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 June 2018

We thank Alberto Alesina, David Autor, Sascha Becker, Sonia Bhalotra, Gilles Duranton, Jon Eguia, Johanna Fajardo, Andreas Ferrara, Markus Frölich, Paola Giuliano, James Heckman, Martin Huber, Kosuke Imai, Ed Leamer, Yi Lu, Craig McIntosh, Ralph Ossa, Sebastian Ottinger, Anne Otto, David Pacini, Bruno Pellegrino, Maria Petrova, Justin Pierce, David Slichter, Daniel Sturm, Peter Schott, Zhigang Tao, Dustin Tingley, Dan Trefler, 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, UCSD, U HK, Warwick, the 2015 Quebec Political Economy conference, the 2016 EEA conference, the 2017 VfS conference, and the 2017 INET conference for valuable comments and discussions. We also thank David Slichter for thoughtful comments and 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.

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 June 2018
JEL No. F1,F6,J2

ABSTRACT

Instrumental variables (IV) are a common means to identify treatment effects. But standard IV methods do not allow us to unpack the complex treatment effects that arise when a treatment and its outcome together cause a second outcome of interest. For example, IV methods have been used to show that import exposure to low-wage countries has adversely affected Western labor markets. Similarly, they have been used to show that import exposure has increased voter polarization. However, standard IV cannot estimate to what extent the latter is a consequence of the former. This paper proposes a new identification framework that allows us to do so, appealing to one additional identifying assumption without requiring additional instruments. The added identifying assumption can be relaxed, and bounds instead of point estimates can be derived. Applying this framework, we estimate that labor market adjustments explain most to all of the effect of import exposure on voting, thereby providing rigorous evidence that the correct policy response to voter polarization has to be focused on labor markets.

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

Instrumental variables (IV) are broadly used to identify the causal effect of a treatment variable on an outcome in observational data. Standard IV methods, however, are unable to unpack the causal chain that arises when the treatment and its outcome jointly cause a second outcome of interest. Our empirical application at the nexus of import competition, labor markets and voting is a case in point: International trade between high and low-wage countries has risen dramatically in the last thirty years (Krugman, 2008). Consumers in high-wage countries have benefited from such import exposure through cheaper manufactured goods. However, IV methods have been used to show that import exposure has also caused real wage stagnation and substantial job losses in Western manufacturing (Autor, Dorn, and Hanson, 2013; Dauth, Findeisen, and Suedekum, 2014; Malgouyres, 2017).[1] The same IV methods have been used to show that import exposure has increased the support for parties and politicians with protectionist, populist, and nationalist agendas (Malgouyres, 2014; Dippel, Gold, and Heblich, 2015; Feigenbaum and Hall, 2015; Autor, Dorn, Hanson, and Majlesi, 2016). While these two findings seem to be related, standard IV methods cannot quantify to what extent import exposure has turned voters towards political populism because it adversely affected labor markets. Unpacking this causal mechanism is important to guide policy makers in designing effective responses to political populism.

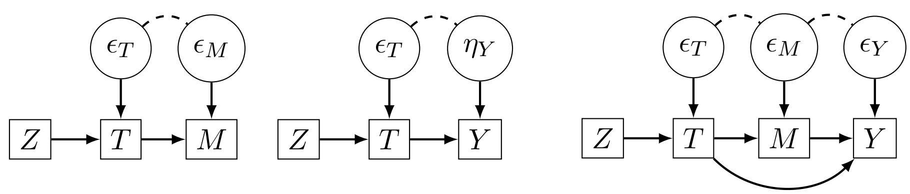
From an econometric perspective, we investigate the problem of identifying causal relations when an endogenous treatment and its outcome together cause a second outcome of interest. We propose a solution to the problem that does not require additional instrumental variables and can be easily implemented using the well-known Two-Stage Least Square (2SLS) estimator. We begin by clarifying the identification challenge. The starting point is to estimate the effect of a non-random treatment T (e.g. import exposure) on an outcome M (e.g. regional labor market adjustments). The Ordinary Least Squares (OLS) estimate of said treatment effect may be biased by omitted variables that affect both T and M (e.g. regional demand shocks may reduce regional imports as well as employment). The solution involves using an instrumental variable Z that affects T (i.e. there is a first stage relation) but is uncorrelated with the omitted variables (i.e. the

[1] Not every paper that addresses this question uses IV. See for example Dix-Carneiro 2014 and Pierce and Schott 2016, who instead focus on exogenous variation in tariff reductions.

Table 1: The Identification Problem of Mediation Analysis with IV

A. Graphical Representation

Model I: IV for Labor M	*Model II: IV for Voting Y*	*Model III: IV for the Mediation Model*



B. Model Equations

$T = f_T(Z, \epsilon_T)$	$T = f_T(Z, \epsilon_T)$	$T = f_T(Z, \epsilon_T),\ M = f_M(T, \epsilon_M)$
$M = f_M(T, \epsilon_M)$	$Y = g_Y(T, \eta_Y)$	$Y = f_Y(T, M, \epsilon_Y)$
$Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M)$	$Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$	$Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$

Notes: (*a*) *Model I* is the standard IV model, which enables the identification of the causal effect of T on M. *Model II* is the standard IV model that enables the identification of the causal effects of T on Y. *Model III* is the IV Mediation Model with an instrumental variable Z. (*b*) Panel A gives the graphical representation of the models. Panel B presents the nonparametric structural equations of each model. Conditioning variables are suppressed for sake of notational simplicity. We use $\perp\!\!\!\perp$ to denote statistical independence.

exclusion restriction holds).[2] This is the standard IV solution and is depicted in *Model I* in Table 1. T is endogenous in a regression of M on T (i.e. $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$), but Z is exogenous (i.e. $Z \perp\!\!\!\perp \epsilon_T, \epsilon_M$).

We are interested in the identification challenge that arises when there is a second outcome of interest Y (e.g. voting) that is likely to be caused by T both through M as well as 'directly'. The most straightforward approach to this is to simply estimate the 'total effect' of T on Y using the same IV approach, as depicted in *Model II* in Table 1: $\epsilon_T \not\perp\!\!\!\perp \eta_Y$, but Z is exogenous (i.e. $Z \perp\!\!\!\perp \epsilon_T, \eta_Y$).[3] In combination, *Model I* and *Model II* estimate the causal effect of T on M and the causal effect of T on Y. However, this leaves unidentified whether and to what extent the former causes the latter. In our empirical setting, there clearly could be other channels that directly link T to Y: On the one hand, if import exposure creates anxiety about the future this may by itself turn voters towards populism (Mughan and Lacy, 2002; Mughan, Bean, and McAllister, 2003).[4] On the

[2] We base our analysis on the well-known instrument proposed by Autor et al. (2013).

[3] It is common to use the same instrument to identify the causal effect of a treatment on several outcomes, and the application studied here is no different. For example, three pairs of papers in the related literature 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, Lu, Pierce, Schott, and Tao (2016).

[4] In our empirical specification, import exposure refers to the net import exposure, i.e. the exposure to imports *after* accounting for export exposure.

other hand, import exposure may be politically moderating if it lowers consumer prices, if targeted government transfers like *Trade Adjustment Assistance* increase, if manufacturers shift production towards more differentiated higher mark-up output varieties (as in Holmes and Stevens, 2014), or if it leads to task-upgrading within industries and occupations (as in Becker and Muendler, 2015). Depending on these factors' relative importance, their aggregate effect on support for populists may be positive or negative. If these direct effects as a whole are negative, the effect of import exposure on voting that is mediated by labor market adjustments could actually be larger than the total effect estimated by *Model II*. We find evidence that this is indeed the case.

The identification challenges that arise from this discussion are depicted in *Model III* in Table 1. T causes Y indirectly through M as well as directly, i.e. through a number of other channels. In a regression of Y on both T and M there are two endogenous regressors (i.e. $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y$, $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$), but there is only one instrument Z to address this endogeneity. *Model III* is a *mediation model*, i.e. one where T (import exposure) causes an intermediate outcome M (labor market adjustments) that is also a 'mediator' in T's effect on a final outcome Y (voting).[5] Most of the approaches to identification in mediation analysis assume that T is as good as randomly assigned (i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_M$), making them not applicable to the IV settings we are interested in. See, e.g., Imai, Keele, Tingley, and Yamamoto 2011a. The only existing approaches to achieving identification in the IV setting of *Model III* require separate dedicated instruments for M, which require additional exogeneity assumptions that are considerably more restrictive than the standard ones (e.g. Jun, Pinkse, Xu, and Yildiz 2016; Frolich and Huber 2017).

Our proposed solution does not assume away endogeneity in any of the key relationships in *Model III* and does not require additional instruments. Instead, we rely on the insight that in many research settings the omitted variable concerns themselves suggest a natural solution. This is the case when T is endogenous in a regression of Y on T primarily because of omitted variables that affect M. For example, in the literature above the main endogeneity concern in a regression of regional manufacturing employment (M) on import exposure (T) is that unobserved adverse regional demand shocks reduce regional imports as well as employment, and it seems likely that such shocks affect voting (Y) primarily to the extent that they affect labor markets. We show

[5] Mediation analysis decomposes the *total effect* of T on Y into the *indirect effect* of T on Y that operates through M and the *direct effect* that does not. The indirect effect may alternatively be labeled as the '*mediated effect*'. For recent works on this literature, see Heckman and Pinto (2015b); Pearl (2014); Imai, Keele, and Tingley (2010).

that this assumption alone is sufficient to unpack the causal channels in *Model III*, allowing us to identify the extent to which T causes Y through M. We further show that under linearity, the resulting identification framework is straightforwardly estimated using three separate 2SLS estimations of the effect of T on M, the effect of T on Y, and and the effect of M on Y conditional on T.[6]

While the identifying assumption of our framework is plausible in our setting, it may be less so in other mediation-type IV settings; and even when it is plausible, it is desirable to know how robust the results are to relaxing it. When we allow unobserved confounders that directly affect T and Y, covariance relations in the data still allow us to provide bounds on the possible range of estimates of the direct and the indirect effects linking T and Y.[7]

We apply our method to data on regional import exposure, labor market adjustments and voting patterns in Germany from 1987–2009. 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 in part because it offers several advantages, especially relative to the U.S., for the question at hand: (*i*) 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*) 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*) 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 socio-economic surveys, the

[6] Our main focus is to estimate the product of the effect T on M and the effect M on Y, i.e. the indirect 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

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

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

[7] In spirit, this relates to the method developed in Conley, Hansen, and Rossi (2012) for calculating plausible bounds on IV estimates, when the IV's exclusion restriction may not be fully satisfied.

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

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.

We combine changes in national sector-specific trade flows with regional labor markets' initial industry mix to determine regional import exposure (T). We then instrument T with a measure based on other high-wage countries' sector-specific trade flows (Z).[9] Estimating *Model I* in Table 1, we corroborate existing results that import exposure significantly reduces total employment (M), particularly in manufacturing, raises unemployment and negatively affects manufacturing wages. With a view towards the mediation analysis that follows, we aggregate these into one index using principal component (PC) analysis.[10] This approach is appealing for mediation settings if the treatment effects are concentrated in one PC, and this PC also has a clear interpretation. This turns out to be the case in the German labor market data, where one PC summarizes the effect of import competition on labor markets.

Estimating *Model II* in Table 1, we find that import exposure (T) increased voter polarization (Y). There is a significant positive effect on the vote share of the nationalist and highly protectionist extreme right.[11] There is no significant effects on turnout, or any of the mainstream parties, small parties, or the far left.[12] These findings are corroborated by the SOEP's individual-level data, where we can 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 import exposure. Using gravity residuals instead of our IV strategy yields similar results.[13]

Next, we estimate the causal links in *Model III*. A single PC, associated with total employment

[9]In 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 3.4 for more details.

[10] Our framework estimates the mediating effect of a single variable M, but in many research settings there will be several observed variables that potentially link a treatment T to an outcome Y. principal component analysis is attractive in this case because it generates orthogonal indices that are purely statistical.

[11] Election 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).

[12] Of course, *all other parties combined* must have a vote share loss of equal magnitude and significance to the extreme right's gain.

[13] We report these for completeness as this is standard in the literature on import exposure and labor markets (Autor et al., 2013; Dauth et al., 2014). However, our focus is naturally on the IV setting to which our identification framework applies.

and manufacturing, drives the effect of import competition on extreme-right voting. The effect of import-exposure-driven adjustments in this PC on voting is larger than the total effect of import exposure on voting, implying that other channels that connect import exposure to voting are moderating in the aggregate. When we relax the identifying assumptions within the possible range implied by the covariance relations in the data, we derive bounds on the indirect effect which suggests it explains between 70 and 128 percent of the total effect. Even at the lower bound, labor market adjustments are thus the primary reason for the populist backlash against import exposure.

Our paper's contribution is two-fold: It answers a relevant substantive question at the nexus of the literatures on trade, local labor markets and politics; in order to do so it makes a methodological contribution to the literature on causal mechanisms and on IV. On the substantive side, our 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.[14] 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 useful for the typical observational data setting where good instrumental variables are scarce. Our model can be estimated by well-known 2SLS methods, its identifying assumption can be relaxed to derive bounds instead of point estimates, and it can be applied to a potentially broad range of empirical research questions in which an endogenous treatment and its primary outcome together cause a second outcome of interest.

Section 2 explains our identification approach.Section 3 describes the data. Section 4 presents the IV results for *Model I* and *Model II*, establishing the causal effects of import exposure on labor markets and voting. Section 5 applies *Model III*. Section 6 concludes.

[14]Our findings relate more broadly 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).

2 Examining the Mediation Model with an Instrumental Variable

Our goal is to evaluate a sequence of causal relations where import exposure T causes labor market adjustments M, and both T and M cause changes in voting behavior Y. Such a sequence of causal relations is called a mediation model (Pearl, 2011). We modify the standard mediation model by adding an instrument Z that causes T. A general nonparametric model that portrays these causal relations is given by:

$$T = f_T(Z, \epsilon_T) \quad (1)$$

$$M = f_M(T, \epsilon_M) \quad (2)$$

$$Y = f_Y(T, M, \epsilon_Y) \quad (3)$$

We use $\epsilon_T, \epsilon_M, \epsilon_Y$ for the unobserved error terms associated with variables T, M, Y respectively. We use $supp(Z)$, $supp(T)$, $supp(M)$, $supp(Y)$ for the support of variables Z, T, M, Y respectively. We are interested in estimating causal effects among variables T, M, and Y. The seminal work of Robins and Greenland (1992) examines the mediation model in which a binary treatment T causes Y via M. They decompose the causal effect of T on Y into the *total*, the *direct* and the *indirect* effects. The total effect TE stands for the average causal effect of T on Y. The direct effect DE stands for the causal effect of T on Y that is not generated by changes in M. The indirect effect IE is the causal effect of T on Y induced by the change in the distribution of the mediator M.

A causal effect is defined by the difference between potential (counterfactual) variables. For instance, let $M(t)$ be the counterfactual variable M when T takes the value $t \in supp(T)$. The average causal effect of T on M when T takes values $t, t' \in supp(T)$ is given by the expected value of the difference $E(M(t) - M(t'))$.[15] Model (1)–(3) yields four counterfactual variables defined by the following equations:

[15] Formally, a causal model is defined by a set of structural equations that define the causal direction among the model. Counterfactual variables are defined by *fixing* an argument of a structural equation to a value. For instance let the structural equation that governs the data generating process of variable M be $M = f_M(T, \epsilon_M)$ where ϵ_M denotes an error term. The the counterfactual variable $M(t)$ is defined by *fixing* the T-input of this structural equation to the value $t \in supp(T)$, namely, $M(t) = f_M(t, \epsilon_M)$. See Heckman and Pinto (2015a) for a discussion on causality and the fixing operator.

$$M(t) = f_M(t, \epsilon_M); t \in supp(T) \tag{4}$$
$$Y(t) = f_Y(t, M(t), \epsilon_Y); t \in supp(T) \tag{5}$$
$$Y(m) = f_Y(T, m, \epsilon_Y); m \in supp(M) \tag{6}$$
$$Y(m,t) = f_Y(t, m, \epsilon_Y); t \in supp(T) \text{ and } m \in supp(M) \tag{7}$$

It is useful to examine a simple Randomized Control Trial (RCT) to show how the total, direct and indirect effect can be written using counterfactual variables. Counterfactual variable $Y(t)$ in (5) denotes the potential outcome Y when T takes the value $t \in supp(T)$, $Y(m)$ is the counterfactual outcome when M is fixed to the value $m \in supp(M)$, and $Y(m,t)$ is the counterfactual outcome that arises when both T, M are fixed to t, m respectively. Consider an RCT whose treatment assignment takes the values in $supp(T) = \{t_0, t_1\}$, where t_0 indicates the control group and t_1 indicates the treatment group. Let $F_{M(t)}(m)$ denote the Cumulative Density Function (CDF) of the counterfactual mediator $M(t)$ conditional on the assignment $t \in \{t_0, t_1\}$. The total effect TE is the expected difference between counterfactual outcome Y when T is fixed at t_1 and t_0. The direct effect $DE(t)$ evaluates the expected difference of counterfactual outcomes between treated (t_1) and control (t_0) group holding the distribution of the mediator fixed at $M(t)$. The indirect effect $IE(t)$ evaluates the expected value of the the difference between counterfactual outcomes $Y(t,m)$ when the distribution of the mediator m varies between treated $M(t_1)$ and control $M(t_0)$ while holding the t-input fixed. Notationally, these effects are defined as:

$$\begin{array}{lclcl}
TE & = & E(Y(t_1) - Y(t_0)) & \equiv & E(Y(t_1, M(t_1)) - Y(t_0, M(t_0))) \\
DE(t) & = & E(Y(t_1, M(t)) - Y(t_0, M(t))) & \equiv & \int E\Big(Y(t_1, m) - Y(t_0, m)\Big) dF_{M(t)}(m) \\
IE(t) & = & E(Y(t, M(t_1)) - Y(t, M(t_0))) & \equiv & \int E\Big(Y(t,m)\Big)\Big[dF_{M(t_1)}(m) - dF_{M(t_0)}(m)\Big]
\end{array}$$

Robins and Greenland's (1992) main contribution is to show that the total effect of T on Y can be decomposed as the sum of the effect of T on Y that is mediated by M (the indirect effect) and the causal effect of T on Y that is not mediated by M (the direct effect).[16] Equations (8)–(9) express the total effect as the sum of direct and indirect effects:[17]

[16] Pearl (2011) makes a distinction between natural (or "descriptive") direct and indirect effects and controlled (or "prescriptive") direct effects.

[17] A large literature on mediation analysis relies on the Sequential Ignorability Assumption **A-3** of Imai, Keele, and Yamamoto (2010) to identify mediation effects. This assumption is discussed in Online Appendix A. See Frolich and Huber (2017) for a recent review of the mediation literature.

$$
\begin{aligned}
TE &= E(Y(t_1, M(t_1)) - Y_i(t_0, M(t_0))) \\
&= \Big(E(Y(t_1, M(t_1))) - E(Y(t_0, M(t_1)))\Big) + \Big(E(Y(t_0, M(t_1)) - Y_i(t_0, M(t_0)))\Big) = DE(t_1) + IE(t_0) \quad (8) \\
&= \Big(E(Y(t_1, M(t_1))) - E(Y(t_1, M(t_0)))\Big) + \Big(E(Y(t_1, M(t_0)) - Y_i(t_0, M(t_0)))\Big) = IE(t_1) + DE(t_0). \quad (9)
\end{aligned}
$$

Robins and Greenland's (1992) decomposition can be extended to the problem we examine by allowing T to be a continuous variable, in which case the decomposition is obtained by the total differentiation of the counterfactual outcome:

$$
\underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}} = \underbrace{\frac{\partial E(Y(t,m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t,m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}}. \tag{10}
$$

Identification of the total, direct and indirect effects hinges on the dependence relation among the error terms $\epsilon_T, \epsilon_M, \epsilon_Y$ in (1)–(3). Suppose that the error terms ϵ_T, ϵ_M are statistically independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_M$. This means that there are no unobserved variables that jointly cause T and M. In this case, T is a exogenous with respect to M, as in an RCT. It is easy to show that the independence conditions $M(t) \perp\!\!\!\perp T$ holds and the expected value of counterfactual variable $M(t)$ is identified by the conditional expectation $E(M(t)) = E(M|T = t)$. In addition, if error terms (ϵ_T, ϵ_M) and ϵ_Y were statistically independent, then independence conditions $\big(Y(t), Y(t,m)\big) \perp\!\!\!\perp T$ and $Y(t,m) \perp\!\!\!\perp M$ would also hold. This means that variables T, M are exogenous with respect to Y and the expected value of counterfactual variables $E(Y(t))$ and $E(Y(t,m))$ would be identified by conditional expectations of observed variables $E(Y(t)) = E(Y|T = t)$ and $E(Y(t,m)) = E(Y|T = t, M = m)$ respectively.

These assumption of the mutual independence among error terms ($\epsilon_T \perp\!\!\!\perp \epsilon_M$, $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, $\epsilon_M \perp\!\!\!\perp \epsilon_Y$) are very strong; in fact they are stronger than the random treatment assumption made in an RCT, which does not include $\epsilon_M \perp\!\!\!\perp \epsilon_Y$. They are therefore unlikely to hold in any observational data, and since our starting point is to derive a mediation model for IV settings, we naturally want to relax them. We now investigate the pairwise dependence relations between the error terms in turn:

(*i*) Our interest is in settings where IV is needed to identify the causal effect of a treatment T on an outcome M. Inherently, we therefore allow $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$. In our empirical application the literature's main endogeneity concern is that unobserved adverse regional demand shocks reduce regional imports (T) as well as employment (M).

(*ii*) Our interest is in settings where T and M jointly cause a second outcome of interest Y. We therefore want to allow $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$. At the same time, it is usually unappealing to assume $\epsilon_M \perp\!\!\!\perp \epsilon_Y$. In our empirical application unobserved industry or worker characteristics that affect labor market outcomes may also affect political preferences. This induces a correlation between error term ϵ_M, associated with labor market outcomes M, and error term ϵ_Y, associated with voting behavior.

(*iii*) When it comes to the relation between ϵ_T and ϵ_Y, we observe that in many research settings the specific omitted variable concerns themselves suggest a natural solution, as follows. In our empirical application for example, the main endogeneity concern in a regression of regional manufacturing employment (M) on import exposure (T) is that unobserved adverse regional demand shocks reduce regional imports as well as employment, and it is plausible that such shocks affect voting (Y) primarily to the extent that they affect labor markets. This implies that $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y$ only because of confounders that affect T and M.

The identification framework we propose uses this intuition. It allows for unobserved shocks ϵ_T to affect voter preferences through the labor markets M, but not through other channels. Notationally, $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, but $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$. This can be stated alternatively as $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | M$, namely, the unobserved variables ϵ_T, ϵ_Y that affect regional import exposure and voting are *not* statistically independent conditional on labor market adjustments M. This assumption does not assume away endogeneity in any of the key relations, but, as we show, suffices to yield point-identification of the total, direct and indirect effects.

It is important to be clear that we are not claiming that the assumption that T is endogenous in Y only because of unobserved confounders that impact T and M is appropriate in all mediation-type settings. There will be settings where this assumption is clearly not plausible and there will be settings where the plausibility of the assumption is debatable. It is therefore important that we can relax $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, which we do in Section 2.3.

Relaxing $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ still allows us to bound the decomposition of the total into direct and indirect effect. In addition, it allows us to estimate bounds for the correlation between error terms. In our empirical application, the estimates of the correlation between error terms (in Section 5) will corroborate our reasoning: While the bounds for the correlation between error terms ϵ_T, ϵ_Y contain the zero correlation, the estimated bounds for ϵ_T, ϵ_M and ϵ_M, ϵ_Y are strictly positive.[18]

The statistical dependence between error terms ϵ_T and ϵ_M precludes the independence between T and counterfactuals $M(t), Y(t)$ i.e. $T \not\perp\!\!\!\perp \big(Y(t), M(t)\big)$. Hence T is endogenous and the observed correlation between treatment T and the outcomes M, Y does not identify causal effects. As well, the correlation between error terms ϵ_M and ϵ_Y turns M endogenous because it invalidates the independence condition between M and counterfactuals $\big(Y(m,t), Y(m)\big)$, i.e. $M \not\perp\!\!\!\perp \big(Y(m), Y(m,t)\big)$. Hence the causal effect of M on Y cannot be identified on the basis of their observed distributions. In particular, the conditional expectation of observed variables $E(M|T=t)$, $E(Y|T=t)$, $E(Y|T=t, M=m)$ do not identify the expected value of the counterfactual outcomes $E(M(t))$, $E(Y(t))$, $E(Y(t,m))$. We utilize the properties of an instrumental variable Z to solve this identification problem.

The standard IV exclusion restriction is that instrument Z affects M and Y only through its impact on T. Otherwise stated, for Z to be an instrument, it must be the case that Z is statistically independent of unobserved error terms $\epsilon_T, \epsilon_M, \epsilon_Y$ which jointly cause T, M, Y. In the interest of clarity, we state this property as Assumption **A-1**.[19]

Assumption A-1 *The independence relation $Z \perp\!\!\!\perp \big(\epsilon_T, \epsilon_M, \epsilon_Y\big)$ holds in the mediation model (1)–(3).*

A-1 merely states the independence condition that characterizes Z as an instrumental variable for T. Lemma **L-1** explains that this independence condition generates the two exclusion restrictions.

Lemma L-1 *Under Assumption **A-1**, the following statistical relations hold:*

Targeted Causal Relation	*IV Relevance*		*Exclusion Restrictions*
$T \to Y$	$Z \not\perp\!\!\!\perp T$	*and*	$Z \perp\!\!\!\perp Y(t)$
$T \to M$	$Z \not\perp\!\!\!\perp T$	*and*	$Z \perp\!\!\!\perp M(t)$

Proof P-1 *See **P-1** in Appendix A.*

[18] While this does not amount to a statistical test of the identifying assumption, it is nonetheless reassuring.

[19] As mentioned, Autor et al. (2013) fostered a large literature that uses import exposure of other countries (say O's) as an instrument Z for one specific country's (say G's) import exposure to low-wage manufacturing countries. See expressions (46) and (47) for the precise definitions of T and Z in our empirical application.

The exclusion restrictions in **L-1** imply that the counterfactual outcomes $M(t)$ and $Y(t)$ can be evaluated using standard IV techniques. This is not surprising. **L-1** simply means that an instrument for T enables the identification of the causal effect of T on M as well as T on Y. **L-1** is most relevant due to its symmetry. The exclusion restriction that applies to the mediator M also applies to outcome Y. The fact that M causes Y (and not the opposite) plays no role in generating exclusion restrictions. Indeed, the lemma would remain the same if the causal relation $M \rightarrow Y$ were reversed to $M \leftarrow Y$. The irrelevance of the causal direction between M and Y exposes a limitation of the instrumental variable: although instrument Z for T identifies the causal effect of T on M, Y, it is not suitable to identify the causal effect of M on Y. This problem is addressed by evoking the following dependence structure among error terms:

Assumption A-2 *The following independence relations hold in the mediation model* (1)–(3)*:*

$$\epsilon_T \not\perp\!\!\!\perp \epsilon_M,\ \epsilon_M \not\perp\!\!\!\perp \epsilon_Y,\ \epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M \text{ and } \epsilon_T \perp\!\!\!\perp \epsilon_Y. \tag{11}$$

Assumption **A-2** allows for the correlation between error terms ϵ_T, ϵ_M, thus T is endogenous. **A-2** also allows for error terms ϵ_M, ϵ_Y, to correlate, thus M is endogenous. **A-2** states that error terms ϵ_T, ϵ_Y are unconditionally independent, but correlate conditional on ϵ_M.

Assumption **A-2** yields a new exclusion restriction stated in Lemma **L-2**. We show that this additional restriction renders our model just identified with one instrumental variable Z. Specifically, this restriction is used to identify the causal effect of M on Y and thereby decompose the total effect of T on Y into its direct and indirect counterparts.

Lemma L-2 *Under* ***A-1*** *–* ***A-2****, the following statistical relation hold:*

Targeted Causal Relation	*IV Relevance*		*Exclusion Restrictions*
for $M \rightarrow Y$	$Z \not\perp\!\!\!\perp M \mid T$	*and*	$Z \perp\!\!\!\perp Y(m) \mid T$

Proof P-2 *See* ***P-2*** *in Appendix A.*

The exclusion restriction of **L-2**, i.e. $Z \perp\!\!\!\perp Y(m)|T$, implies that the instrumental variable 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. Corollary **C-1** states that the counterfactual outcome $Y(m)$ conditioned on $T = t$, i.e. $(Y(m)|T = t)$, is equal in distribution to the counterfactual outcome $Y(m, t)$. Therefore Z can be used to identify $Y(m, t)$ using standard IV methods.

Corollary C-1 *Under* **A-1***–***A-2***, the counterfactual outcome* $Y(m)$ *conditioned on* $T = t$ *is equal in distribution to the counterfactual outcome* $Y(m,t)$*, i.e.,* $(Y(m)|T = t) \overset{d}{=} Y(m,t)$.

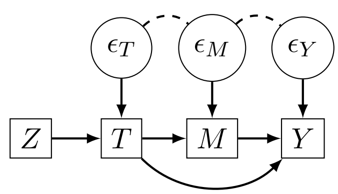
Proof P-3 *See* ***P-3*** *in Appendix A.*

As discussed, we relax Assumption **A-2** in Section 2.3 to derive bounds instead of point estimates.

Panel A of Table 2 represents the causal relations of the mediation model (1)–(3) as a Directed Acyclic Graph (DAG). Squares represent observed variables while circles denote unobserved variables. Causal relations are denoted by solid lines while the dependence structure among error variables is depicted by dashed lines. Table 2 is a version of *Model III* in Table 1, with the specific assumptions on the error terms that we just discussed.

Table 2: The Mediation Model with IV

A. DAG Representation



B. Model Equations

Treatment variable: $T = f_T(Z, \epsilon_T)$
Observed Mediator: $M = f_M(T, \epsilon_M)$
Outcome: $Y = f_Y(T, M, \epsilon_Y)$
where: $\epsilon_T \not\perp\!\!\!\perp \epsilon_M,\ \epsilon_M \not\perp\!\!\!\perp \epsilon_Y,\ \epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$
and: $Z \perp\!\!\!\perp \big(\epsilon_T, \epsilon_M, \epsilon_Y\big), \epsilon_T \perp\!\!\!\perp \epsilon_Y$

Remark 2.1 The exclusion restrictions in **L-1**–**L-2** also hold for a more general model that allows for an unobserved mediator U that is caused by T and causes both M and Y. Notationally, this model is characterized by the following equations: $T = f_T(Z, \epsilon_T)$, $U = f_U(T, \epsilon_U)$, $M = f_M(T, U, \epsilon_M)$, $Y = f_Y(T, M, U, \epsilon_Y)$. We investigate this model in Online Appendix B.

Remark 2.2 The independence condition in **A-2** enables the use of the IV to identify the counterfactual outcome $Y(m)$. The identification is possible by conditioning on the treatment T. Another identification approach could be formulated if additional instrumental variables were available.

Consider an additional variable $\tilde{Z}$ that plays the role of an instrumental variable that is exclusively dedicated to M. This means that variable $\tilde{Z}$ is characterized by two properties: (1) $\tilde{Z}$ does not cause T; and (2) $\tilde{Z}$ has no impact on Y other than through M. This instrument could be used to evaluate the causal effect of M on Y, however the availability of such instrument is unlikely in most empirical settings. See Online Appendix C for a discussion on this topic, where we discuss the potential applicability as well as pitfalls of using 'automation' as a candidate for $\tilde{Z}$ in our empirical application.

Remark 2.3 Exclusion restrictions, such as **L-1**–**L-2**, are necessary but not sufficient to identify causal effects. An extensive IV literature exists on the additional assumptions that grant the identification of causal effects.[20] Examples of these additional assumptions are monotonicity (Imbens and Angrist, 1994; Heckman and Pinto, 2017), separability of the choice equation Heckman and Vytlacil (2005) or control functions (Blundell and Powell, 2003, 2004), or revealed preference analysis (Pinto, 2015).

2.1 The IV Mediate Model under Linearity

We adopt the assumption of linearity, which is prevalent in many literatures, including the local labor markets literature, which our empirical application belongs to. Our model is thus characterized by the following equations:

$$Z = \epsilon_Z, \tag{12}$$
$$T = \beta_T^Z \cdot Z + \epsilon_T, \tag{13}$$
$$M = \beta_M^T \cdot T + \epsilon_M, \tag{14}$$
$$Y = \beta_Y^T \cdot T + \beta_Y^M \cdot M + \epsilon_Y, \tag{15}$$

where $\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y$, are error terms whose variances are denoted by $\sigma^2_{\epsilon_Z}, \sigma^2_{\epsilon_T}, \sigma^2_{\epsilon_M}, \sigma^2_{\epsilon_Y}$ respectively. Let ρ_{TM} stands for the correlation between ϵ_T, ϵ_M. Likewise, let ρ_{TY}, ρ_{MY} stand for the correlations between ϵ_T, ϵ_Y and ϵ_M, ϵ_Y respectively. For sake of notational simplicity, we assume that each variable has mean zero. This assumption does not incur a loss of generality. The direct effect is given by the coefficient $DE = \beta_Y^T$, the indirect effect is given by the coefficient multiplication $IE = \beta_M^T \cdot \beta_Y^M$, an the total effect is the sum of these two terms $TE = \beta_Y^T + \beta_M^T \cdot \beta_Y^M$.[21]

[20] See Dahl, Huber, and Mellace (2017) for a recent review.

[21] The linear model lacks treatment-mediator interactions and therefore has homogeneous effects (natural and controlled effects coincide). This is called the no-interaction assumption that the mediation and direct effects do not depend on the values of the Treatment T.

We are interested in evaluating the linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^T, \beta_Y^M$. The identification of these coefficients depends on the covariance matrix of observed data. Therefore it is useful to represent model (12)–(15) in matrix form. Let $\mathbf{X} = [Z, T, M, Y]'$ be the vector of observed random variables and $\epsilon = [\epsilon_Z, \epsilon_T, \epsilon_M, \epsilon_Y]'$ be the vector of unobserved error terms. Matrix $\mathbf{\Psi}$ in (16) stands for the arrangement of linear coefficients. Model (12)–(15) is then written as $\mathbf{X} = \mathbf{\Psi} \cdot \mathbf{X} + \boldsymbol{\epsilon}$ in (17):

$$\underbrace{\begin{bmatrix} Z \\ T \\ M \\ Y \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_T^Z & 0 & 0 & 0 \\ 0 & \beta_M^T & 0 & 0 \\ 0 & \beta_Y^T & \beta_Y^M & 0 \end{bmatrix}}_{\mathbf{\Psi}} \cdot \underbrace{\begin{bmatrix} Z \\ T \\ M \\ Y \end{bmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{bmatrix}}_{\boldsymbol{\epsilon}} \tag{16}$$

$$\therefore \mathbf{X} = \mathbf{\Psi} \cdot \mathbf{X} + \boldsymbol{\epsilon} \tag{17}$$

Equation (18) presents the covariance matrix $\Sigma_{\mathbf{X}}$ of observed variables $\mathbf{X}$:

$$\Sigma_{\mathbf{X}} \equiv \mathbf{Var}\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix} = \begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} & \sigma_{ZM} & \sigma_{ZY} \\ \cdot & \sigma_{TT} & \sigma_{TM} & \sigma_{TY} \\ \cdot & \cdot & \sigma_{MM} & \sigma_{MY} \\ \cdot & \cdot & \cdot & \sigma_{YY} \end{bmatrix}. \tag{18}$$

Assumption **A-1** states that Z is an instrumental variable. It implies that error term ϵ_Z is statistically independent of $\epsilon_T, \epsilon_M, \epsilon_Y$. Thus the covariance matrix Σ_ϵ of unobserved error terms $\boldsymbol{\epsilon}$ is given by:

$$\Sigma_\epsilon \equiv \mathbf{Var}\begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix} = \begin{bmatrix} \sigma^2_{\epsilon_Z} & 0 & 0 & 0 \\ \cdot & \sigma^2_{\epsilon_T} & \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} & \rho_{TY}\sigma_{\epsilon_T}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \sigma^2_{\epsilon_M} & \rho_{MY}\sigma_{\epsilon_M}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \cdot & \sigma^2_{\epsilon_Y} \end{bmatrix}. \tag{19}$$

The identification of linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^T, \beta_Y^M$ and the unobserved parameters in Σ_ϵ as defined by (19) is based on the equality between the covariance matrices of the observed and unobserved random variables, namely:

$$\mathbf{X} = \mathbf{\Psi} \cdot \mathbf{X} + \boldsymbol{\epsilon} \quad \Rightarrow \quad (\mathbf{I} - \mathbf{\Psi})\,\mathbf{X} = \boldsymbol{\epsilon} \quad \Rightarrow \quad (\mathbf{I} - \mathbf{\Psi})\,\Sigma_{\mathbf{X}}\,(\mathbf{I} - \mathbf{\Psi})' = \Sigma_{\boldsymbol{\epsilon}}. \tag{20}$$

Assumption **A-2** states that $\epsilon_T \perp\!\!\!\perp \epsilon_Y$. In the linear model (12)–(15), this is equivalent to assuming that the correlation of error terms ϵ_T, ϵ_Y is zero, i.e. $\rho_{TY} = 0$. Let $\widetilde{\Sigma}_{\boldsymbol{\epsilon}}$ be the covariance matrix (19) under the assumption that $\rho_{TY} = 0$. This covariance matrix is displayed in (22).

$$\widetilde{\Sigma}_{\mathbf{X}} \equiv \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_T^Z & 1 & 0 & 0 \\ 0 & -\beta_M^T & 1 & 0 \\ 0 & -\beta_Y^T & -\beta_Y^M & 1 \end{bmatrix}}_{\mathbf{I}-\boldsymbol{\Psi}} \cdot \underbrace{\begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} & \sigma_{ZM} & \sigma_{ZY} \\ \cdot & \sigma_{TT} & \sigma_{TM} & \sigma_{TY} \\ \cdot & \cdot & \sigma_{MM} & \sigma_{MY} \\ \cdot & \cdot & \cdot & \sigma_{YY} \end{bmatrix}}_{\Sigma_{\mathbf{X}}} \cdot \underbrace{\begin{bmatrix} 1 & -\beta_T^Z & 0 & 0 \\ 0 & 1 & -\beta_M^T & -\beta_Y^T \\ 0 & 0 & 1 & -\beta_Y^M \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{(\mathbf{I}-\boldsymbol{\Psi})'} = \tag{21}$$

$$= \underbrace{\begin{bmatrix} \sigma_{\epsilon_Z}^2 & 0 & 0 & 0 \\ \cdot & \sigma_{\epsilon_T}^2 & \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} & 0 \\ \cdot & \cdot & \sigma_{\epsilon_M}^2 & \rho_{MY}\sigma_{\epsilon_M}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \cdot & \sigma_{\epsilon_Y}^2 \end{bmatrix}}_{\Sigma_{\boldsymbol{\epsilon}} \text{ under } \rho_{TY}=0} \equiv \widetilde{\Sigma}_{\boldsymbol{\epsilon}}. \tag{22}$$

The equality (21)–(22) compares two covariance matrices of dimension four. Each matrix has $4 \cdot 4 = 16$ elements. We use $\widetilde{\Sigma}_{\boldsymbol{\epsilon}}[i,j]$ to denote the element in the i-th row and j-th column of matrix $\widetilde{\Sigma}_{\boldsymbol{\epsilon}}$. For sake of notational simplicity, we define $\widetilde{\Sigma}_{\mathbf{X}} \equiv (\mathbf{I}-\boldsymbol{\Psi})\,\Sigma_{\mathbf{X}}\,(\mathbf{I}-\boldsymbol{\Psi})'$ where $\widetilde{\Sigma}_{\mathbf{X}}[i,j]$ denotes the the element in the i-th row and j-th column of the matrix $(\mathbf{I}-\boldsymbol{\Psi})\,\Sigma_{\mathbf{X}}\,(\mathbf{I}-\boldsymbol{\Psi})'$. The matrix is symmetric, thus the equality generates ten equations: four diagonal equations and six off-diagonal ones. Notationally, these ten equalities are defined by $\widetilde{\Sigma}_{\mathbf{X}}[i,j] = \widetilde{\Sigma}_{\boldsymbol{\epsilon}}[i,j]$ for $i \le j; i,j \in \{1,2,3,4\}$. Four out of the six off-diagonal equation are equal to zero, namely, $\widetilde{\Sigma}_{\mathbf{X}}[1,j] = 0$ for $j \in \{2,3,4\}$ and $\widetilde{\Sigma}_{\mathbf{X}}[2,4] = 0$. What follows are the equations associated with the four zero elements in the covariance matrix (22):

$$\widetilde{\Sigma}_{\mathbf{X}}[1,2] = 0 \Rightarrow \quad \sigma_{ZT} - \beta_T^Z\sigma_{ZZ} = 0 \qquad \Rightarrow \quad \beta_T^Z = \tfrac{\sigma_{ZT}}{\sigma_{ZZ}} \tag{23}$$

$$\widetilde{\Sigma}_{\mathbf{X}}[1,3] = 0 \Rightarrow \quad \sigma_{ZM} - \beta_M^T\sigma_{ZT} = 0 \qquad \Rightarrow \quad \beta_M^T = \tfrac{\sigma_{ZM}}{\sigma_{ZT}} \tag{24}$$

$$\left.\begin{matrix} \widetilde{\Sigma}_{\mathbf{X}}[1,4] = 0 \\ \widetilde{\Sigma}_{\mathbf{X}}[2,4] = 0 \end{matrix}\right\} \Rightarrow \quad \left.\begin{matrix} \sigma_{ZY} - \beta_Y^M\sigma_{ZM} - \beta_Y^T\sigma_{ZT} = 0 \\ \sigma_{TY} - \beta_Y^M\sigma_{TM} - \beta_Y^T\sigma_{TT} = 0 \end{matrix}\right\} \quad \Rightarrow \quad \left\{\begin{matrix} \beta_Y^M = \tfrac{\sigma_{ZT}\sigma_{TY} - \sigma_{TT}\sigma_{ZY}}{\sigma_{ZT}\sigma_{TM} - \sigma_{TT}\sigma_{ZM}} \\ \beta_Y^T = -\tfrac{\sigma_{ZM}\sigma_{TY} - \sigma_{TM}\sigma_{ZY}}{\sigma_{ZT}\sigma_{TM} - \sigma_{TT}\sigma_{ZM}} \end{matrix}\right. \tag{25}$$

The four equalities in (23)–(25) suffice to identify all linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ of model (12)–(15). There are six remaining equalities in (21)–(22). The four diagonal equations generated upon (21)–(22) identify the variances of the error terms. Those equations are listed in (26).The left-hand side of each equation consists of observed covariances or identified parameters. The right-hand side of each equation consist of the error variances.

$$
\begin{aligned}
\widetilde{\Sigma}_{\mathbf{X}}[1,1] = \widetilde{\Sigma}_{\mathbf{e}}[1,1] \quad &\Rightarrow & \sigma_{ZZ} &= \sigma^2_{\epsilon_Z} \\
\widetilde{\Sigma}_{\mathbf{X}}[2,2] = \widetilde{\Sigma}_{\mathbf{e}}[2,2] \quad &\Rightarrow & \left(\sigma_{TT} - \beta_T^Z\sigma_{ZT}\right) - \beta_T^Z\left(\sigma_{ZT} - \beta_T^Z\sigma_{ZZ}\right) &= \sigma^2_{\epsilon_T} \\
\widetilde{\Sigma}_{\mathbf{X}}[3,3] = \widetilde{\Sigma}_{\mathbf{e}}[3,3] \quad &\Rightarrow & \left(\sigma_{MM} - \beta_M^T\sigma_{TM}\right) - \beta_M^T\left(\sigma_{TM} - \beta_M^T\sigma_{TT}\right) &= \sigma^2_{\epsilon_M} \\
\widetilde{\Sigma}_{\mathbf{X}}[4,4] = \widetilde{\Sigma}_{\mathbf{e}}[4,4] \quad &\Rightarrow & \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix}' \begin{bmatrix} \sigma_{YY} & \sigma_{MY} & \sigma_{TY} \\ \sigma_{MY} & \sigma_{MM} & \sigma_{TM} \\ \sigma_{TY} & \sigma_{TM} & \sigma_{TT} \end{bmatrix} \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix} &= \sigma^2_{\epsilon_Y}
\end{aligned}
\tag{26}
$$

The last two equalities can be extracted from (21)–(22) to identify the correlations ρ_{TM}, ρ_{MY}. Those are described in (27).

$$
\begin{aligned}
\widetilde{\Sigma}_{\mathbf{X}}[2,3] = \widetilde{\Sigma}_{\mathbf{e}}[2,3] \quad &\Rightarrow & \sigma_{TM} - \beta_T^Z\sigma_{ZM} - \beta_M^T(\sigma_{TT} - \beta_T^Z\sigma_{ZT}) &= \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} \\
\widetilde{\Sigma}_{\mathbf{X}}[3,4] = \widetilde{\Sigma}_{\mathbf{e}}[3,4] \quad &\Rightarrow & \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix}' \begin{bmatrix} \sigma_{MY} & \sigma_{TY} \\ \sigma_{MM} & \sigma_{TM} \\ \sigma_{TM} & \sigma_{TT} \end{bmatrix} \begin{pmatrix} 1 \\ -\beta_M^T \end{pmatrix} &= \rho_{MY}\sigma_{\epsilon_M}\sigma_{\epsilon_Y}
\end{aligned}
\tag{27}
$$

With $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ identified by the four equalities in (23)–(25), and $\sigma_{\epsilon_Z}, \sigma_{\epsilon_T}, \sigma_{\epsilon_M}, \sigma_{\epsilon_Y}$ identified by the four equalities in (26), the model in (27) is therefore just-identified for ρ_{TM}, ρ_{MY}.

The identification formulas to identify all linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ are described in the right-hand side of the expressions (23)–(25). Each identifying formula is associated with a well-known econometric estimator:

1. Parameter β_T^Z is identified by (23) as the covariance between Z, T divided by the variance of Z. This formula implies that β_T^Z can be estimated by the Ordinary Least Square (OLS) regression of T on Z.

$$\text{OLS: } T = \beta_T^Z \cdot Z + \epsilon_T. \tag{28}$$

2. Parameter β_M^T is identified by (24) as the ratio of the covariance between Z, M divided by the covariance of Z, T. This formula implies that β_M^T can be estimated by a Two-stage Least

Squares (2SLS) where Z is the instrumental variable, T is the endogenous explanatory variable and M is the outcome variable. Namely, β_M^T can be estimated by evaluating the standard 2SLS model:

$$\text{First Stage: } T = \beta_T^Z \cdot Z + \epsilon_T, \tag{29}$$

$$\text{Second Stage: } M = \beta_M^T \cdot T + \epsilon_M. \tag{30}$$

3. Parameters β_Y^M, β_Y^T are jointly identified by the two remaining equations in (25). In Online Appendix D we show that β_Y^M and β_Y^T are the expected values of the estimators of a 2SLS regression where T plays the role of a conditioning variable, Z is the instrument, M is the endogenous variable and Y is the dependent variable. Namely, β_Y^M and β_Y^T can be estimated by evaluating the following two-stage model:

$$\text{First Stage: } M = \gamma_M^Z \cdot Z + \gamma_M^T \cdot T + \epsilon_T, \tag{31}$$

$$\text{Second Stage: } Y = \beta_Y^M \cdot M + \beta_Y^T \cdot T + \epsilon_Y. \tag{32}$$

The estimation procedure associated with identification formulas (31) and (32) is not as common as those in (29) and (30). In fact, it is a novel property of the framework laid out here that 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 illustrate it with our empirical application to the context of import exposure, labor markets and voting behavior: As previously noted, the main endogeneity concern in a regression of regional manufacturing employment (M) on import exposure (T) is that unobserved adverse regional demand shocks reduce regional imports as well as employment. We therefore instrument for regional import exposure. For instance, an industry-j-specific domestic demand shock will reduce both local import exposure (T) and local employment (M) in regions that are specialized in industry j. 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. For illustration, consider industry-specific imports from China to a specific country, say Australia. The identifying assumption in Autor et al. (2013) is that Australian industry-specific imports are independent of German domestic demand conditions (at least conditional on controls). The key piece of intuition is that high Australian industry-specific imports from China conditional on (i.e. relative to) German industry-specific imports from China (T) will partly reflect or proxy for German industry-specific demand conditions, i.e. the source of the bias. Conditional on German imports (T), higher Australian imports from China (Z) in a given sector therefore 'causes' additional reductions in German employment, by virtue of proxying for negative German demand conditions.[22]

It is worth making the link to Table 1 in the Introduction explicit. *Model I* stands for the standard IV model that evaluates the effect of T on M using Z as the instrument. This model is estimated by the 2SLS regression defined by equations (29) and (30). *Model III* is the mediation model with instrumental variables. This model is estimated by *Model I* plus the 2SLS regression represented by the linear equations (31)–(32). *Model II* stands for the IV model that evaluates the total effect (TE) of T on Y. In Table 1, we have $T = f_T(Z, \epsilon_T)$ on $Y = g_Y(T, \eta_Y)$. The independence relation $Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$ induces the exclusion restriction $Y(t) \perp\!\!\!\perp Z$ and T is endogenous due to the statistical dependence between error terms ϵ_T and η_Y. *Model II* is obtained from *Model III* by substitution of the mediation variable M in (30) into the outcome equation in (32):

$$Y = \beta_Y^M \cdot M + \beta_Y^T \cdot T + \epsilon_Y \text{ and } M = \beta_M^T \cdot T + \epsilon_M \tag{33}$$

$$\Rightarrow Y = \beta_Y^M \cdot \left(\beta_M^T \cdot T + \epsilon_M\right) + \beta_Y^T \cdot T + \epsilon_Y \tag{34}$$

$$= \underbrace{\left(\beta_Y^M \cdot \beta_M^T + \beta_Y^T\right)}_{TE} \cdot T + \underbrace{\beta_Y^M \epsilon_M + \epsilon_Y}_{\eta_Y} \equiv g_Y(T, \eta_Y). \tag{35}$$

Error term η_Y of *Model II* is mapped into $\beta_Y^M \epsilon_M + \epsilon_Y$ in (35). Thus the correlation between η_Y and ϵ_T results from the correlation between ϵ_M and ϵ_T and the independence $Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$ is due to $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$ in **A-1**.[23]

2.2 Controlling for Additional Covariates

These methods can be easily extended to enable conditioning on additional covariates. Let $\mathbf{K}$ denote a set of variables that we wish to control for. By this we mean that variables K cause

[22] Whether they indeed do so is a question of explanatory power, not econometric identification.

[23] In Online Appendix E, we investigate the particular case in which the instrument Z consists of a single variable. We show that the estimate of the total effect $TE = \beta_Y^M \cdot \beta_M^T + \beta_Y^T$ evaluated by the 2SLS regressions in (29)–(30) and (31)–(32) is numerically the same as the standard 2SLS estimate of the causal effect of T on Y in *Model II*, namely, the 2SLS estimate of θ_Y^T that uses $T = \beta_T^Z \cdot Z + \epsilon_T$ for the first stage and $Y = \theta_Y^T \cdot T + \epsilon_Y$ for the second stage.

observed variables $\mathbf{X} = [Z, T, M, Y]'$. In other words, part of the covariation in $\mathbf{X}$ is not due to the causal relations of model (12)–(15), but due to the correlation induced by covariates $\mathbf{K}$. Thus, our task is to isolate the portion of covariance $\mathbf{\Sigma_X}$ that is caused by variables $\mathbf{K}$.

It is standard to use orthogonal projections to decompose a covariance matrix $\mathbf{\Sigma_X}$ into the part due to variables $\mathbf{K}$ and its complement. Let $\mathcal{K}$ be the linear space spanned by variables $\mathbf{K}$ and $\mathcal{K}^{\perp}$ be the orthogonal complement to $\mathcal{K}$. Thus the covariance $\mathbf{\Sigma_X}$ can be decomposed as:

$$\Sigma_{\mathbf{X}} = \mathbf{\Sigma}_{\mathbf{X}|\mathcal{K}} + \mathbf{\Sigma}_{\mathbf{X}|\mathcal{K}^{\perp}} \tag{36}$$

$$\text{where } \mathbf{\Sigma}_{\mathbf{X}|\mathcal{K}^{\perp}} = \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X},\mathbf{K}}\Sigma_{\mathbf{K}}^{-1}\Sigma'_{\mathbf{X},\mathbf{K}}. \tag{37}$$

Controlling for covariates $\mathbf{K}$ is achieved by simply replacing the covariance matrix $\mathbf{\Sigma_X}$ in (19) by $\mathbf{\Sigma}_{\mathbf{X}|\mathcal{K}^{\perp}}$ in (37). All previous identification results follow. In practice, replacing $\mathbf{\Sigma_X}$ by $\mathbf{\Sigma}_{\mathbf{X}|\mathcal{K}^{\perp}}$ is equivalent to adding covariates $\mathbf{K}$ as conditioning variables in the linear regressions (28)–(32) that estimate the model coefficients. For instance, instead of (28), parameter β_T^Z can now be estimated by the following OLS regression:

$$\text{New OLS for } \beta_T^Z\text{: } T = \beta_T^Z \cdot Z + \beta_T^K \cdot K + \epsilon_T, \tag{38}$$

and equivalently for regressions (29)–(32).[24]

2.3 Allowing for General Error Dependency

Assumption **A-2** allows for statistical dependence between error terms $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$ and $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$. **A-2** also allows for the conditional dependence of error terms $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y|\epsilon_M$, and assumes these error terms ϵ_T, ϵ_Y are unconditionally independent $\epsilon_T \perp\!\!\!\perp \epsilon_Y$. This error structure has appeal: it renders T, M and Y endogenous, but still enables the identification of all causal effects in a just-identified model. As previously noted, it is equivalent to assuming that T is endogenous in a regression of Y on T only because of confounders that affect T and M. We view this assumption as quite plausible in our setting, but there is clearly a concern that our framework may also be applied to settings where the identifying assumption is harder to defend. It is therefore important to extend

[24] Equations (48) and (49) in section 4 are the empirical equivalent of equations (29) and (30), with controls K added. Equations (50) and (51) in section 5 are the empirical equivalent of equations (31) and (32).

the framework to relax Assumption **A-2**, in order to gauge the importance of the independence $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ for the results.

Allowing $\rho_{TY} \neq 0$ is equivalent to stating that the statistical dependence among error terms is unrestricted. This generates a model with eleven parameters: four coefficients $\beta_T^Z, \beta_M^T, \beta_Y^T, \beta_Y^M$, four error variances $\sigma^2_{\epsilon_Z}, \sigma^2_{\epsilon_T}, \sigma^2_{\epsilon_M}, \sigma^2_{\epsilon_Y}$, and three correlations $\rho_{TM}, \rho_{TY}, \rho_{MY}$. The identification of these coefficients relies on the matrix equality in (20). An identification problem arises because (20) yields only ten identifying equalities. We show that these equalities render six of the eleven model parameters point-identified. Those are the coefficients β_T^Z, β_M^T, the correlation ρ_{TM} and the variances $\sigma^2_{\epsilon_Z}, \sigma^2_{\epsilon_T}, \sigma^2_{\epsilon_M}$.

The remaining five parameters are not point-identified. Those are the coefficients β_Y^M, β_Y^T, the correlations ρ_{TY}, ρ_{MY} and the variance $\sigma^2_{\epsilon_Y}$. As a consequence, neither the direct effect $DE = \beta_Y^T$ nor the indirect effect $IE = \beta_M^T \cdot \beta_Y^M$ is point-identified. We can still evaluate bounds for these effects. We estimate the range of possible values that these treatment effects can take when error terms are allowed to have any dependence relation. The subsequent identification follows the same steps as in the previous section.

Model identification relies on matrix equation $\widetilde{\Sigma}_{\mathbf{X}} = \Sigma_{\boldsymbol{\epsilon}}$, where $\widetilde{\Sigma}_{\mathbf{X}} \equiv (\mathbf{I} - \mathbf{\Psi})\, \Sigma_{\mathbf{X}}\, (\mathbf{I} - \mathbf{\Psi})'$ as in (20). This yields ten linear equalities given by $\widetilde{\Sigma}_{\mathbf{X}}[i,j] = \Sigma_{\boldsymbol{\epsilon}}[i,j]$ for $i \leq j; i,j \in \{1,2,3,4\}$. The independence relation $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M)$ implies that $\Sigma_{\boldsymbol{\epsilon}}[1,2] = \Sigma_{\boldsymbol{\epsilon}}[1,3] = 0$. Thereby the equalities $\widetilde{\Sigma}_{\mathbf{X}}[1,2] = 0$ and $\widetilde{\Sigma}_{\mathbf{X}}[1,3] = 0$ as in (23)–(24) still hold. As a consequence, the coefficients β_T^Z, β_M^T remain unchanged and are still identified by $\beta_T^Z = \frac{\sigma_{ZT}}{\sigma_{ZZ}}$ and $\beta_M^T = \frac{\sigma_{ZM}}{\sigma_{ZT}}$. Coefficient β_T^Z and β_M^T can still be evaluated by the OLS regression in (28) and the 2SLS in (29)–(30) respectively. Coefficients β_T^Z, β_M^T refer to *Model I* in Table 1. The model is not altered by the causal relation between T and Y and thereby β_T^Z, β_M^T are not affected by relaxing $\rho_{TY} \neq 0$.

Error variances are identified by the diagonal of the matrix equality $\widetilde{\Sigma}_{\mathbf{X}} = \Sigma_{\boldsymbol{\epsilon}}$. The equality $\widetilde{\Sigma}_{\mathbf{X}}[1,1] = \Sigma_{\mathbf{e}}[1,1]$ implies that $\sigma_{ZZ} = \sigma^2_{\epsilon_Z}$. Furthermore, the identification of β_T^Z and the observed covariances enable the identification of error variance $\sigma^2_{\epsilon_T}$ by the following equation:

$$\widetilde{\Sigma}_{\mathbf{X}}[2,2] = \Sigma_{\mathbf{e}}[2,2] \quad \Rightarrow \quad \left(\sigma_{TT} - \beta_T^Z \sigma_{ZT}\right) - \beta_T^Z \left(\sigma_{ZT} - \beta_T^Z \sigma_{ZZ}\right) \quad = \quad \sigma^2_{\epsilon_T} \tag{39}$$

In addition, the identification of β_M^T enables the identification of $\sigma^2_{\epsilon_M}$ by the following equation:

$$\widetilde{\Sigma}_{\mathbf{X}}[3,3] = \Sigma_{\mathbf{e}}[3,3] \quad \Rightarrow \quad \left(\sigma_{MM} - \beta_M^T \sigma_{TM}\right) - \beta_M^T \left(\sigma_{TM} - \beta_M^T \sigma_{TT}\right) \quad = \quad \sigma^2_{\epsilon_M} \tag{40}$$

Parameters β_T^Z, β_M^T and variances $\sigma^2_{\epsilon_T}, \sigma^2_{\epsilon_M}$ enable the identification of correlation ρ_{TM} via the following equation:

$$\widetilde{\Sigma}_{\mathbf{X}}[2,3] = \Sigma_{\mathbf{e}}[2,3] \quad \Rightarrow \quad \sigma_{TM} - \beta_T^Z \sigma_{ZM} - \beta_M^T(\sigma_{TT} - \beta_T^Z \sigma_{ZT}) \quad = \quad \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} \tag{41}$$

We are left with four equalities $\widetilde{\Sigma}_{\mathbf{X}}[i,j] = \Sigma_{\mathbf{e}}[i,j]$ such that $(i,j) \in \{(2,3),(1,4),(2,4),(3,4)\}$, and five parameters $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma^2_{\epsilon_Y}$ that are not point-identified. If any of these five parameters were known, the remaining four parameters would be just identified. It is useful to define an auxiliary variable $\kappa \equiv \rho_{MY} \cdot \sigma_{\epsilon_Y}$ to examine the identification of the model. Suppose κ was known, then the model parameters could be identified by the following procedure:

1. The equalities $\widetilde{\Sigma}_{\mathbf{X}}[1,4] = \Sigma_{\mathbf{e}}[1,4]$, $\widetilde{\Sigma}_{\mathbf{X}}[2,4] = \Sigma_{\mathbf{e}}[2,4]$ generate the following equations:
$$\sigma_{ZY} - \beta_Y^M \sigma_{ZM} - \beta_Y^T \sigma_{ZT} = 0 \tag{42}$$
$$\sigma_{TY} - \beta_Y^M \sigma_{TM} - \beta_Y^T \sigma_{TT} = \underbrace{\rho_{TY}\sigma_{\epsilon_Y}}_{\kappa}\sigma_{\epsilon_T}. \tag{43}$$

   If κ was known, the coefficients β_Y^M, β_Y^T could be obtained by the following formula:
$$\begin{bmatrix} \beta_Y^M \\ \beta_Y^T \end{bmatrix} = \left(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \cdot \left(\mathbf{B}'\mathbf{A}^{-1}\mathbf{C}\right), \tag{44}$$
$$\text{where } \mathbf{A} = \begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} \\ \sigma'_{ZT} & \sigma_{TT} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \sigma_{ZM} & \sigma_{ZT} \\ \sigma'_{TM} & \sigma_{TT} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \sigma_{ZY} \\ \sigma_{TY} + \kappa \cdot \sigma_{\epsilon_T} \end{bmatrix}. \tag{45}$$

2. Coefficients β_Y^M, β_Y^T enable the evaluation of error variance $\sigma^2_{\epsilon_Y}$ via equality $\widetilde{\Sigma}_{\mathbf{X}}[4,4] = \Sigma_{\mathbf{e}}[4,4]$ in (26).

3. The evaluation of σ_{ϵ_Y} in addition to $\beta_Y^T, \beta_Y^M, \beta_M^T, \sigma_{\epsilon_M}$ enables the identification of the correlation ρ_{MY} via the equality $\widetilde{\Sigma}_{\mathbf{X}}[2,3] = \Sigma_{\mathbf{e}}[2,3]$ in (27).

4. Finally, the correlation ρ_{TY} can be identified through the ratio $\rho_{TY} = \kappa/\sigma_{\epsilon_Y}$.

The identification of model parameters is anchored on value of the parameter κ. Under no model constrains, the parameter κ could take any value in the real line, which would render uninformative bounds for parameters $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma^2_{\epsilon_Y}$. Fortunately, we can rely on additional model restrictions that delimit the range of values that κ can take. For instance, κ is defined by $\kappa \equiv \rho_{MY} \cdot \sigma_{\epsilon_Y}$, where the outcome error variance $\sigma^2_{\epsilon_Y}$ is smaller or equal than the variance of the outcome σ_{YY} itself, i.e. $\sigma_{YY} \geq \sigma^2_{\epsilon_Y}$. This model property limits the range of possible values of κ to $|\kappa| \leq \sqrt{\sigma_{YY}}$. Another restriction arises from the fact that the covariance $\Sigma_{\mathbf{e}}$ must be a positive-definite matrix. That is to say that its eigenvalues[25] are strictly positive. We also benefit from the restrictions on the model correlations $0 \leq |\rho_{TY}| \leq 1$ and $0 \leq \rho_{TY} \leq 1$. We evaluate the κ-interval that complies with all these model restrictions. We then use this interval to generate the bounds for: (1) the direct effect $DE = \beta_Y^T$; (2) the indirect effect $IE = \beta_M^T \cdot \beta_Y^M$; (3) the total effect $TE = \beta_Y^T + \beta_M^T \cdot \beta_Y^M$; (4) the share of the total effect that is mediated by the indirect effect, namely, $S = IE/TE = \left(\beta_M^T \cdot \beta_Y^M\right) \cdot \left(\beta_Y^T + \beta_M^T \cdot \beta_Y^M\right)^{-1}$.

3 Data

Our data is organized as a 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),[26] and again tripled between 1998 and 2009.

Our data is observed at the county (*Landkreis*) level.[27] 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 two separate *Landkreise*-maps for periods 1 and 2 in Appendix

[25] A a square matrix $\mathbf{A}$ of dimension N has N eigenvalues $\lambda_n; n = 1, ..., N$ that are defined as the root-solutions of polynomial generated by $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$, where $\det(\cdot)$ is the matrix determinant and $\mathbf{I}$ stands for the identity matrix of dimension N.

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

[27] 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.

B. We include period-specific fixed effects for four broad regions in all estimations (North, West, South, and East Germany). 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 2, we need the following variables: *Treatment* T_{it} is our measure of local labor market i's import 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.[28]

3.1 Import Exposure (Treatment T)

We follow Autor et al. (2013) and Dauth et al. (2014) and calculate T as *net import exposure*:[29]

$$T_{it} = \sum_j \frac{L_{ijt}}{L_{jt}} \frac{\Delta IM_{Gjt} - \Delta EX_{Gjt}}{L_{it}}. \tag{46}$$

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.[30] 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 import exposure in equation (46) 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

[28] Conditioning variables K_{it} are discussed with the results in section 4.1.

[29] Throughout the paper, we will refer to net import exposure as import exposure for short.

[30] The *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.

known, and is solved by always conditioning on region i's initial share of manufacturing employment in all our regressions (Autor et al., 2013).

3.2 Labor Market Variables (Mediator M)

We use the *Institut für Arbeitsmarkt- und Berufsforschung* (IAB)'s Historic Employment and Establishment Statistics (HES) database to glean information on workers' industry of employment, occupation, and place of work for all German workers subject to social insurance.[31] Individual-level data are aggregated up to the *Landkreis* level to match our voting data. We consider decadal changes in (*i*) total employment, (*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 F provides additional information on data sources and variable construction.

3.3 Voting (Final Outcome Y)

To measure how import exposure affects voting behavior, we focus on party-votes in federal elections in Germany (*Bundestagswahlen*).[32] 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.[33] 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

[31] see 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.

[32] The 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.

[33] This 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.

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).[34]

3.4 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 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}):[35]

$$Z_{it}^{IM} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta IM_{Ojt}}{L_{it-1}} \;, \qquad Z_{it}^{EX} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta EX_{Ojt}}{L_{it-1}}. \tag{47}$$

Finally, 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 $t-1$.

4 Baseline Results

In this section, we estimate *Model I* (trade effect on labor markets) and *Model II* (trade effect on voting) using the standard IV approach.

4.1 Model I

Under linearity, *Model I* can be estimated by evaluating the following two-stage model:

$$T_{it} = \beta_T^Z \cdot Z_{it} + \beta_T^K \mathbf{K}_{it} + \eta_{it}, \tag{48}$$

$$M_{it} = \beta_M^T \cdot T_{it} + \beta_M^K \mathbf{K}_{it} + \epsilon_{it}. \tag{49}$$

[34] Online Appendix G provides additional background on the German political system and party landscape, and in Online Appendix H we present descriptive patterns on the vote share variables by period.

[35] We 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.

Equations (48) and (49) are equations (29) and (30) with control variables $\mathbf{K}_{it}$ included and subscripts added to reflect the panel nature of the data. *Model II* will be estimated by the same model, with Y_{it} replacing M_{it} in the second stage (49).

$\mathbf{K}_{it}$ includes i's start-of-period manufacturing employment share; the employment share in the largest sector;[36] along with separate controls for the employment share in car manufacturing and the chemical industry;[37] the start-of-period employment share that is foreign born, and the start-of-period unemployment rate; finally, the start-of-period vote-shares for all parties (with the SPD as the ommitted party share. Further included is 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).[38] 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).

We use the exact same set of controls variables in *Model I* and *Model II*. The core economic controls are well-motivated and routinely included in the related literature. Adding social and baseline voting controls is motivated by *Model II* and turns out to have more bearing on the estimates of interest in *Model II* than in *Model I*, as one would expect. We emphasize that *Model I* has been examined in detail in other papers including Autor et al. (2013), whose results have already been replicated for Germany by Dauth et al. (2014). We therefore keep the results for *Model I* brief. There is, however, one aspect that we need to explore for the purposes of our analysis. There is naturally more than one observed labor market outcome that is likely to be impacted by import competition. In addition to (*i*) total employment, we additionally observe (*ii*) manufacturing's employment share, (*iii*) manufacturing wages, (*iv*) non-manufacturing wages, and (*v*) unemployment. total employment (*i*) is probably the most direct measure of the potential consequences of import exposure, but we find that (*ii*) and (*iii*) are also significantly impacted by import competition, with weak effects also for (*iv*) and (*v*).[39]

Without additional separate dedicated sets of instruments for each potential mediator, our

[36]It 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.

[37]The latter account for those industries' outstanding importance for the German economy.

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

[39] This is consistent with prior research that clearly shows that the labor market effects of import exposure are concentrated in manufacturing employment (Autor et al., 2013; Dauth et al., 2014).

method can only identify the effect of as many mechanisms as there are treatments. Indeed, it is likely to be a common problem for researchers interested in applying the method developed here that there will often be a number of observed variables that potentially link a treatment T to an outcome Y. 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.[40] 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).

Table 3: German Labor Markets' Principal Components' Factor-Loadings

	(1) Δ log employment	(2) Δ Share Manuf. Empl.	(3) Δ log Manuf. Wage	(4) Δ log Non-Manuf. Wage	(5) Δ Share Unempl.
LMC_1	0.1711	-0.3632	0.5108	0.5486	0.5261
LMC_2	0.7625	0.6004	0.2104	0.0607	-0.1012
LMC_3	-0.5389	0.397	0.5311	0.3251	-0.4053

Notes: The table reports on the factor loadings of the five labor market variables on LMC_1 and LMC_2. See discussion in text. LMC_1's eigenvalue is 2.707, explaining 54.1 percent of the total variation. LMC_2's eigenvalue is 1.281, explaining 25.6 percent of the total variation. LMC_3's eigenvalue is 0.509, explaining 10.2 percent of the total variation.

One can best interpret the LMCs through their relation to the labor market outcomes we observe, specifically through their factor loadings. Table 3 reports on the LMCs' factor loadings. By construction, there are always as many LMCs as variables but to keep this section brief, we report only the first three. The convention is to report these only for the LMCs with an eigenvalue larger than 1, i.e. only the first two. In our data, LMC_1 and LMC_2 together explain about 80 percent of the variation in the labor market data $(0.541 + 0.256)$. The third LMC only explains about 10.2 percent, the fourth and fifth together explain the remaining 10.1 percent.

LMC_1 is somewhat ambiguous: it loads positively on changes in wages, but negatively on

[40]By 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).

manufacturing employment and also positively on unemployment.[41] By contrast, LMC_2's interpretation is unambiguous: Its factor loadings are strongly positive for changes in manufacturing employment and total employment, and negative for changes in unemployment. LMC_2 is clearly associated with the labor market aspects that we know to be most affected by import exposure.

The estimation results of *Model I* are displayed in table 4, each cell reporting on a different regression specification. In column 1, our least conservative specification considers only the start-of-period manufacturing employment share as a control.[42] In column 2, 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 3, we add controls for the start-of-period employment share that is foreign born and the start-of-period unemployment rate. In column 4, we add start-of-period vote-shares for all parties, with the SPD being the omitted baseline. The addition of controls in columns 3 and 4 is motivated by *Model II* but we will need to maintain the same control variables for labor market outcomes in our mediation framework and therefore apply the same here.

In the upper part of Panel A, we investigate the five individual labor market outcomes, where we view total employment as the primary one. Import exposure has a significant negative effect on total employment, and on manufacturing employment and wages, as well as a weak effect on the unemployment rate. In our preferred specification in column 4, a one-standard-deviation increase in T_{it} (€1,350 per worker) decreases total employment by about 3 percent ($e^{-0.024 \cdot 1.35} - 1 = -0.032$).[43] In the lower part of Panel A, we investigate the effect of T_{it} on the five principal components LMC_1–LMC_5. We argued that PCA is appealing if the mediating effects turn out to be sharply concentrated in one PC, and this PC has a clear interpretation. Indeed, this turns out to be the case here: Comparing the results for LMC_2 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

[41] Our interpretation of LMC_1 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 LMC_1 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).

[42] As is common in this literature, 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 3.4.

[43] We report corresponding OLS results in Online Appendix I (table 3).

Table 4: Effect of Import Exposure T_{it} on Labor Markets M_{it}

	(1)	(2)	(3)	(4)
Panel A: Second Stage (49), for Individual Labor Market Outcomes				
M: Δ log employment	-0.023***	-0.021**	-0.023***	-0.023***
	[0.004]	[0.012]	[0.005]	[0.004]
M: Δ Share Manuf. Empl.	-0.440**	-0.515**	-0.719***	-0.757***
	[0.048]	[0.022]	[0.001]	[0.000]
M: Δ log Manuf. Wage	-0.006**	-0.007***	-0.006***	-0.006***
	[0.013]	[0.004]	[0.006]	[0.008]
M: Δ log Non-Manuf. Wage	-0.005***	-0.005***	-0.003	-0.002
	[0.004]	[0.008]	[0.112]	[0.265]
M: Δ Share Unempl.	0.076	0.049	0.107	0.099
	[0.271]	[0.575]	[0.140]	[0.134]
M: LMC_1	-0.033	-0.036	0.019	0.030
	[0.489]	[0.441]	[0.624]	[0.410]
M: LMC_2	-0.264***	-0.266***	-0.317***	-0.324***
	[0.003]	[0.004]	[0.001]	[0.000]
M: LMC_3	-0.028	-0.051	-0.072*	-0.066*
	[0.517]	[0.277]	[0.052]	[0.062]
M: LMC_4	0.024	-0.002	-0.015	-0.027
	[0.409]	[0.957]	[0.621]	[0.366]
M: LMC_5	-0.016	0.002	0.006	0.008
	[0.433]	[0.912]	[0.773]	[0.718]
Panel B: First Stage Equation (48)				
β_M^{IM}	0.225	0.212***	0.214***	0.217***
	[0.000]	[0.000]	[0.000]	[0.000]
β_M^{EX}	-0.211	-0.206***	-0.206***	-0.202***
	[0.000]	[0.000]	[0.000]	[0.000]
Controls	Baseline	+ Industry	+Socio	+ Voting
R-Squared	0.524	0.551	0.555	0.566
F-Stat. of excluded Instruments	43.81	43.64	40.15	38.77
Period-by-region FE	Yes	Yes	Yes	Yes
Observations	730	730	730	730

Notes: (*a*) Each cell reports results from a separate IV 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. (*b*) All specifications include region-by-period fixed effects. Columns incrementally add controls. (1) controls only for start-of-period manufacturing and period-specific region fixed effects. (2) adds controls for dominant industries (employment share of the largest industry, in automobiles, and chemicals). (3) adds controls for the structure of the workforce (share foreign workers, and population share unemployed). (4) adds beginning-of-period voting controls. (*c*) Across rows, we investigate different outcomes. For example, the top-row reports a semi-elasticity where a one-standard-deviation increase in T_{it} (€1,350 per worker) decreased total employment by about 3 percent, ($e^{-0.024 \cdot 1.35} - 1 = -0.032$). (*d*) The bottom panel reports the first stage results. *p-values* are reported in square brackets, standard errors are clustered at the level of 96 commuting zones. *** p<0.01, ** p<0.05, * p<0.1.

0.001 in all specifications. By contrast, the other four LMCs are largely unaffected by import competition, with only the third one showing a weak response.[44]

For the estimation of *Model III* we will later focus on LMC_2 because it appears to summarize the effects of import exposure on labor markets very well in this data.

Panel B reports on the first-stage results of estimating equation (48). These results are highly significant, have the expected sign and are in line with estimates in the existing literature.[45] We do not study any individual-level labor market results, but refer the interested reader to Dauth et al. (2014) who present such evidence for Germany.

4.2 Model II

We now turn to estimating *Model II*, which shares the first stage equation (48) with *Model I*, but replaces M_{it} with Y_{it} in the second stage equation (49). We devote more attention to *Model II* because our empirical application is an investigation of the causes of political populism and the subsequent results are original to this study. In Table 1 in the Introduction *Model II* represents causal equations $T = f_T(Z, \epsilon_T)$ and $Y = g_Y(T, \eta_Y)$. The endogeneity implied by $\epsilon_T \not\perp\!\!\!\perp \eta_Y$ can be solved with the same instrument as in *Model I* because $Z \perp\!\!\!\perp \epsilon_T, \eta_Y$.

Table 5 presents our baseline results of this 2SLS estimation. Each cell reports results from a different regression. While we are primarily interested in political support for populists, our voting 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. Each row in table 5 pertains to one of these different outcome variables. Columns refer to different regression specifications, which are introduced in exactly the same manner as in table 4. In addition, column 5 reports the results from our preferred specification in column 4 as standardized coefficients to facilitate a comparison of the magnitudes of the effect on different election outcomes.

The effects are broadly consistent across all specifications, though we see that the stepwise inclusion of controls reduces the estimated magnitude. Table 5 suggests no effect on turnout;

[44] Given our interpretation of the LMCs, these results resonate closely with Autor, Dorn, and Hanson (2015) who show that import exposure has had large effects on (overall and manufacturing) employment whilst the polarization of work and the rise of service jobs (i.e. our LMC_1) were explained by other factors, primarily automation.

[45] For added clarity, we break the instrument $\beta_T^Z \cdot Z_{it}$ in the first stage equation (48) into a separate import and export instrument $\beta_T^{IM} \cdot Z_{it}^{IM} + \beta_T^{EX} \cdot Z_{it}^{EX}$. Whether we have one instrument or a vector of instruments has no bearing on our argument that we develop a method that requires instruments dedicated to only one endogenous variable T.

Table 5: Effect of Import Exposure (T_{it}) on Voting (Y_{it})

	(1)	(2)	(3)	(4)	(5)
Δ Turnout	0.002	0.003	0.003	0.002	0.030
	[0.348]	[0.244]	[0.219]	[0.289]	[0.289]
<u>Established Parties:</u>					
Δ voteshare CDU,CSU	-0.128	-0.206	-0.211	-0.066	-0.016
	[0.457]	[0.271]	[0.246]	[0.630]	[0.630]
Δ voteshare SPD	-0.020	0.015	0.023	-0.010	-0.001
	[0.897]	[0.928]	[0.881]	[0.941]	[0.941]
Δ voteshare FDP	0.215***	0.232***	0.222***	0.122	0.022
	[0.005]	[0.004]	[0.004]	[0.109]	[0.109]
Δ voteshare Greens	-0.132**	-0.131**	-0.073	-0.042	-0.015
	[0.022]	[0.029]	[0.188]	[0.374]	[0.374]
<u>Non-Established Parties:</u>					
Δ voteshare Extreme-Right	0.118***	0.140***	0.131***	0.102**	0.051**
	[0.001]	[0.001]	[0.002]	[0.032]	[0.032]
Δ voteshare Far-Left	-0.037	-0.064	-0.095	-0.079	-0.020
	[0.773]	[0.650]	[0.417]	[0.501]	[0.501]
Δ voteshare Other Small	-0.015	0.013	0.002	-0.028	-0.021
	[0.696]	[0.739]	[0.957]	[0.528]	[0.528]
Controls	Baseline	+ Industry	+Socio	+ Voting	∼(4)
Period-by-region F.E.	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 identical fashion as table 4. Column 4 is our preferred specification, which includes socioeconomic controls. Column 5 reports on the same specification with standardized outcome variables to facilitate comparison of magnitudes across rows. (*c*) Across rows, we report on different voting outcomes. Our focus will be on the most significant, and in column 5 also most sizeable, voting response, which occurs on the extreme right. (*d*) The first stage is identical to that in table 4. *p-values* are reported in square brackets, standard errors are clustered at the level of 96 commuting zones. *** p<0.01, ** p<0.05, * p<0.1.

and looking at reactions across the political spectrum, we see no significant effects on established, small, or far-left parties in our preferred specification in column 4. The only segment of the party spectrum that responds consistently to trade shocks across all specifications is the vote-share of extreme-right parties.[46] In our preferred specification in column 4, a one-standard-deviation increase in T_{it} (€1,350 per worker) increases the extreme-right vote share by 0.14 ($0.102 \cdot 1.35$) percentage points, roughly 35 percent of the average per-decade increase of 0.43 percentage points during the 22 years we study. Column 5 reports the results from our preferred specification as beta coefficients, which shows 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.[47]

Interpreting the Effects: Whether far-left or extreme-right populists capture protectionist sentiments 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."[48] In Appendix B 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 im-

[46] There is also some evidence of polarization in the response of the vote share of the market-liberal FDP is only marginally insignificant. 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 protectionist sentiments in their region. Our focus is on the extreme-right's vote share.

[47] We present corresponding OLS estimates in table 4 of Online Appendix I.

[48] For 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).

portantly, however, Germany did not have a populist 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, the political backlash to import exposure may be more strongly reflected in changing vote shares. The coefficient size is thus specific to the political context. Our focus is not on the magnitude of the effect of import exposure on voting behavior but on the causal mechanisms underlying it.[49]

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.[50] The gravity results are reported in Appendix C and are in line with those in table 5.

Individual-Level Analysis: One benefit of using German data is that the Socio-Economic Panel (SOEP) has a long-run panel structure that is unique amongst attitudinal socio-economic surveys, starting in 1984 (GSOEP, 2007).[51] 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 import 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.[52] 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 5, 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 D, mimic closely our main table 5. A county's import exposure shifts individuals' preferences to the 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

[49]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 H.

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

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

[52]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.

sectors, who are also most likely to experience adverse labor market effects from import exposure.

Without *Model III*, this is is far as we can go. Standard IV methods generate a causally identified effect of import exposure on total employment (-0.024) and a causally identified effect of import exposure on voting (0.102). We now estimate the mediation model to identify to what extent the former explains the latter.

5 *Model III*: Mediation Analysis

In this section, we apply the estimation framework developed in Section 2 to estimate *Model III*. This allows us to estimate the indirect effect IE of import exposure on voting that runs through labor markets. The extent to which import exposure polarized voters because it caused labor market adjustments is identified by a comparison of this indirect effect with the total effect of import exposure on voting.[53]

Under linearity, *Model III* can be estimated by evaluating the following two-stage model:

$$M_{it} = \gamma_M^Z \cdot Z_{it} + \gamma_M^T \cdot T_{it} + \gamma_M^K \mathbf{K}_{it} + \epsilon_{it}, \tag{50}$$

$$Y_{it} = \beta_Y^M \cdot M_{it} + \beta_Y^T \cdot T_{it} + \beta_Y^K \mathbf{K}_{it} + \eta_{it}. \tag{51}$$

Equations (50) and (51) are exactly equations (31) and (32), except that control variables are included and subscripts added to reflect the panel nature of the data.[54]

Table 6 summarizes the results of applying our mediation model. There is good reason to only focus on the fourth specification because it includes baseline voting controls. As a point of comparison, however, we also report on the third specification, in columns 1 and 3. Furthermore Table 4 suggests that LMC_2 should be viewed as the main summary measure of labor market adjustments in our data. However, since principal components are harder to interpret, we begin in columns 1–2 with the log of total employment as the mediator, before then focusing on LMC_2 as the mediator in columns 3–4. Ultimately, column 4 presents the most important evidence in the table since it includes all relevant controls and uses our main summary measure of labor market

[53] The total effect is estimated in *Model II* as 0.102, reported in table 5. The indirect effect is $\hat{\beta}_M^T$, reported in table 4, multiplied by the effect of M on Y, which we now estimate.

[54] The first-stage coefficients are denoted by γ's instead of β's because they do not correspond to parameters in the causal model represented by equations (12)–(15).

Table 6: Main Estimates of the Mediation Model

	(1)	(2)	(3)	(4)
Mediating Variables:	Δ log(employment)		LMC_2	
	Panel A: Second Stage (51)			
β_Y^M	-6.275***	-6.113***	-0.441**	-0.401***
	[0.009]	[0.000]	[0.017]	[0.003]
$DE = \beta_Y^T$	-0.009	-0.050	-0.036	-0.071**
	[0.820]	[0.107]	[0.344]	[0.014]
$IE = \beta_M^T \cdot \beta_Y^M$	0.142*	0.140**	0.140**	0.130**
	[0.056]	[0.024]	[0.049]	[0.023]
IE/TE	$1.078 = \frac{0.141}{0.131}$	$1.372 = \frac{0.140}{0.102}$	$1.067 = \frac{0.140}{0.131}$	$1.274 = \frac{0.130}{0.102}$
	Panel B: First Stage Equation (50)			
γ_M^{EX}	0.006***	0.006***	0.079***	0.082***
	[0.000]	[0.000]	[0.000]	[0.000]
γ_M^{IM}	-0.004**	-0.004**	-0.018	-0.024
	[0.046]	[0.030]	[0.344]	[0.187]
γ_M^T	0.001	0.001	-0.086	-0.074
	[0.837]	[0.747]	[0.128]	[0.171]
Controls	+Socio	+ Voting	+Socio	+ Voting
Observations	730	730	730	730
R-Squared	0.471	0.506	0.354	0.401

Notes: (*a*) Columns 1–2 and 3–4 present results for two different mediators, Δlog(employment) and LMC_2. (*b*) Panel A reports estimates from the second-stage equation (51). The point estimate $\hat{\beta}_Y^M$ in column 2 indicates that a one-percent drop in employment raises the change in the extreme right's vote share by 0.06 percentage points ($6.113/100$). The implied IE is $-6.113 \times -0.023 = 0.140$, which is 137% of the total effect TE of 0.102 reported in table 5. In our main specification in column 4, the implied IE is $-0.401 \times -0.324 = 0.130$. (*c*) Panel B presents results of the the first-stage equation (50). The estimated coefficients have the expected sign. (*d*) *p-values* are reported in square brackets, with standard errors are clustered at the level of 96 commuting zones (*** p<0.01, ** p<0.05, * p<0.1).

adjustments as the mediator.

The bottom panel reports on estimating the first stage equation (50). It is helpful to estimate the first stage with a prior about coefficients' expected signs. As discussed, the literature worries about domestic industry-specific demand conditions as a source of confounding bias. German industries that experience negative domestic demand shocks will see fewer imports and less employment. The discussion in section 2 suggests that, conditional on Germany's imports T, other countries' imports may proxy for negative German demand. This is what we find in the top-panel of table 6. Other countries' imports from China worth €1,000 per worker reduce German employment by 0.4 percent ($\hat{\gamma}_M^{IM} = -0.004$), while the same amount in exports to China increases German employment by 0.6 percent ($\hat{\gamma}_M^{EX} = 0.006$).[55]

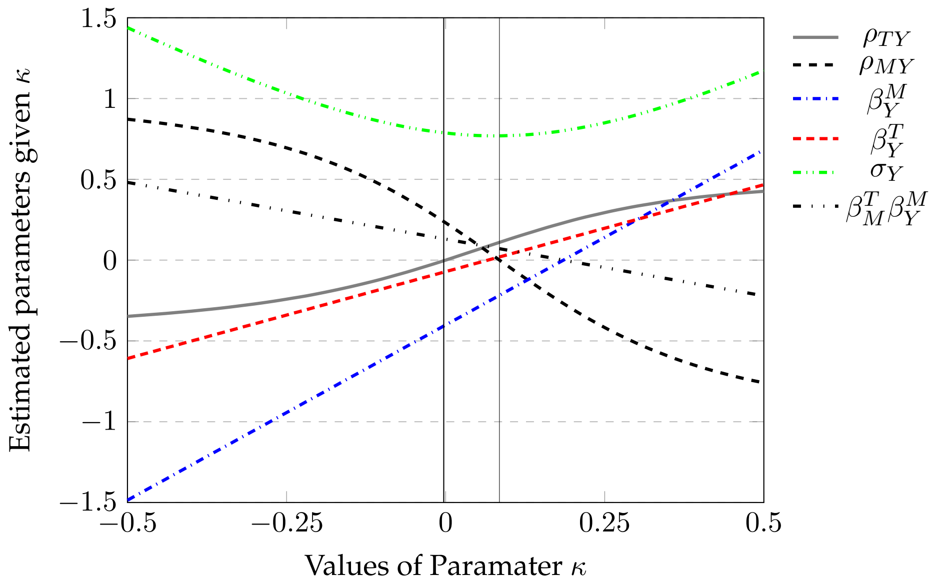
The top panel reports on estimating the second stage equation (51). The point estimate $\hat{\beta}_Y^M$ in column 2 indicates that a one-percent drop in employment raises the change in the extreme right's vote share by 0.06 percentage points, i.e. $6.113/100$, relative to an average per-decade increase of 0.43 percentage points during the 22 years we study. The indirect effect is the product of $\hat{\beta}_Y^M \times \hat{\beta}_M^T$, with the latter reported in table 4 ($-6.113 \times -0.023 = 0.140$). In column 2, the indirect effect is 137 percent of the *total effect* reported in table 5. Table 4 showed that LMC_2 is the only principal component that responded to import competition in the data. In column 4, the indirect effect implied by LMC_2 is 127 percent of the *total effect*. This is our preferred estimate since LMC_2 provides a summary measure of the impact of import exposure on German labor markets. The relative size of the estimated indirect effect to the total effect is less in columns 1 and 3 (108 and 107 percent respectively), but this is fully explained by a larger total effect in the denominator when baseline voting controls are not included.

In summary, our estimation framework suggests that the percentage of the total populist effect that is explained by labor markets ranges from 107 percent to 138 percent. When we consider only our preferred fourth specification, the range is tighter, from 127 percent to 138 percent. For the indirect effect to be more than 100 percent of the total, it needs to be partly offset by the direct effect. In other words, other channels linking import exposure to voting need to be politically moderating in the aggregate. Indeed, $\hat{\beta}_M^T$ in Table 6 consistently has a negative sign.[56] This implies that if

[55] As before, we use two instruments Z_{it}^{IM}, Z_{it}^{EX}, but the important point of our theory is that we need instruments for only one endogenous variable, namely T.

[56] While we can only speculate, other channels that are potentially moderating are import exposure's effect on task-

Figure 1: Graphing the Bounds



Notes: The horizontal axis is for the auxiliary variable κ; see section 2.3. On the horizontal axis are the five model parameters $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma^2_{\epsilon_Y}$ that are not point identified when we relax *Assumption **A-2***. The κ-interval that complies with the model restrictions of Section 2.3, namely $\kappa \in [-.003, 0.085]$, is delimited by vertical lines. This interval generates the bounds for the correlations ρ_{TY}, ρ_{MY}; the coefficients β_Y^M, β_Y^T; the indirect effect $\beta_M^T \cdot \beta_Y^M$. and the standard error σ_Y. The estimates in table 5 refer to the point estimates generated by setting κ to zero. Indeed, the parameter κ is defined as $\kappa = \rho_{YT} \cdot \sigma_Y$, thus $\rho_{YT} = 0 \Rightarrow \kappa = 0$.

the only effect of import exposure was to decrease employment, the political response would be stronger than the one observed in the data.

Our framework applies to mediation settings where T is endogenous in a regression of Y on T only because of confounding factors that influence M (*Assumption **A-2***). In our setting, this means that the effect of import exposure on voting is confounded only by unobserved factors that simultaneously impact imports and local labor markets, such as for example negative local demand shocks. The identifying assumptions underlying IV are always strong, and this is no different. While we view the stated assumption as plausible in our research setting, there are invariably many mediation settings our method could be applied to, where the identifying assumption is less plausible, or even implausible.

It is therefore important to be able to ask how our results would change if we relaxed this

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.

assumption,by allowing $\rho_{TY} \neq 0$. In section 2.3 we showed that we can do this and still calculate bounds on the model parameters. Formally, we define an auxiliary variable $\kappa = \rho_{MY} \cdot \sigma_{\epsilon_Y}$. We show that all coefficients are point-identified for a given value of κ. We thus investigate the values of κ that comply with the model restrictions discussed in Section 2.3. In our empirical setting, the κ-interval that accords with model restriction is given by $\kappa \in [-.003, 0.085]$. This interval in turn implies bounds on the five model parameters $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma^2_{\epsilon_Y}$. Figure 1 shows this graphically, and Table 7 reports the results.

Table 7: Bounds on the Mediation Model

Estimated Parameters		*Assumption A-2* holds	Bounds (No *Assumption A-2*)
Correlation	ρ_{TY}	0	$[-0.003, 0.110]$
Correlation	ρ_{MY}	0	$[0.001, 0.236]$
Correlation	ρ_{TM}	0.502	0.502
Coefficient	β_Y^M	-0.401	$[-0.218, -0.406]$
Direct Effect	$DE = \beta_Y^T$	-0.071	$[-0.073, 0.020]$
Indirect effect	$IE = \beta_M^T \cdot \beta_Y^M$	0.130	$[0.071, 0.131]$
Share	$S = IE/TE$	1.274	$[0.690, 1.284]$

Notes: (*a*) The reported values of the direct and indirect effect when *Assumption A-2* holds are those reported in column 4 of of Table 6. Relaxing *Assumption A-2* corresponds to allowing $\rho_{TY} \neq 0$. The possible range of values of ρ_{TY} is dictated by the model restrictions in section 2.3.

There is a range of possible correlations $\rho_{TY} \in [-0.003, 0.110]$. Different correlations ρ_{TY} imply different direct and indirect effects (while the total effect is unchanged by ρ_{TY}). The range $\rho_{TY} \in [-0.003, 0.110]$ implies the following bounds on the direct effect $\beta_Y^T \in [-0.073, 0.020]$ and indirect effect $\beta_M^T \cdot \beta_Y^M \in [0.071, 0.131]$. The range $\rho_{TY} \in [-0.003, 0.110]$ contains zero. Therefore, all causal effects evaluated under the assumption that $\rho_{TY} = 0$ (column 4 of Table 6) belong to the causal effect intervals generated by relaxing this correlation assumption. By contrast, the estimated bounds for ϵ_T, ϵ_M and ϵ_M, ϵ_Y are strictly positive. This is reassuring in corroborating our identification reasoning, although it should not be interpreted as a statistical test of *Assumption A-2*, being as these are not statistical bounds but implied by the model restrictions in section 2.3.[57]

[57] We do not impose the bounds for ρ_{TY} to contain 0. Suppose that zero did not belong to the bounds of ρ. As a consequence, the estimates of causal effects under the assumption $\rho_{TY} = 0$ would not belong to their respective bounds when the assumption $\rho_{TY} = 0$ is relaxed. This would not necessarily amount to a statistical rejection of the identifying assumption. It is still useful to consider the full range of values implied by the bounds in addition to the

The correlation structure in the data is consistent with labor market adjustments explaining between 69 and 128 percent of the observed political response to import exposure. Even at the lower end of this range, the evidence presented here clearly suggests that effective responses to political populism need to focus on labor markets first and foremost.

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 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 'direct 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. While our identifying assumption plausibly holds in our setting, we caution that

point estimate.

researchers need to critically evaluate its validity before applying it to other mediation-type IV settings.

References

Acemoglu, D. and P. Restrepo (2017). Robots and jobs: Evidence from us labor markets. *MIT Umpublished Mimeo.*.

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 ageempirical 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.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012). Plausibly exogenous. *Review of Economics and Statistics 94*(1), 260–272.

Dahl, C. M., M. Huber, and G. Mellace (2017). It's never too late: A new look at local average treatment effects with or without defiers. *Discussion Papers on Business and Economics, University of Southern Denmark, 2/2017*.

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).

Dix-Carneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica*, 825–885.

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 wissensgesellschaft: 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 Umpublished 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. Mdrc 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. Merkens, 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.

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. Vytlacil (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.

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.

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. (2011). The mediation formula: A guide to the assessment of causal pathways in nonlinear models. Forthcoming in *Causality: Statistical Perspectives and Applications*.

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.

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*.

Yamamoto, T. (2014, March). Identification and estimation of causal mediation effects with treatment noncompliance. Manuscript. Department of Political Science, Massachusetts Institute of Technology, Cambridge.

Appendix A Proofs

Proof P-1 *The treatment equation $T = f_T(Z, \epsilon_T)$ in (1) implies that $Z \not\perp\!\!\!\perp T$. Thus our task is to proof two exclusion restrictions: $Z \perp\!\!\!\perp M(t)$ and $Z \perp\!\!\!\perp Y(t)$. According to (4), the counterfactual mediation is given by $M(t) = f_M(t, \epsilon_M)$. But Assumption **A-1** states that $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$. In particular, we have that:*

$$Z \perp\!\!\!\perp \epsilon_T \Rightarrow Z \perp\!\!\!\perp f_M(t, \epsilon_M) \Rightarrow Z \perp\!\!\!\perp M(t). \tag{52}$$

We can use iterated substitution to express the outcome counterfactual $Y(t)$ in (5) as the following function of error terms:

$$Y(t) = f_Y(t, M(t), \epsilon_Y) = f_Y(t, f_M(t, \epsilon_M), \epsilon_Y) \text{ by (4)}, \tag{53}$$

$$\text{by } \textbf{A-1} \text{ we have that: } Z \perp\!\!\!\perp (\epsilon_M, \epsilon_Y) \Rightarrow Z \perp\!\!\!\perp f_Y(t, f_M(t, \epsilon_M), \epsilon_Y) \Rightarrow Z \perp\!\!\!\perp Y(t). \tag{54}$$

Proof P-2 *The lemma requires two proofs. The first consists in showing that Z is not independent of M conditioned on T, that is, $Z \not\perp\!\!\!\perp M|T$. The second consists in showing that the exclusion restriction $Z \perp\!\!\!\perp Y(m)|T$ holds under the independence condition $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ in Assumption **A-2**.*

A intuitive justification for $Z \not\perp\!\!\!\perp M|T$ relies on interpreting the correlations generated by condition on T. Recall that the treatment equation is given by $T = f_T(Z, \epsilon_T)$. Thus, conditioning on $T = t$ is equivalent to conditioning on the values of Z, ϵ_T such that $f_T(Z, \epsilon_T) = t$. This induces a correlation between Z and ϵ_T and thereby $Z \not\perp\!\!\!\perp \epsilon_T|T$. Moreover, ϵ_T correlates with ϵ_M and therefore we also have that $Z \not\perp\!\!\!\perp \epsilon_M|T$. But if $Z \not\perp\!\!\!\perp \epsilon_M|T$, then $Z \not\perp\!\!\!\perp f_M(T, \epsilon_M)|T$ and therefore we have that $Z \not\perp\!\!\!\perp M|T$ as $M = f_M(T, \epsilon_M)$. In summary, conditioning on T induces a correlation between Z and ϵ_T, but error term ϵ_T correlates with ϵ_M, which in turn generates a correlation between Z and M.

*It remains to show that the the independence relation $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ generate the exclusion restriction $Z \perp\!\!\!\perp Y(m)|T$ where the outcome counterfactual $Y(m)$ is given by $Y(m) = f_Y(T, m, \epsilon_Y)$ as in (6). The following rationale justify this assessment. Assumptions **A-1**–**A-2** generate the unconditional independence relation $(\epsilon_T, \epsilon_Z) \perp\!\!\!\perp \epsilon_Y$. Let $f_1(\cdot), f_2(\cdot), f_3(\cdot)$ be three arbitrary non-degenerate functions such that $f_1 : supp(\epsilon_Z) \times supp(\epsilon_T) \to \mathbb{R}$, $f_2 : supp(\epsilon_Z) \to \mathbb{R}$, $f_3 : supp(\epsilon_Y) \to \mathbb{R}$. Under this notation, we have that:*

$$(\epsilon_T, \epsilon_Z) \perp\!\!\!\perp \epsilon_Y \quad \Rightarrow \quad \epsilon_Z \perp\!\!\!\perp \epsilon_Y | f_1(\epsilon_Z, \epsilon_T) \quad \Rightarrow \quad f_2(\epsilon_Z) \perp\!\!\!\perp f_3(\epsilon_Y) | f_1(\epsilon_Z, \epsilon_T). \tag{55}$$

In particular, we can set functions $f_1(\epsilon_T), f_2(\epsilon_Y), f_3(\epsilon_Z, \epsilon_T)$ *in* (55) *to the following expressions:* $f_1(\epsilon_Z) = f_Z(\epsilon_Z)$, $f_2(\epsilon_Y) = f_Y(t, m, \epsilon_Y)$, *and* $f_3(\epsilon_Z, \epsilon_T) = f_T(f_Z(\epsilon_T), \epsilon_Z)$. *Thus:*

$$f_2(\epsilon_Z) \perp\!\!\!\perp f_3(\epsilon_Y) \mid f_1(\epsilon_Z, \epsilon_T) \tag{56}$$

$$\Rightarrow f_Z(\epsilon_Z) \perp\!\!\!\perp f_Y(t, m, \epsilon_Y) \mid \big(f_T(f_Z(\epsilon_T), \epsilon_Z) = t\big) \quad \forall (t, m) \in supp(T) \times supp(M) \tag{57}$$

$$\Rightarrow Z \perp\!\!\!\perp f_Y(t, m, \epsilon_Y) \mid \big(T = t\big) \quad \forall (t, m) \in supp(T) \times supp(M) \tag{58}$$

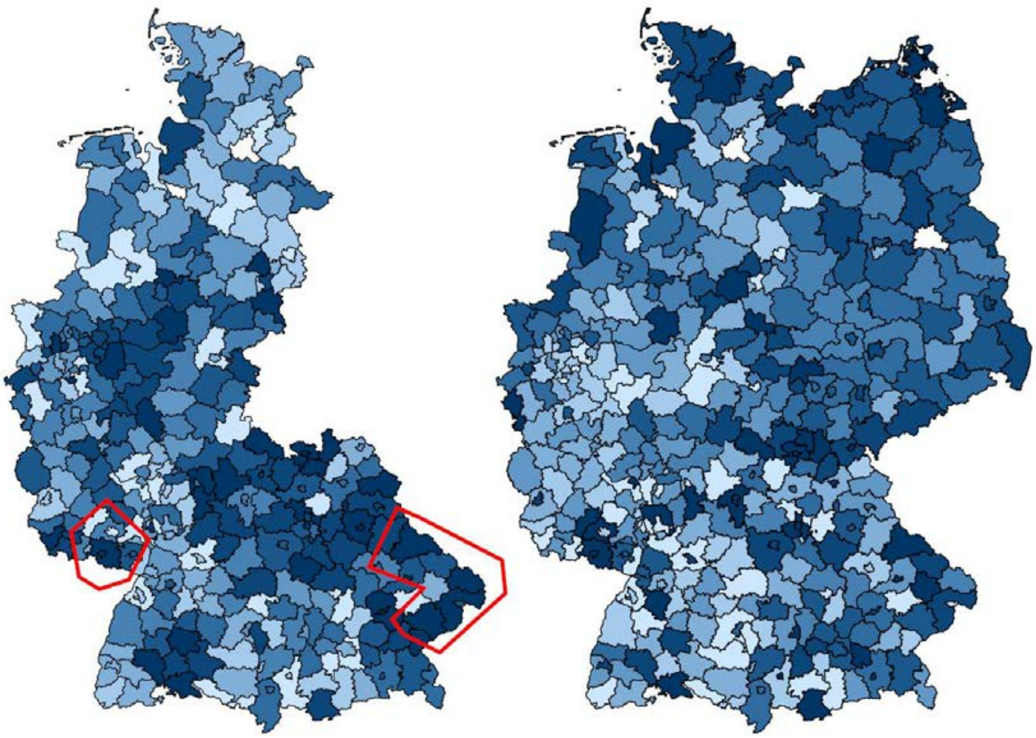
$$\Rightarrow Z \perp\!\!\!\perp Y(m)|T. \tag{59}$$

Proof P-3 *Assumptions* ***A-1***–***A-2*** *implies that* $\epsilon_Y \perp\!\!\!\perp (Z, \epsilon_T)$. *According to Equations* (6)*, we have that:*

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

where the third equality comes from $\epsilon_Y \perp\!\!\!\perp (Z, \epsilon_T)$.

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 B 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.[58] 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 southwest 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 the local *Republikaner* managed

[58]We 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.

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 H.

Appendix C 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') industries[59], 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^{grav}_{Gjt} to replace ΔIM_{Gjt}.[60] Online Appendix J shows how to obtain the gravity-residuals ΔIM^{grav}_{Gjt} that replace ΔIM_{Gjt} the gravity-residuals ΔEX^{grav}_{Gjt} that replace ΔEX_{Gjt}. An exogenous measure for changes in in German industries' import exposure can be obtained from netting out both effects such that $\Delta Trade^{grav}_{Gjt} = \Delta IM^{grav}_{Gjt} - \Delta EX^{grav}_{Gjt}$. Substituting $\Delta Trade_{Gjt}$ in equation (46) with $\Delta Trade^{grav}_{Gjt}$ provides an exogenous measure of regional import exposure based on the gravity approach as

$$T^{grav}_{it} = \sum_j \frac{L_{ijt}}{L_{jt}} \frac{\Delta Trade^{grav}_{Gjt}}{L_{it}} \tag{60}$$

We now substitute T_{it} from our baseline regression with T^{grav}_{it} directly. Otherwise, we run exactly the same specifications as before.

[59] Competitiveness increases due to productivity increases, better market access, and decreasing relative trade cost.

[60] As 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).

Table 8: Gravity Results for Effect of T_{it} on Voting

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

Results are reported in Table 8. 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 5. The key observation is that the positive effect of import exposure on extreme-right party votes is confirmed by this alternative identification strategy. In addition, a few other effects that were insignificant in table 5 become sharper here: there is more evidence for a positive effect of import 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 D Individual-Level Evidence

In this section, we test whether our regional-level results can be confirmed at the individual level. For each party P, we aggregate individuals' self-reported voting intentions into a decadal cumulative share of years in which a respondent stated he would vote for it. We calculate Y^P_{wt} 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.[61] For each party P, the dependent variable is a share between 0 and 1 for individual w in time period t and we separately estimate

$$Y^P_{wt} = \gamma^Y_{Y-1} \cdot Y^P_{wt-1} + \gamma^Y_T \cdot T_{it} + \gamma^Y_X \cdot X_{it-1} + \epsilon_{wt}. \tag{61}$$

for each party outcome. With a slight abuse of notation, Y^P_{wt-1} 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 5. Our focus is on estimating γ^Y_T, the effect of region i's import exposure T_{it} on a resident worker w's reported party support. Table 9 reports the results. Across rows it mimics closely our main table 5, except that there is no turnout measure in the SOEP. Every coefficient in table 9 reports the estimate of γ^Y_T from a separate regression. T_{it} is always instrumented as before, we

[61] 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. 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.

do not report the first stage regressions again. Column 1 includes the period-region fixed effects and the regional economic controls from table 5. We also add region i's base-year socio-economic and voting controls X_{it-1} from table 5 for each period. To better gauge magnitudes, column 2 reports the same specification with standardized outcomes. Import exposure shifts individuals' preferences to the extreme right, though the effect is weaker in the SOEP than it was in the actual voting data in table 5, 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 spectrum shows a response that is close to being significant. We also find distinctive results on what types of workers are driving these patterns. 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.[62] 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.[63] 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 most likely to experience adverse labor market effects from trade are the ones most likely to turn towards the extreme right because of increasing import exposure in their region. By contrast, the reduction in the change in the SPD's vote share is driven by low-skill *non*-manufacturing workers. A possible 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-initiated labor market reforms.

[62]In an earlier working paper, we focused on comparing the effect of individuals' import exposure due to their industry of employment relative to their regions' import exposure (Dippel et al., 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 import exposure.

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

Table 9: Individual-Level Analysis

	(1) All Controls	(2) Standardized	(3) High-Skill	(4) Low-Skill & Manuf.	(5) Low-Skill & Not Manuf.
Established Parties:					
Would Vote CDU/CSU	0.001	0.003	-0.007	-0.013	0.008
	(0.292)	(0.292)	(0.278)	(0.794)	(0.827)
Would Vote SPD	-0.008*	-0.016*	-0.013	-0.011	-0.017*
	(1.901)	(1.901)	(0.460)	(0.400)	(1.930)
Would Vote FDP	0.001	0.005	-0.018	0.011	0.007
	(0.459)	(0.459)	(0.420)	(0.664)	(0.568)
Would Vote Green Party	0.003	0.012	0.070	0.025	0.002
	(1.000)	(1.000)	(1.474)	(0.909)	(0.152)
Non-Established Parties:					
Would Vote Extreme-Right Parties	0.003	0.023	0.010	0.083**	0.006
	(1.619)	(1.619)	(0.875)	(2.206)	(0.475)
Would Vote Far-Left Parties	-0.001	-0.007	-0.051	0.019	-0.009
	(1.059)	(1.059)	(1.358)	(1.356)	(1.055)
Would Vote Other Small Parties	-0.001	-0.007	0.005	-0.026	-0.003
	(0.642)	(0.642)	(0.182)	(1.053)	(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 differnt 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 import 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 5, 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 import exposure. Column 4 shows that it is the population most affected by import 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 The Sequential Ignorability Assumption

A large literature on mediation analysis relies on the Sequential Ignorability Assumption A-3 of Imai et al. (2010) to identify mediation effects.

Assumption A-3 *Sequential Ignorability (Imai et al., 2010):*

$$\big(Y(t', m), M(t)\big) \perp\!\!\!\perp T|X \tag{62}$$
$$Y(t', m) \perp\!\!\!\perp M(t)|(T, X), \tag{63}$$

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 $x \in supp(X)$ and $m \in supp(M)$.

Under Sequential Ignorability A-3, 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:

$$ADE(t) = \int \Bigg(\big(E(Y|T = t_1, M = m, X = x) - E(Y|T = t_0, M = m, X = x, X = x)\big) dF_{M|T=t,X=x}(m) \Bigg) dF_X(x) \tag{64}$$

$$AIE(t) = \int \Bigg(E(Y|T = t, M = m, X = x) \Big[dF_{M|T=t_1,X=x}(m) - dF_{M|T=t_0,X=x}(m) \Big] \Bigg) dF_X(x). \tag{65}$$

Imai, Tingley, Keele and Yamamoto offer a substantial line of research that explores the identifying properties of Sequential Ignorability Assumption A-3. See Imai, Keele, Tingley, and Yamamoto (2011b) for a comprehensive discussion of the benefits and limitations of the sequential ignorability assumption.

The main critics of Sequential Ignorability A-3 is that it does not hold under the presence of either *Confounders* or *Unobserved Mediators* (Heckman and Pinto, 2015b).

The independence relation (62) 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-3 holds for the model defined in (**??**) because:

$$(\epsilon_Y, \epsilon_M) \perp\!\!\!\perp \epsilon_T \Rightarrow \quad \big(f_Y(t', m, \epsilon_Y), f_M(t, \epsilon_M)\big) \perp\!\!\!\perp f_T(\epsilon_T) \Rightarrow \quad \big(Y(t', m), M(t)\big) \perp\!\!\!\perp T. \tag{66}$$
$$\epsilon_Y \perp\!\!\!\perp \epsilon_M|\epsilon_T \Rightarrow \quad f_Y(t', m, \epsilon_Y) \perp\!\!\!\perp f_M(t, \epsilon_M)|f_T(\epsilon_T) \Rightarrow \quad Y(t', m) \perp\!\!\!\perp M(t)|T, \tag{67}$$

where the initial independence relation in (66) and (67) 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 (63) assumes that M is exogenous conditioned on X and T. It assumes that no confounding variable causing M and Y. Sequential Ignorability A-3 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.

Sequential Ignorability A-3 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:

$$\begin{aligned} ADE(t') &= \Big(\lambda_{YT} \cdot t_1 + \lambda_{YM} \cdot E(M(t'))\Big) - \Big(\lambda_{YT} \cdot t_0 + \lambda_{YM} \cdot E(M(t'))\Big) \\ \therefore ADE &= \lambda_{YT} \cdot \Delta(t) \qquad (68) \\ \text{and } AIE(t') &= \Big(\lambda_{YT} \cdot t' + \lambda_{YM} \cdot E(M(t_1))\Big) - \Big(\lambda_{YT} \cdot t' + \lambda_{YM} \cdot E(M(t_0))\Big) \\ &= \Big(\lambda_{YT} \cdot t' + \lambda_{YM}\lambda_M \cdot t_1\Big) - \Big(\lambda_{YT} \cdot t' + \lambda_{YM}\lambda_M \cdot t_0\Big) \\ \therefore AIE &= \lambda_{YM} \cdot \lambda_M \cdot \Delta(t) \qquad (69) \end{aligned}$$

Online Appendix B Identification of Causal Parameters

When we additionally allow for an unobserved mediator U that is caused by T and causes both M and Y (see **Remark** *2.1*), the linear mediation model we investigate can be fully described by the following equations:

$$\begin{aligned}
\text{Instrumental Variable } Z &= \epsilon_Z, && (70)\\
\text{Treatment } T &= \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T, && (71)\\
\text{Unobserved Mediator } U &= \zeta_T \cdot T + \epsilon_U, && (72)\\
\text{Observed Mediator } M &= \varphi_T \cdot T + \varphi_U \cdot U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, && (73)\\
\text{Outcome } Y &= \beta_T \cdot T + \beta_M \cdot M + \beta_U \cdot U + \beta_V \cdot V_Y + \epsilon_Y, && (74)\\
\text{Exogenous Variables} \quad & Z, V_T, V_M, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y \text{ are statistically independent variables,} && (75)\\
\text{Scalar Coefficients} \quad & \xi_Z, \xi_V, \zeta_T, \varphi_T, \varphi_U, \delta_Y, \delta_T, \beta_T, \beta_M, \beta_U, \beta_V && (76)\\
\text{Unobserved Variables} \quad & V_T, V_M, U, \epsilon_Z, \epsilon_T, \epsilon_U, \epsilon_M, \epsilon_Y. && (77)
\end{aligned}$$

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.

We first eliminate the unobserved mediator U from Equations (73)–(74) by iterated substitution. Equations (74)–(74) are then expressed as:

$$M = (\varphi_T + \varphi_U \zeta_T) \cdot T + \varphi_U \cdot \epsilon_U + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \tag{78}$$
$$Y = (\beta_T + \beta_U \zeta_T) \cdot T + \beta_M \cdot M + \beta_U \cdot \epsilon_U + \beta_V \cdot V_Y + \epsilon_Y. \tag{79}$$

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

$$\widetilde{\varphi}_T = \varphi_T + \varphi_U \zeta_T, \tag{80}$$
$$\widetilde{\beta}_T = \beta_T + \beta_U \zeta_T, \tag{81}$$
$$\widetilde{U} = \epsilon_U. \tag{82}$$

We use equations (78)–(82) to simplify Model (70)–(74) into the following equations:

$$\begin{aligned}
\text{Instrumental Variable } Z &= \epsilon_Z, && (83)\\
\text{Treatment } T &= \xi_Z \cdot Z + \xi_V \cdot V_T + \epsilon_T, && (84)\\
\text{Observed Mediator } M &= \widetilde{\varphi}_T \cdot T + \varphi_U \cdot \widetilde{U} + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, && (85)\\
\text{Outcome } Y &= \widetilde{\beta}_T \cdot T + \beta_M \cdot M + \beta_U \cdot \widetilde{U} + \beta_V \cdot V_Y + \epsilon_Y. && (86)
\end{aligned}$$

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

$$M(t) = \widetilde{\varphi}_T \cdot t + \varphi_U \cdot \widetilde{U} + \delta_Y \cdot V_Y + \delta_T \cdot V_T + \epsilon_M, \tag{87}$$
$$Y(m) = \widetilde{\beta}_T \cdot T + \beta_M \cdot m + \beta_U \cdot \widetilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \tag{88}$$
$$Y(t,m) = \widetilde{\beta}_T \cdot t + \beta_M \cdot m + \beta_U \cdot \widetilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \tag{89}$$
$$\begin{aligned} Y(t) &= \widetilde{\beta}_T \cdot t + \beta_M \cdot M(t) + \beta_U \cdot \widetilde{U} + \beta_V \cdot V_Y + \epsilon_Y. \\ &= (\widetilde{\beta}_T + \beta_M \widetilde{\varphi}_T) \cdot t + (\beta_U + \beta_M \varphi_U) \cdot \widetilde{U} + (\beta_V + \beta_M \delta_Y) \cdot V_Y + \beta_M \delta_T \cdot V_T + \beta_M \cdot \epsilon_M + \epsilon_Y. \end{aligned} \tag{90}$$

We claim that the coefficients associated with unobserved variables $V_T, \widetilde{U}, V_Y$ may only be identified up a linear transformation. Consider the coefficients δ_T, β_V that multiply the unobserved variable V_T in Equations (84) and (85) 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 $V_T, \widetilde{U}, V_Y$ has unit variance:

$$\text{var}(V_T) = \text{var}(\widetilde{U}) = \text{var}(V_Y) = 1. \tag{91}$$

Assumption (91) 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 B.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 (87)–(90), these causal effects are given by:

$$\text{Total Effect of } T \text{ on } Y: \frac{dE(Y(t))}{dt} = \widetilde{\varphi}_T \cdot \beta_M + \widetilde{\beta}_T. \tag{92}$$
$$\text{Direct Effect of } T \text{ on } Y: \frac{\partial E(Y(t,m))}{\partial t} = \widetilde{\beta}_T. \tag{93}$$
$$\text{Effect of } M \text{ on } Y: \frac{dE(Y(m))}{dm} = \beta_M. \tag{94}$$
$$\text{Effect of } T \text{ on } M: \frac{dE(M(t))}{dt} = \widetilde{\varphi}_T. \tag{95}$$
$$\text{Indirect Effect of } T \text{ on } Y: \frac{\partial E(Y(t,m))}{\partial m} \cdot \frac{dE(M(t))}{dt} = \beta_M \cdot \widetilde{\varphi}_T. \tag{96}$$

Online Appendix B.2 Identifying Equations

Model (83)–(86) can be conveniently expressed in matrix notation. In Equation (97) we define $\mathbf{X} = [Z, T, M, Y]'$ as the vector of observed variables, $\mathbf{V} = [V_T, V_Y, \widetilde{U}]'$ 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 (75), the random vectors $\mathbf{V}$ and $\boldsymbol{\epsilon}$ are independent, that is, $\mathbf{V} \perp\!\!\!\perp \boldsymbol{\epsilon}$. We use $\mathbf{K}$ in (97) for the matrix of parameters that multiply $\mathbf{X}$ and $\mathbf{A}$ for the matrix of parameters that multiply $\mathbf{V}$.

$$\mathbf{X} = \begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} V_T \\ V_Y \\ \widetilde{U} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}, \quad \mathbf{K} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \widetilde{\varphi}_T & 0 & 0 \\ 0 & \widetilde{\beta}_T & \beta_M & 0 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ \xi_V & 0 & 0 \\ \delta_Y & \delta_Y & \varphi_U \\ 0 & \beta_V & \beta_U \end{bmatrix}. \tag{97}$$

Using the notation in (97), we can express the linear system (83)–(86) as following:

$$\underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ \xi_Z & 0 & 0 & 0 \\ 0 & \widetilde{\varphi}_T & 0 & 0 \\ 0 & \widetilde{\beta}_T & \beta_M & 0 \end{bmatrix}}_{\mathbf{K}} \cdot \underbrace{\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ \xi_V & 0 & 0 \\ \delta_T & \delta_Y & \varphi_U \\ 0 & \beta_V & \beta_U \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{pmatrix} V_T \\ V_Y \\ \widetilde{U} \end{pmatrix}}_{\mathbf{V}} + \underbrace{\begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix}}_{\boldsymbol{\epsilon}}, \tag{98}$$

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon}. \tag{99}$$

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 (91), 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 (102) determines the relation between the covariance matrices of observed and unobserved variables:

$$\mathbf{X} = \mathbf{K} \cdot \mathbf{X} + \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon} \Rightarrow (\mathbf{K} - \mathbf{I})\, \mathbf{X} = \mathbf{A} \cdot \mathbf{V} + \boldsymbol{\epsilon}, \tag{100}$$

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

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

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 (102) 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 (102). The remaining six equalities from the off-diagonal relation of the covariance matrices in (102).

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 (102) 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 (102). Those are listed below:

$$\mathrm{cov}(Z,T) - \mathrm{cov}(Z,Z)\cdot \xi_Z = 0 \tag{103}$$
$$\mathrm{cov}(Z,M) - \mathrm{cov}(Z,T)\cdot \widetilde{\varphi}_T = 0 \tag{104}$$
$$\mathrm{cov}(Z,Y) - \mathrm{cov}(Z,M)\cdot \beta_M - \mathrm{cov}(Z,T)\cdot \widetilde{\beta}_T = 0 \tag{105}$$
$$\mathrm{cov}(T,Y) - \mathrm{cov}(T,T)\cdot \widetilde{\beta}_T - \mathrm{cov}(T,M)\cdot \beta_M = 0 \tag{106}$$
$$\mathrm{cov}(M,Y) - \mathrm{cov}(T,M)\cdot \widetilde{\beta}_T - \mathrm{cov}(M,M)\cdot \beta_M = \beta_U\cdot\varphi_U + \beta_V\cdot\delta_Y \tag{107}$$
$$\mathrm{cov}(T,M) - \mathrm{cov}(T,T)\cdot \widetilde{\varphi}_T = \delta_T\cdot\xi_V \tag{108}$$

Simple manipulation of Equations (103)–(108) generate the identification of the following parameters:

$$\xi_Z = \frac{\mathrm{cov}(Z,T)}{\mathrm{cov}(Z,Z)} \qquad \text{from Eq.(103)} \tag{109}$$
$$\widetilde{\varphi}_T = \frac{\mathrm{cov}(Z,M)}{\mathrm{cov}(Z,T)} \qquad \text{from Eq.(104)} \tag{110}$$
$$\beta_M = \frac{\mathrm{cov}(Z,T)\,\mathrm{cov}(T,Y) - \mathrm{cov}(T,T)\,\mathrm{cov}(Z,Y)}{\mathrm{cov}(T,M)\,\mathrm{cov}(Z,T) - \mathrm{cov}(T,T)\,\mathrm{cov}(Z,M)} \qquad \text{from Eqs.(105)–(106)} \tag{111}$$
$$\widetilde{\beta}_T = \frac{\mathrm{cov}(Z,M)\,\mathrm{cov}(T,Y) - \mathrm{cov}(Z,Y)\,\mathrm{cov}(T,M)}{\mathrm{cov}(T,T)\,\mathrm{cov}(Z,M) - \mathrm{cov}(Z,T)\,\mathrm{cov}(T,M)} \qquad \text{from Eqs.(105)–(106)} \tag{112}$$
$$\beta_U\cdot\varphi_U + \beta_V\cdot\delta_Y = \mathrm{cov}(M,Y) - \mathrm{cov}(M,M)\cdot\beta_M - \mathrm{cov}(T,M)\cdot\widetilde{\beta}_T \qquad \text{from Eq.(107)} \tag{113}$$
$$\delta_T\cdot\xi_V = \frac{\mathrm{cov}(T,M)\,\mathrm{cov}(Z,M) - \mathrm{cov}(T,T)\,\mathrm{cov}(Z,Y)}{\mathrm{cov}(Z,M)} \qquad \text{from Eq.(108)} \tag{114}$$

Moreover, if we divide Equation (105) by $\mathrm{cov}(Z,T)$ we obtain:

$$\frac{\mathrm{cov}(Z,Y)}{\mathrm{cov}(Z,T)} - \frac{\mathrm{cov}(Z,M)}{\mathrm{cov}(Z,T)}\cdot\beta_M - \frac{\mathrm{cov}(Z,T)}{\mathrm{cov}(Z,T)}\cdot\widetilde{\beta}_T = 0 \tag{115}$$
$$\Rightarrow \frac{\mathrm{cov}(Z,Y)}{\mathrm{cov}(Z,T)} - \widetilde{\varphi}_T\cdot\beta_M - \widetilde{\beta}_T = 0 \tag{116}$$
$$\Rightarrow \widetilde{\varphi}_T\cdot\beta_M + \widetilde{\beta}_T = \frac{\mathrm{cov}(Z,Y)}{\mathrm{cov}(Z,T)}. \tag{117}$$

The four causal of interest parameters defined in (92)–(95) are respectively identified by Equations (110), (111), (112) and (117):

$$\frac{dE(M(t))}{dt} = \widetilde{\varphi}_T = \frac{\mathrm{cov}(Z,M)}{\mathrm{cov}(Z,T)}, \tag{118}$$
$$\frac{dE(Y(m))}{dm} = \beta_M = \frac{\mathrm{cov}(Z,Y)\,\mathrm{cov}(T,T) - \mathrm{cov}(Y,T)\,\mathrm{cov}(Z,T)}{\mathrm{cov}(Z,M)\,\mathrm{cov}(T,T) - \mathrm{cov}(M,T)\,\mathrm{cov}(Z,T)}, \tag{119}$$
$$\frac{\partial E(Y(t,m))}{\partial t} = \widetilde{\beta}_T = \frac{\mathrm{cov}(Z,M)\,\mathrm{cov}(T,Y) - \mathrm{cov}(Z,Y)\,\mathrm{cov}(T,M)}{\mathrm{cov}(T,T)\,\mathrm{cov}(Z,M) - \mathrm{cov}(Z,T)\,\mathrm{cov}(T,M)}, \tag{120}$$
$$\frac{dE(Y(t))}{dt} = \widetilde{\varphi}_T\cdot\beta_M + \widetilde{\beta}_T = \frac{\mathrm{cov}(Z,Y)}{\mathrm{cov}(Z,T)}. \tag{121}$$

Next section explains that each causal effect (118)–(121) can be evaluated by standard Two-stage Least Squares regressions.

Online Appendix C 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 et al., 2010, 2011b,a).[64] In this case, treatment T is as good as randomly assigned and the resulting model is equivalent to assuming no confounding variables and no unobserved mediators U in *Model III* of Table 1.[65] 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.[66]

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) 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).[67] 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.[68] Because of its natural appeal, we discuss this approach here and contrast its identification requirements explicitly to ours. A standard mediation model with confounding variables V and two separate dedicated instrumental variables (for separate endogenous variables) is described as follows:

$$\text{Treatment variable: } T = f_T(Z_T, V, \epsilon_T), \tag{122}$$
$$\text{Observed Mediator: } M = f_M(T, Z_M, V, \epsilon_M), \tag{123}$$
$$\text{Outcome: } Y = f_Y(T, M, V, \epsilon_Y), \tag{124}$$
$$\text{where: } (Z_T, Z_M) \perp\!\!\!\perp V. \tag{125}$$

This model is presented as a DAG in Table 1. 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.[69]

The empirical challenge in evaluation Model (122)–(125) 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)*

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

[65] If 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).

[66] In 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.

[67] Both 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.

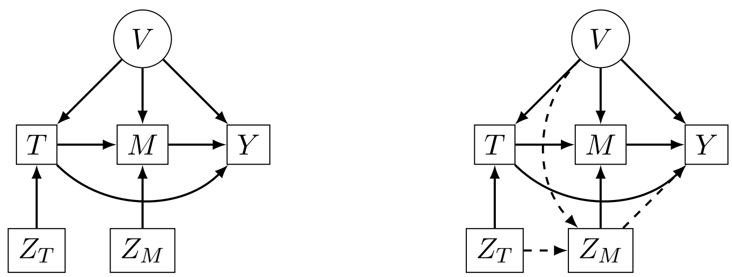
[68] Recently, Frolich and Huber (2017) provide an important contribution on the mediation model with two dedicated instruments.

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

Online Appendix Table 1: 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.

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 1. 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 (125) 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.

Online Appendix D Estimation of Causal Parameters

Our goal is to show that the four causal parameters listed in Equations (118)–(121) 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 (126)–(127) 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.

$$\text{First Stage: } T = \kappa_1 + \beta_1 \cdot Z + \epsilon_1, \tag{126}$$

$$\text{Second Stage: } Y = \kappa_2 + \beta_2 \cdot T + \epsilon_2. \tag{127}$$

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 :

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

Consider the inclusion of additional covariates X in both stages of the 2SLS method. Variables X in (129)–(130) play the role of control covariates in the first stage and second stages of the 2SLS estimator. Control covariates X directly causes Y in (130) while the instrument Z only causes Y though it impact on T.

$$\text{First Stage: } T = \kappa_1 + \beta_1 \cdot Z + \psi_1 \cdot X + \epsilon_1, \tag{129}$$

$$\text{Second Stage: } Y = \kappa_2 + \beta_2 \cdot T + \psi_2 \cdot X + \epsilon_2. \tag{130}$$

The 2SLS model (129)–(130) 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 (131) and it is a consistent estimator under model assumptions.

$$\text{plim}(\hat{\beta}_2) = \frac{\text{cov}(Z, Y)\,\text{cov}(X, X) - \text{cov}(Y, X)\,\text{cov}(Z, X)}{\text{cov}(Z, T)\,\text{cov}(X, X) - \text{cov}(T, X)\,\text{cov}(Z, X)} = \beta_2. \tag{131}$$

The 2SLS estimator $\hat{\psi}_2$ for parameter ψ_2 is expressed by Equation (132) and it is a consistent estimator under model assumptions.

$$\text{plim}(\hat{\psi}_2) = -\frac{\text{cov}(Z, Y)\,\text{cov}(T, X) - \text{cov}(Y, X)\,\text{cov}(Z, T)}{\text{cov}(Z, T)\,\text{cov}(X, X) - \text{cov}(T, X)\,\text{cov}(Z, X)} = \psi_2. \tag{132}$$

Each of the identification formulas for the causal effects in (118)–(121) describes a ratio of covariances that corresponds to one of the three 2SLS formulas (128), (131) or (131).

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

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

According to Equation (128), this effect can be estimated by the 2SLS regression (126)–(127) 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:

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

According to Equation (128),this effect can be estimated by the 2SLS regression (126)–(127) 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:

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

which can be estimated by the 2SLS regression (126)–(127) 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:

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

According to the 2SLS estimator in (131), this causal effect can be estimated by $\hat{\beta}_2$ in the 2SLS regression (129)–(130) 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:

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

According to the 2SLS estimator in (132), this causal effect can be estimated by $\hat{\psi}_2$ in the 2SLS regression (129)–(130) 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 E Total, Indirect and Direct Effects under One Instrument

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

$$\text{Total Effect of } T \text{ on } Y: \frac{dE(Y(t))}{dt} = \frac{\text{cov}(Z,Y)}{\text{cov}(Z,T)}. \tag{133}$$

$$\text{Direct Effect of } T \text{ on } Y: \frac{\partial E(Y(t,m))}{\partial t} = \frac{\text{cov}(Z,M)\,\text{cov}(T,Y) - \text{cov}(Z,Y)\,\text{cov}(T,M)}{\text{cov}(T,T)\,\text{cov}(Z,M) - \text{cov}(Z,T)\,\text{cov}(T,M)}. \tag{134}$$

$$\text{Effect of } M \text{ on } Y: \frac{\partial E(Y(t,m))}{\partial m} = \frac{\text{cov}(Z,T)\,\text{cov}(T,Y) - \text{cov}(T,T)\,\text{cov}(Z,Y)}{\text{cov}(T,M)\,\text{cov}(Z,T) - \text{cov}(T,T)\,\text{cov}(Z,M)} \tag{135}$$

$$\text{Effect of } T \text{ on } M: \frac{dE(M(t))}{dt} = \frac{\text{cov}(Z,M)}{\text{cov}(Z,T)} \tag{136}$$

$$\text{Indirect Effect of } T \text{ on } Y: \frac{\partial E(Y(t,m))}{\partial m} \cdot \frac{dE(M(t))}{dt}. \tag{137}$$

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 is stated as following:

$$\underbrace{\frac{dE(Y(t))}{dt}}_{\text{Total Effect}} = \underbrace{\frac{\partial E(Y(t,m))}{\partial t}}_{\text{Direct Effect}} + \underbrace{\frac{\partial E(Y(t,m))}{\partial m} \cdot \frac{dE(M(t))}{dt}}_{\text{Indirect Effect}}. \tag{138}$$

We show that the decomposition described in (138) 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 (133) is equal to the covariance ratio that identifies the direct effect in Equations (134) plus the multiplication of the covariance ratios that identify the effect of T on M in (136) and the effect of M on Y

described in Equation (135). We thank David Slichter for pointing out 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{\operatorname{cov}(Z,M)\operatorname{cov}(T,Y) - \operatorname{cov}(Z,Y)\operatorname{cov}(T,M)}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)}}_{\frac{\partial E(Y(t,m))}{\partial t}} + \underbrace{\frac{\operatorname{cov}(Z,T)\operatorname{cov}(T,Y) - \operatorname{cov}(T,T)\operatorname{cov}(Z,Y)}{\operatorname{cov}(T,M)\operatorname{cov}(Z,T) - \operatorname{cov}(T,T)\operatorname{cov}(Z,M)}}_{\frac{\partial E(Y(t,m))}{\partial m}} \cdot \underbrace{\frac{\operatorname{cov}(Z,M)}{\operatorname{cov}(Z,T)}}_{\frac{dE(M(t))}{dt}} \\
&= \frac{\operatorname{cov}(Z,M)\operatorname{cov}(T,Y) - \operatorname{cov}(Z,Y)\operatorname{cov}(T,M)}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)} + \frac{\operatorname{cov}(Z,M)\operatorname{cov}(T,Y) - \operatorname{cov}(T,T)\operatorname{cov}(Z,Y)\frac{\operatorname{cov}(Z,M)}{\operatorname{cov}(Z,T)}}{\operatorname{cov}(T,M)\operatorname{cov}(Z,T) - \operatorname{cov}(T,T)\operatorname{cov}(Z,M)} \\
&= \frac{\operatorname{cov}(Z,M)\operatorname{cov}(T,Y) - \operatorname{cov}(Z,Y)\operatorname{cov}(T,M)}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)} + \frac{\operatorname{cov}(T,T)\operatorname{cov}(Z,Y)\frac{\operatorname{cov}(Z,M)}{\operatorname{cov}(Z,T)} - \operatorname{cov}(Z,M)\operatorname{cov}(T,Y)}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)} \\
&= \frac{\operatorname{cov}(T,T)\operatorname{cov}(Z,Y)\frac{\operatorname{cov}(Z,M)}{\operatorname{cov}(Z,T)} - \operatorname{cov}(Z,Y)\operatorname{cov}(T,M)}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)} \\
&= \frac{\operatorname{cov}(T,T)\operatorname{cov}(Z,M)\frac{\operatorname{cov}(Z,Y)}{\operatorname{cov}(Z,T)} - \operatorname{cov}(Z,Y)\operatorname{cov}(T,M)}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)} \\
&= \frac{\operatorname{cov}(T,T)\operatorname{cov}(Z,M)\frac{\operatorname{cov}(Z,Y)}{\operatorname{cov}(Z,T)} - \operatorname{cov}(Z,Y)\operatorname{cov}(T,M)\frac{\operatorname{cov}(Z,T)}{\operatorname{cov}(Z,T)}}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)} \\
&= \frac{\operatorname{cov}(T,T)\operatorname{cov}(Z,M)\frac{\operatorname{cov}(Z,Y)}{\operatorname{cov}(Z,T)} - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)\frac{\operatorname{cov}(Z,Y)}{\operatorname{cov}(Z,T)}}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)} \\
&= \left(\frac{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)}{\operatorname{cov}(T,T)\operatorname{cov}(Z,M) - \operatorname{cov}(Z,T)\operatorname{cov}(T,M)}\right) \cdot \left(\frac{\operatorname{cov}(Z,Y)}{\operatorname{cov}(Z,T)}\right) \\
&= \frac{\operatorname{cov}(Z,Y)}{\operatorname{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 (134), (135) and (133). The third equation isolates and eliminates the common term $\operatorname{cov}(Z,M)$ 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 $\operatorname{cov}(Z,M)$ and $\operatorname{cov}(Z,Y)$ of the first term of the numerator. The seventh equation includes the term $\frac{\operatorname{cov}(Z,T)}{\operatorname{cov}(Z,T)}$ which is equal to one. The eight equation exchange the covariances $\operatorname{cov}(Z,Y)$ and $\operatorname{cov}(Z,T)$ of the second term of the numerator. The night 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 $\frac{\operatorname{cov}(Z,Y)}{\operatorname{cov}(Z,T)}$ which, according to (133), is equal to the total effect of choice T on outcome Y.

Online Appendix F Data Sources

Online Appendix F.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 form 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 friinge 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 F.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 F.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 import 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

[70] See 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 G Background on German Politics 1987 to 2009

Online Appendix G.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 G.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 German*y (KPD); Leninist, Stalinist, and Maoist

[71] In 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.

[72] Voters 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.

[73] In Online Appendix G.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 G.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 G.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

[74] To 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.

[75] The 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 G.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 infratructure 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 in combination with subsequent economic hardship, provided a fertile ground for extreme-right parties.

Online Appendix H Descriptive Statistics

Table 2 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 import 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, import 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.

[76] Note 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

[78] A 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`

[79] The 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 G for more details.

Online Appendix Table 2: The Core Variables in 1987–1998 and in 1998–2009

		(1)	(2)	(3)	(4)	(5)	(6)
		Period 1 (1987-1998),			Period 2 (1998-2009),		
percentile:		25th	median	75th	25th	median	75th
$\boldsymbol{T_{it}}$		-0.264	0.068	0.521	-1.222	-0.663	-0.144
$\widehat{\boldsymbol{T}}_{\boldsymbol{it}}$ (instrumented with Z_{it})		-0.068	0.143	0.402	-1.150	-0.574	-0.113
$\mathbf{Y_{it}}$:	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
$\mathbf{M_{it}}$:	Δ log(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 Table 3: OLS Version of Table 4

	(1) Baseline OLS	(2) + Industry OLS	(3) +Socio OLS	(4) + Voting OLS
Δ log(Total Employment)	-0.013***	-0.011**	-0.009**	-0.009*
	(-3.138)	(-2.514)	(-2.070)	(-1.919)
Δ Share Manufacturing Employment	-0.502***	-0.524***	-0.496***	-0.502***
	(-3.348)	(-3.486)	(-3.289)	(-3.362)
Δ log(Mean Manufacturing Wage)	-0.003**	-0.004**	-0.003**	-0.003**
	(-2.122)	(-2.262)	(-2.094)	(-2.152)
Δ log(Mean Non-Manufacturing Wage)	-0.001	-0.001	-0.000	-0.000
	(-0.934)	(-0.853)	(-0.351)	(-0.433)
Δ Share Unemployment	0.089*	0.095	0.102*	0.125***
	(1.659)	(1.617)	(1.732)	(2.674)
Period-by-region FE	Yes	Yes	Yes	Yes
Observations	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 I Robustness and Further Results

Online Appendix I.1 Labor Market Outcomes

Online Appendix I table 3 presents the OLS results corresponding to the paper's table 4.

Online Appendix I.2 Additional Results on the Vote Share Table 5

Online Appendix I table 4 presents the OLS results corresponding to the paper's table 5.

Online Appendix Table 4: OLS Version of Table 5

	(1) Baseline OLS	(2) + Industry OLS	(3) +Socio OLS	(4) + Voting OLS	(5) Standard. OLS
Δ Turnout	0.004***	0.004***	0.003**	0.003**	0.040**
	(2.932)	(3.059)	(2.337)	(2.430)	(2.430)
Established Parties:					
Δ Vote Share CDU/CSU	-0.081	-0.113	-0.062	-0.067	-0.016
	(-1.015)	(-1.423)	(-0.963)	(-1.020)	(-1.020)
Δ Vote Share SPD	-0.037	-0.044	0.061	0.062	0.005
	(-0.416)	(-0.471)	(0.884)	(0.929)	(0.929)
Δ Vote Share FDP	0.094**	0.105**	0.081*	0.088**	0.016**
	(1.971)	(2.398)	(1.805)	(2.097)	(2.097)
Δ Vote Share Green Party	0.046	0.063*	0.062*	0.068**	0.024**
	(1.221)	(1.755)	(1.835)	(2.042)	(2.042)
Non-established Parties					
Δ Vote Share Extreme-Right Parties	0.038*	0.036	-0.009	-0.004	-0.002
	(1.703)	(1.522)	(-0.483)	(-0.240)	(-0.240)
Δ Vote Share Far-Left Parties	-0.108*	-0.109	-0.138**	-0.153**	-0.039**
	(-1.669)	(-1.597)	(-2.159)	(-2.491)	(-2.491)
Δ Vote Share Other Small Parties	0.048	0.062**	0.003	0.007	0.005
	(1.586)	(2.186)	(0.138)	(0.259)	(0.259)
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.

Online Appendix J Constructing Gravity Residuals

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

$$log(EX_{djt}^{CE-O}) - log(EX_{djt}^{G-O}) = \alpha_d + \alpha_j + \epsilon_{djt}^{IM}, \quad (139)$$

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_{Git}^{grav} for ΔEX_{Gjt}. It can be obtained from the residuals of the regression

$$log(EX_{djt}^{G-CE}) - log(EX_{djt}^{G-O}) = \alpha_d + \alpha_j + \epsilon_{djt}^{EX}, \quad (140)$$

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.

[80]Since 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 K Subsample Results for the Effect of Trade T on Labor M and Voting Y

Column 1 of Online Appendix K table 5 reports on total employment (as in the paper's table 4), 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 K table 5 similarly decompose the same eight political outcomes as reported in table 5. The sample sizes are 322, 408, and 322 respectively.

The effect of trade shocks on labor markets should be more pronounced in the second period, when German labor markers were more flexible. We found some evidence for this pattern in the individual results in table 9. This motivates us to decompose the effect of import exposure on local labor markets by period in this section.

Online Appendix Table 5: Decomposing the Results by Period

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

Notes: The table reports subsample estimations. Column 1 reports on the log of total employment, as in table 4. Columns 2–9 report on the same eight political outcomes as in table 5. 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 (47). Every panel additional 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.

Table 5 reports on the manufacturing 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 import 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 9 by period. We therefore report the

Online Appendix Table 6: Individual-Level Analysis

	(1) 1990-1998	(2) 1998-2009	(3) Low-Skill & Manuf., 1998-2009	(4) Low-Skill & Not Manuf., 1998-2009
Established Parties:				
Would Vote CDU/CSU	-0.025	0.002	-0.006	0.003
	(-0.743)	(0.227)	(-0.350)	(0.257)
Would Vote SPD	0.027	-0.019**	0.001	-0.022**
	(0.761)	(-2.217)	(0.031)	(-2.352)
Would Vote FDP	-0.038	0.015	0.002	0.021
	(-0.720)	(1.177)	(0.116)	(1.431)
Would Vote Green Party	0.019	0.016	0.007	0.007
	(0.409)	(1.295)	(0.363)	(0.565)
Non-Established Parties:				
Would Vote Extreme-Right Parties	0.029	0.028*	0.088**	0.016
	(0.735)	(1.802)	(2.013)	(1.035)
Would Vote Far-Left Parties	-0.008	-0.005	0.026	-0.010
	(-0.670)	(-0.751)	(1.579)	(-1.043)
Would Vote Other Small Parties	0.018	-0.012	-0.048*	-0.001
	(0.340)	(-1.072)	(-1.674)	(-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 9 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 import 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 9. (*b*) Standard errors are clustered at the region level. *** p<0.01, ** p<0.05, * p<0.1.

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.