.	-0.440 [0.048]	-0.618 [0.002]	-0.738 [0.000]	-0.745 [0.000]	-0.755 [0.000]
	M: $\Delta \log(\text{Mean Manufact. Wage})$	-0.006 [0.013]	-0.005 [0.032]	-0.006 [0.014]	-0.005 [0.012]	-0.006 [0.010]
	M: $\Delta \log(\text{Mean Non-Manuf. Wage})$	-0.005 [0.004]	-0.002 [0.096]	-0.002 [0.304]	-0.001 [0.433]	-0.001 [0.419]
	M: Δ Share Unemployment	0.076 [0.271]	0.097 [0.124]	0.076 [0.359]	0.084 [0.302]	0.110 [0.090]
<i>Second Stage: M on Y, conditional on T</i>						
$\widehat{\Gamma}_M^{\widehat{Y} T}$	M: Δ Share Manufacturing Employ.	-0.070 [0.447]	-0.202 [0.113]	-0.214 [0.058]	-0.198 [0.028]	-0.193 [0.029]
$\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{\widehat{Y} T}$		0.031 [0.478]	0.125 [0.159]	0.158 [0.094]	0.148 [0.059]	0.146 [0.060]
$\widehat{\Gamma}_M^{\widehat{Y} T}$	M: $\Delta \log(\text{Mean Manufacturing Wage})$	-6.17 [0.054]	-6.24 [0.066]	-6.72 [0.042]	-4.12 [0.081]	-3.80 [0.099]
$\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{\widehat{Y} T}$		0.035 [0.127]	0.028 [0.163]	0.038 [0.117]	0.022 [0.153]	0.022 [0.164]
$\widehat{\Gamma}_M^{\widehat{Y} T}$	M: $\Delta \log(\text{Mean Non-Manuf. Wage})$	-20 [0.035]	-45 [0.322]	-102 [0.497]	-103 [0.449]	-101 [0.454]
$\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{\widehat{Y} T}$		0.095 [0.090]	0.108 [0.395]	0.167 [0.571]	0.128 [0.586]	0.132 [0.583]
$\widehat{\Gamma}_M^{\widehat{Y} T}$	M: Δ Share Unemployment	0.341 [0.541]	0.615 [0.328]	0.450 [0.334]	0.148 [0.632]	0.133 [0.744]
$\widehat{\Gamma}_T^M \times \widehat{\Gamma}_M^{\widehat{Y} T}$		0.026 [0.593]	0.060 [0.409]	0.034 [0.506]	0.012 [0.664]	0.015 [0.749]

Notes: (a) Columns 1–5 introduce controls in the same way as table 2. *p-values* are reported in square brackets, standard errors are clustered at the level of 96 commuting zones. (b) The top panel reports the results of estimating (5) for four other labor market outcomes. The first two outcomes (manufacturing’s employment share and manufacturing wages) are significantly negatively impacted by import competition. Results for the other two outcomes (non-manufacturing wages and unemployment) are much weaker. (c) The bottom panel reports the results of estimating (47) for these four labor market outcomes. Both manufacturing’s employment share and manufacturing wages are estimated to have significant causal effects on voting, with *p-values* in columns 2–5 ranging from 0.028–0.113 for manufacturing’s employment share and from 0.042–0.099 for manufacturing wages.

Online Appendix Table 9: Principal Component Analysis

			(1)	(2)	(3)	(4)	(5)
			Factor-Loadings (Eigen-Vectors)				
	Eigen-value	Eigen-value: Proportion	Δ Log(Total Empl.)	Δ Share Manuf. Empl.	$\Delta \log(\text{Avg.}$ Manuf. Wage)	$\Delta \log(\text{Avg.}$ Non- Manuf. Wage)	Δ Share Unempl
LMC ₁	2.707	0.541	0.1711	-0.3632	0.5108	0.5486	0.5261
LMC ₂	1.281	0.256	0.7625	0.6004	0.2104	0.0607	-0.1012
LMC ₃	0.509	0.102	-0.5389	0.397	0.5311	0.3251	-0.4053
LMC ₄	0.289	0.058	-0.3075	0.5916	-0.2521	0.0517	0.6994
LMC ₅	0.215	0.043	0.0664	0.0042	-0.5908	0.7661	-0.2439

Notes: The first column shows the eigenvalues of the five principal components (LMCs). The second column shows the share of total data variation they explain. Together, the first two LMCs explain almost 80 percent of the variation in the data (0.541 + 0.256). Columns 1–5 report on each LMC’s eigen-vector, or ‘factor loadings’. LMC₁ is primarily associated with changes in wages and unemployment. LMC₂ is strongly associated with changes in total employment and in manufacturing’s share of employment.