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

Christian Dippel† Robert Gold‡ Stephan Heblich§ Rodrigo Pinto¶

July 28, 2018

Abstract

Instrumental variables (IV) are commonly used to identify treatment effects, but standard IV estimation cannot unpack the complex treatment effects that arise when a treatment and its outcome together cause a second outcome of interest. For example, IV estimations have been used to show that import exposure to low-wage countries has adversely affected high-wage countries' labor markets. They have also been used to show that such import exposure has polarized voters. However, they cannot answer the question to what extent the latter is a consequence of the former. We propose a new identification strategy that allows us to do so, appealing to one additional identifying assumption and requiring no additional instruments. Applying this framework, we estimate that across specifications labor market adjustments explain virtually all of the effect of import exposure on voting. This enables us to provide rigorous evidence that the correct policy response to voter polarization has to be focused on labor markets.

Keywords: Instrumental Variables, Causal Mediation Analysis, Import Exposure, Voting, Local Labor Markets

JEL Codes: C36, D72, F6, J2

*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, LMU, 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.

†University of California, Los Angeles, CCPR, and NBER.

‡IfW - Kiel Institute for the World Economy and CESifo.

§University of Bristol, CESifo, IZA, and SERC.

¶University of California, Los Angeles, CCPR, and NBER.

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 estimation, however, is 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 toward political populism because such exposure adversely affected labor markets. Unpacking this causal mechanism is important to guide policy-makers in designing effective responses to political populism.

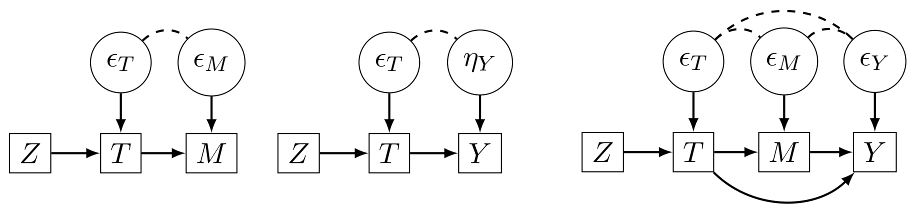
From an econometric perspective, we investigate the problem of identifying causal relations when an endogenous treatment and its outcome together cause a second outcome of interest. We propose a solution to the problem that does not require additional instrumental variables and can be easily implemented using the well-known two-stage least squares (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 and Kovak (2017) 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 non-parametric 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" (which is to say, through other channels). 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 does not identify 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 created anxiety about the future, this may by itself turn voters toward populism (Mughan and Lacy, 2002; Mughan,

[2]We base our analysis on the IV strategy proposed by Autor et al. (2013), which instruments for one high-wage country's trade exposure to China with other high-wage countries' industry-specific trade exposure to China.

[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).

Bean and McAllister, 2003).[4] On the other hand, import exposure may be politically moderating if it lowered consumer prices, if targeted government transfers like *Trade Adjustment Assistance* increased, if manufacturers shifted production toward more differentiated higher-markup output varieties (as in Holmes and Stevens, 2014), or if it lead 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 were negative (i.e. the direct effect has the opposite sign of the total effect) the effect of import exposure on voting that is mediated by labor market adjustments would 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. Equations $M = f_M(T, \epsilon_M)$ and $Y = f_Y(T, M, \epsilon_Y)$ imply that T causes Y indirectly through M as well as directly, i.e. through other channels (that are graphically represented by the arrow directly linking T to Y). 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 (2011). 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; Frölich 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

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

[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 (2010a).

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 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 the effect of M on Y conditional on T.[6] We also develop a procedure to bound the possible range of the direct and the indirect effects linking T and Y when the identifying assumption of our framework is relaxed.

We apply our method to data on regional import exposure, labor market adjustments, and voting patterns in Germany from 1987 to 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 multiparty 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 as opposed to by district, so local voting patterns are not confounded by local variation in political messaging.[7] (*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 among attitudinal socioeconomic surveys, the German *Socio-Economic Panel*'s (SOEP) long-running panel structure allows us to cross-validate the aggregate results with an individual-level panel-analysis, relating decadal

[6] Our main focus lies on estimating 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,

$$Y = \delta_Y^T \cdot T + \eta_Y, \quad M = \beta_M^T \cdot T + \epsilon_M, \qquad \text{and} \quad Y = \beta_Y^T \cdot T + \beta_Y^M \cdot M + \epsilon_Y,$$

and then compare the total effect δ_Y^T to the indirect effect $\beta_Y^M \cdot \beta_M^T$. See Baron and Kenny (1986) and MacKinnon (2008) for an overview, as well as the discussion in our section 2.1.

[7] This vote, called *Zweitstimme*, mainly determines a party's share of parliamentary seats.

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

changes in individual workers' stated party preferences to changes in their local labor market's 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), and instrument T with a measure based on other high-wage countries' sector-specific trade flows (Z). 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. 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.[9] There are no significant effects on turnout, or on any of the mainstream parties, the far left, or other small parties.[10] 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.[11]

Next, we estimate *Model III*. The effect of trade-induced labor market adjustments on voting is at least as large as the total effect of import exposure on voting, suggesting that other channels though which import exposure affects voting are moderating the total effect. When we relax the identifying assumptions and derive bounds on the indirect effect, we find a lower bound of 70 percent. This, even at the lower bound, labor market adjustments appear to be the primary reason for the populist backlash against import exposure.

Our paper's contribution is twofold: It addresses 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 methodological side, we offer a mediation model that relies on a single instrumental variable Z that directly causes T to identify three causal effects, while allowing for endogenous variables caused by confounders

[9] 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).

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

[11] 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 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. 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.

The rest of the paper proceeds as follows: 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.

2 Mediation with an Instrumental Variable

In the following, section 2.1 lays out the definition and identification of causal effects in a mediation model, i.e. a model where a treatment T and its outcome M jointly cause another outcome Y. In section 2.2, we present the exclusion restrictions needed to obtain identification of all causal effects using an instrument Z for T. These include the standard IV assumptions needed to identify *Model I* and *Model II*, as well as the novel exclusion restriction needed to unpack *Model III*. Under the assumption of linearity, section 2.3 derives an easily operationalized estimation procedure that relies on a series of 2SLS estimations. Finally, in section 2.4 we derive a complementary bounding exercise that relaxes our identifying assumption.

2.1 Causal Effects in the Mediation Model

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). In a mediation model, the total effect (TE) of T on Y can be decomposed as the sum of the 'indirect effect' (IE) of T on Y that is mediated by M and the 'direct effect' (DE) of T on Y that is not mediated by M. Let $\epsilon_T, \epsilon_M, \epsilon_Y$ be the unobserved error terms associated with variables T, M, Y, respectively.

A general nonparametric model that portrays these causal relations is then given by

$$T = f_T(\epsilon_T), \tag{1}$$
$$M = f_M(T, \epsilon_M), \tag{2}$$
$$Y = f_Y(T, M, \epsilon_Y). \tag{3}$$

Causal effects are defined by the difference between counterfactual variables.[12] For instance, let $M(t)$ be the counterfactual variable M when T takes the value t. The causal relation in equation (2) yields the counterfactual variable $M(t)$; while causal relation (3) yields the counterfactual variables $Y(t)$, $Y(m)$, and $Y(m,t)$ as defined below:[13]

$$M(t) = f_M(t, \epsilon_M), \tag{4}$$
$$Y(t) = f_Y(t, M(t), \epsilon_Y), \tag{5}$$
$$Y(m) = f_Y(T, m, \epsilon_Y), \tag{6}$$
$$Y(m,t) = f_Y(t, m, \epsilon_Y). \tag{7}$$

The causal effect of T on M when T takes values t_1, t_0 is given by the expected value of the difference $E(M(t_1) - M(t_0))$. The three main causal effects of interest in a mediation setting (i.e. the TE, DE, and IE) are defined as follows:[14]

$$TE = E(Y(t_1) - Y(t_0)) \equiv \int E(Y(t_1, M(t_1)) - Y(t_0, M(t_0))) \tag{8}$$
$$DE(t) = E(Y(t_1, M(t)) - Y(t_0, M(t))) \equiv \int E\Big(Y(t_1, m) - Y(t_0, m)\Big) dF_{M(t)}(m) \tag{9}$$
$$IE(t) = E(Y(t, M(t_1)) - Y(t, M(t_0))) \equiv \int E\Big(Y(t, m)\Big)\Big[dF_{M(t_1)}(m) - dF_{M(t_0)}(m)\Big] \tag{10}$$

The seminal work of Robins and Greenland (1992) showed that under the assumption of the mutual independence of the error terms, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_M$, $\epsilon_M \perp\!\!\!\perp \epsilon_Y$, $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, the TE, IE, and DE can be causally identified in equations (1)–(3). See Online Appendix A. Under linearity, this

[12] Causal effects and counterfactual outcomes, as well as exclusion restrictions are defined non-parametrically. As is standard, we therefore use a non-parametric model when discussing causality in sections 2.1 and 2.2. Counterfactual variables are defined by *fixing* an argument of a structural equation to a value. The counterfactual variable $M(t)$, for instance, is defined by *fixing* the T-input of this structural equation to the value t, namely, $M(t) = f_M(t, \epsilon_M)$. See Heckman and Pinto (2015a) for a discussion.

[13] Counterfactual variable $Y(t)$ in (5) denotes the potential outcome Y when T takes the value t, $Y(m)$ is the counterfactual outcome when M is fixed to the value m, and $Y(m,t)$ is the counterfactual outcome that arises when both T and M are fixed to t and m, respectively.

[14] Pearl (2011) makes a distinction between controlled (or "prescriptive") and natural (or "descriptive") effects. He uses the terms *controlled* direct effect and *controlled* indirect effect for equations (9) and (10) respectively. In the linear case, which will be our focus, this distinction disappears. See also footnote 21.

amounts to the series of three OLS regressions in footnote 6.[15]

The assumption of the mutual independence among all error terms is clearly a strong one: Assuming that the error terms ϵ_T and ϵ_M are statistically independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_M$, implies that there are no unobserved variables that jointly cause T and M. In this case, T is exogenous with respect to M, i.e. we have random assignment as in a randomized control trial (RCT), a condition that is rarely satisfied in observational data. For example, in our empirical application, the literature's concern that unobserved adverse regional demand shocks reduce regional imports (T) as well as employment (M) implies $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, which precludes the independence between T and counterfactuals $M(t)$ and $Y(t)$ in equations (4), (5), i.e. $T \not\perp\!\!\!\perp \big(Y(t), M(t)\big)$. Hence, T is endogenous and the causal effect of T on the outcomes M and Y cannot be identified on the basis of their observed distributions. This is the identification challenge that *Model I* and *Model II* in Table 1 solve using IV.[16]

Relatedly, assuming the mutual independence among all error renders both the DE and IE identified because it makes variables T and M exogenous with respect to Y (See Online Appendix A). This assumption does not follow even when treatment T is randomly assigned, as in an RCT. Drawing further on our empirical application to illustrate the pitfalls of this assumption, it is likely that $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$ because unobserved industry or worker characteristics that affect labor market outcomes may also affect political preferences. The statistical dependence $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$ precludes the independence between M and counterfactuals $\big(Y(m,t), Y(m)\big)$ in equations (6), (7), 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. This is the additional identification challenge that arises when trying to estimate *Model III* in Table 1, which our framework will allow us to solve by using an instrumental variable.

In summary, the mutual independence among all error terms ϵ_T, ϵ_M, and ϵ_Y is unlikely to hold in observational data. However, without this assumption causal effects are not identified on the basis of observed distributions. To gain identification, we modify the standard mediation

[15] A large literature on mediation analysis relies on the Sequential Ignorability Assumption **A-3** of Imai, Keele and Yamamoto (2010b) to identify mediation effects. We discuss this assumption in Online Appendix B, which is related to the assumption of the mutual independence of the error terms $\epsilon_T, \epsilon_M, \epsilon_Y$. See Frölich and Huber (2017) for a recent review of the mediation literature.

[16] *Model II* in Table 1 amounts to a direct estimate of the TE in equation (8). By contrast, *Model III* in Table 1, which makes the mediation setup explicit, estimates the TE as the sum of the DE and IE in equations (9)–(10).

model by adding an instrument Z that causes T, thus equation (1) becomes $T = f_T(Z, \epsilon_T)$. Stated more succinctly, we study identification of the mediation model in an IV setting. In section 2.2 we derive two well-known exclusion restrictions and one novel one, which together allow for the identification of all causal effects of interest while allowing for the dependence among all key error terms ϵ_T, ϵ_M, and ϵ_Y, and requiring an instrument only for T.

2.2 Exclusion Restrictions

In the following, section 2.2.1 discusses the well-known exclusion restrictions needed for identification of *Model I* and *Model II*; section 2.2.2 presents the novel exclusion restriction needed for the identification of *Model III*, without the use of additional instruments.

2.2.1 Exclusion Restriction to Identify *Model I* and *Model II* in Table 1

The standard property of an instrumental variable Z is that it affects outcomes M and Y only through its impact on treatment T. Put another way, for Z to be an instrument, it must be the case that Z is statistically independent of unobserved error terms ϵ_T, ϵ_M and ϵ_Y, which jointly cause T, M and Y. This property is of course well-established, but in the interest of clarity we state it again as Assumption **A-1**.[17]

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

Assumption **A-1** merely states the independence condition that characterizes Z as an instrumental variable for T in $M(t)$ as well as $Y(t)$. Lemma **L-1** states the well-established result that Assumption **A-1** generates the two exclusion restrictions needed to estimate *Model I* and *Model II*.[18]

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

Targeted Causal Relation	*IV Relevance (First Stage)*		*Exclusion Restriction*
$T \to Y$	$Z \not\perp\!\!\!\perp T$	*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.*

[17] 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 (35) and (36) for the precise definitions of T and Z in our empirical application.

[18] *Model I* stands for the standard IV model that evaluates the effect of T on M using Z as the instrument. *Model II* stands for the IV model that evaluates the total effect (TE) of T on Y.

The exclusion restrictions in **L-1** imply that Z is a valid instrument for T, and thus 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. 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 the exclusion restrictions. Indeed, the lemma would remain the same if the causal relation $M \to Y$ were reversed to $M \leftarrow Y$. This irrelevance of the causal direction between M and Y highlights the fact that identification of *Model I* and *Model II* does not lead to identification of *Model III*. In Section 2.2.2 we address this problem by invoking the error dependence structure discussed therein.

Remark *2.1* Exclusion restrictions such as **L-1** 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.[19]

2.2.2 Exclusion Restriction to Identify *Model III* in Table 1

Our interest is in settings where T and M jointly cause a second outcome of interest Y, as in equation (3). What we will exploit for identification is that in many such settings the omitted variable concerns themselves suggest a natural solution to the identification challenges involved. This is best illustrated in our empirical application, where 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. The critical observation is that the endogeneity concerns in the relation between import exposure and voting (Y) emanate primarily from the labor market channel. In other words, the main concern about unobservables in this relation is that regional demand shocks that affect import exposure also affect voting through their labor market effects. Using this intuition, the framework we propose hinges on allowing unobserved shocks that impact import exposure to also affect voter preferences through labor markets, but not through unrelated channels. This is made precise in Assumption **A-2**.

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

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

$$\epsilon_T \perp\!\!\!\perp \epsilon_Y, \textit{ but } \epsilon_T \not\perp\!\!\!\perp \epsilon_M, \ \epsilon_M \not\perp\!\!\!\perp \epsilon_Y, \textit{ and } \epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M \tag{11}$$

The only assumption in **A-2** that is needed for identification is that error terms ϵ_T, ϵ_Y are unconditionally independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_Y$. None of the remaining assumptions $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$, and $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$ is needed for identification. These remaining assumption merely clarify that we do not assume away endogeneity in any of the key relations: We allow for error terms ϵ_T, ϵ_M to correlate, we allow for error terms ϵ_M, ϵ_Y to correlate, and we allow for ϵ_T, ϵ_Y to correlate conditional on ϵ_M, i.e. $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$, which implies M is endogenous in equation (3).

Remark 2.2 Under $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, either $\epsilon_T \perp\!\!\!\perp \epsilon_M$ or $\epsilon_M \perp\!\!\!\perp \epsilon_Y$ would imply $\epsilon_T \perp\!\!\!\perp \epsilon_Y | \epsilon_M$:
(*i*) If $\epsilon_T \perp\!\!\!\perp \epsilon_M$ and $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$, then the causal effects of T on M and T on Y (*Model I* and *Model II*) could be identified using OLS. However, identification of the causal effect of M on Y would require a separate dedicated instrument $\tilde{Z}$ that has no impact on Y other than through M.
(*ii*) If $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$ and $\epsilon_M \perp\!\!\!\perp \epsilon_Y$, then the causal effects of T on M and T on Y could only be identified using IV. However, the causal effect of M on Y could be identified using OLS.
(*iii*) Finally, if $\epsilon_T \perp\!\!\!\perp \epsilon_M$ and $\epsilon_M \perp\!\!\!\perp \epsilon_Y$, then the causal relations (1)–(3) could be identified by the three OLS regressions in footnote 6.

It is important to note that we are not claiming that the assumption $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ is appropriate in all mediation-type settings characterized by causal relations (1)–(3). Instead, our identification approach is appropriate for settings like ours where T is endogenous in the relation of T and Y primarily because of omitted variables that affect M, and through M also Y. There will be settings where this assumption is implausible or at least debatable. It is therefore desirable to be able to relax $\epsilon_T \perp\!\!\!\perp \epsilon_Y$. We do this in Section 2.4, where we show that under $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y$ we can still bound the correlations between error terms and the estimates of all key causal relations in the data.

A-2 yields a non-obvious new exclusion restriction, which allows for the use of Z as an instrument for M, when conditioned on T, as stated in Lemma **L-2**. Critically, this implies that we can identify three causal effects using only a single instrument (or set of instruments) dedicated for T.

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

Targeted Causal Relation	*IV Relevance (First Stage)*		*Exclusion Restriction*
for $M \to Y$	$Z \not\perp\!\!\!\perp M \mid T$	*and*	$Z \perp\!\!\!\perp Y(m) \mid T$

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

Lemma **L-2** is novel, and it is worth discussing both Z's explanatory power for M, conditional on T, as well as Z's validity as an instrument for M, conditional on T: The relevance of the IV for M conditional on T, i.e. $Z \not\perp\!\!\!\perp M|T$, comes from a residual variation argument, which we again illustrate in our empirical application: there, the endogeneity concern in the relation between T and M is that 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 the literature is to use other (high-wage) countries' imports as the basis of an instrument (Z) that is orthogonal to Germany-specific demand conditions. The key piece of intuition for Z's explanatory power for M, conditional on T, is that when other countries' industry-specific imports are high relative to (i.e. conditional on) German industry-specific imports, these will partly reflect the unobserved German industry-specific demand conditions, i.e. the source of the bias. Conditional on German imports (T), other countries' higher imports in a given sector therefore "cause" additional reductions in German employment because they reflect negative German demand conditions.[20] The exclusion restriction of **L-2** 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. In other words, Z is a valid instrument for M only when conditioning on T. In Online Appendix C we summarize the error dependences of **A-2** once more visually in a directed acyclic graph (DAG), as in *Model III* in Table 1.

Remark 2.3 The exclusion restrictions in **L-1** and **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: $T = f_T(Z, \epsilon_T)$, $U = f_U(T, \epsilon_U)$, $M = f_M(T, U, \epsilon_M)$, and $Y = f_Y(T, M, U, \epsilon_Y)$. We investigate this model in Online Appendix D.

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., $(Y(m)|T = t) \stackrel{d}{=} Y(m,t)$.*

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

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 $Y(m,t)$. As a consequence, the total effect (equation (8)) can be decomposed into the direct effect (equation (9)), and the indirect effect (equation (10)).

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

Remark 2.4 An alternative identification approach to identify the counterfactual outcome $Y(m)$ is with an additional dedicated IV. Consider an additional variable $\tilde{Z}$ that plays the role of an IV that is exclusively dedicated to M. This means that variable $\tilde{Z}$ is characterized by two properties: (*i*) $\tilde{Z}$ does not cause T; and (*ii*) $\tilde{Z}$ 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 e.g. Frölich and Huber (2017), and also Online Appendix E, where we discuss the potential applicability as well as pitfalls of using a second dedicated instrument $\tilde{Z}$.

2.3 The IV Mediation Model under Linearity

As is standard, in section 2.2 we have used a non-parametric model to discuss causality. To derive an easily operationalized estimation procedure from the identification results we derived, we will assume linearity from here on. Linearity is commonly assumed in many applied literatures, in particular is it pervasive in the local labor markets literature which our empirical application belongs to. Under linearity, the exclusion restrictions in section 2.2 simply translate into a lack of correlation between error terms. We will show that this renders a just-identified model. We will show that we can use standard 2SLS estimation procedures to identify the causal effect of T on M, T on Y, and M on Y, thereby decomposing the total effect of T on Y into its direct and indirect effect. Under linearity (and adding the instrument), the causal relations in equations (1)–(3) can be written as follows:

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

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} stand for the correlation between ϵ_T and 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 a mean of 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 $IE = \beta_M^T \cdot \beta_Y^M$, and the total effect is the sum of these two terms $TE = \beta_Y^T + \beta_M^T \cdot \beta_Y^M$.[21]

[21] The linear model lacks treatment-mediator interactions and therefore has homogeneous effects, i.e. Pearl's nat-

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 equations (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.

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

Equations (12)–(15) are thus written as $\mathbf{X} = \mathbf{\Psi} \cdot \mathbf{X} + \boldsymbol{\epsilon}$. Equation (17) presents the covariance matrix $\Sigma_{\mathbf{X}}$ of observed variables $\mathbf{X}$:

$$\Sigma_{\mathbf{X}} \equiv \mathbf{Var}\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix} = \begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} & \sigma_{ZM} & \sigma_{ZY} \\ \cdot & \sigma_{TT} & \sigma_{TM} & \sigma_{TY} \\ \cdot & \cdot & \sigma_{MM} & \sigma_{MY} \\ \cdot & \cdot & \cdot & \sigma_{YY} \end{bmatrix}. \tag{17}$$

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

$$\Sigma_\epsilon \equiv \mathbf{Var}\begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix} = \begin{bmatrix} \sigma_{\epsilon_Z}^2 & 0 & 0 & 0 \\ \cdot & \sigma_{\epsilon_T}^2 & \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} & \rho_{TY}\sigma_{\epsilon_T}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \sigma_{\epsilon_M}^2 & \rho_{MY}\sigma_{\epsilon_M}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \cdot & \sigma_{\epsilon_Y}^2 \end{bmatrix}. \tag{18}$$

Assumption **A-2** states that $\epsilon_T \perp\!\!\!\perp \epsilon_Y$, but allows for $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$, and $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$. Under linearity, this translates into $\rho_{TY} = 0$, but $\rho_{TM} \neq 0$ and $\rho_{MY} \neq 0$ in Σ_ϵ. As well, $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$ implies $\rho_{TY|\epsilon_M} \neq 0$. In Appendix B, we describe how to straightforwardly generate a simulated dataset with these dependence relations.

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 (18) is based on the equality between the covariance matrices of the observed random

ural and controlled effects (see our footnote 14) coincide. This is called the no-interaction assumption, in which the mediation and direct effects do not depend on the values of the treatment T.

variables and unobserved error terms, namely

$$\mathbf{X} = \mathbf{\Psi} \cdot \mathbf{X} + \boldsymbol{\epsilon} \quad \Rightarrow \quad (\mathbf{I} - \mathbf{\Psi})\, \mathbf{X} = \boldsymbol{\epsilon} \quad \Rightarrow \quad (\mathbf{I} - \mathbf{\Psi})\, \Sigma_{\mathbf{X}}\, (\mathbf{I} - \mathbf{\Psi})' = \Sigma_{\boldsymbol{\epsilon}}. \tag{19}$$

Let $\widetilde{\Sigma}_{\boldsymbol{\epsilon}}$ be the covariance matrix (18) under our key assumption that $\rho_{TY} = 0$. This covariance matrix is displayed in (21).

$$\widetilde{\Sigma}_{\mathbf{X}} \equiv \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_T^Z & 1 & 0 & 0 \\ 0 & -\beta_M^T & 1 & 0 \\ 0 & -\beta_Y^T & -\beta_Y^M & 1 \end{bmatrix}}_{\mathbf{I}-\mathbf{\Psi}} \cdot \underbrace{\begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} & \sigma_{ZM} & \sigma_{ZY} \\ \cdot & \sigma_{TT} & \sigma_{TM} & \sigma_{TY} \\ \cdot & \cdot & \sigma_{MM} & \sigma_{MY} \\ \cdot & \cdot & \cdot & \sigma_{YY} \end{bmatrix}}_{\Sigma_{\mathbf{X}}} \cdot \underbrace{\begin{bmatrix} 1 & -\beta_T^Z & 0 & 0 \\ 0 & 1 & -\beta_M^T & -\beta_Y^T \\ 0 & 0 & 1 & -\beta_Y^M \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{(\mathbf{I}-\mathbf{\Psi})'} = \tag{20}$$

$$= \underbrace{\begin{bmatrix} \sigma_{\epsilon_Z}^2 & 0 & 0 & 0 \\ \cdot & \sigma_{\epsilon_T}^2 & \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} & 0 \\ \cdot & \cdot & \sigma_{\epsilon_M}^2 & \rho_{MY}\sigma_{\epsilon_M}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \cdot & \sigma_{\epsilon_Y}^2 \end{bmatrix}}_{\Sigma_{\boldsymbol{\epsilon}} \text{ under } \rho_{TY}=0} \equiv \widetilde{\Sigma}_{\boldsymbol{\epsilon}}. \tag{21}$$

The equality (20)–(21) 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 $\widetilde{\Sigma}_{\mathbf{X}} \equiv (\mathbf{I} - \mathbf{\Psi})\, \Sigma_{\mathbf{X}}\, (\mathbf{I} - \mathbf{\Psi})'$, where $\widetilde{\Sigma}_{\mathbf{X}}[i,j]$ denotes the element in the i-th row and j-th column of the matrix $(\mathbf{I} - \mathbf{\Psi})\, \Sigma_{\mathbf{X}}\, (\mathbf{I} - \mathbf{\Psi})'$. 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 $i \leq j; i,j \in \{1,2,3,4\}$. Four of the six off-diagonal equations 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 (21):

$$\widetilde{\Sigma}_{\mathbf{X}}[1,2] = 0 \Rightarrow \sigma_{ZT} - \beta_T^Z \sigma_{ZZ} = 0 \qquad \Rightarrow \qquad \beta_T^Z = \frac{\sigma_{ZT}}{\sigma_{ZZ}}. \tag{22}$$

$$\widetilde{\Sigma}_{\mathbf{X}}[1,3] = 0 \Rightarrow \sigma_{ZM} - \beta_M^T \sigma_{ZT} = 0 \qquad \Rightarrow \qquad \beta_M^T = \frac{\sigma_{ZM}}{\sigma_{ZT}}. \tag{23}$$

$$\left.\begin{matrix} \widetilde{\Sigma}_{\mathbf{X}}[1,4] = 0 \\ \widetilde{\Sigma}_{\mathbf{X}}[2,4] = 0 \end{matrix}\right\} \Rightarrow \left.\begin{matrix} \sigma_{ZY} - \beta_Y^M \sigma_{ZM} - \beta_Y^T \sigma_{ZT} = 0 \\ \sigma_{TY} - \beta_Y^M \sigma_{TM} - \beta_Y^T \sigma_{TT} = 0 \end{matrix}\right\} \qquad \Rightarrow \qquad \left\{\begin{matrix} \beta_Y^M = \frac{\sigma_{ZT}\sigma_{TY} - \sigma_{TT}\sigma_{ZY}}{\sigma_{ZT}\sigma_{TM} - \sigma_{TT}\sigma_{ZM}}. \\ \beta_Y^T = -\frac{\sigma_{ZM}\sigma_{TY} - \sigma_{TM}\sigma_{ZY}}{\sigma_{ZT}\sigma_{TM} - \sigma_{TT}\sigma_{ZM}}. \end{matrix}\right. \tag{24}$$

The four equalities in (22)–(24) suffice to identify all linear coefficients $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ of equations (12)–(15). There are six remaining equalities in (20)–(21). The four diagonal equations generated upon (20)–(21) identify the variances of the error terms. Those equations are listed in (25). The left-hand side of each equation consists of observed covariances or identified parameters. The right-hand side of each equation consists of the error variances.

$$\begin{aligned}
\widetilde{\Sigma}_{\mathbf{X}}[1,1] = \widetilde{\Sigma}_{\mathbf{e}}[1,1] \;&\Rightarrow & \sigma_{ZZ} &= \sigma_{\epsilon_Z}^2. \\
\widetilde{\Sigma}_{\mathbf{X}}[2,2] = \widetilde{\Sigma}_{\mathbf{e}}[2,2] \;&\Rightarrow & \left(\sigma_{TT} - \beta_T^Z\sigma_{ZT}\right) - \beta_T^Z\left(\sigma_{ZT} - \beta_T^Z\sigma_{ZZ}\right) &= \sigma_{\epsilon_T}^2. \\
\widetilde{\Sigma}_{\mathbf{X}}[3,3] = \widetilde{\Sigma}_{\mathbf{e}}[3,3] \;&\Rightarrow & \left(\sigma_{MM} - \beta_M^T\sigma_{TM}\right) - \beta_M^T\left(\sigma_{TM} - \beta_M^T\sigma_{TT}\right) &= \sigma_{\epsilon_M}^2. \\
\widetilde{\Sigma}_{\mathbf{X}}[4,4] = \widetilde{\Sigma}_{\mathbf{e}}[4,4] \;&\Rightarrow & \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix}' \begin{bmatrix} \sigma_{YY} & \sigma_{MY} & \sigma_{TY} \\ \sigma_{MY} & \sigma_{MM} & \sigma_{TM} \\ \sigma_{TY} & \sigma_{TM} & \sigma_{TT} \end{bmatrix} \begin{pmatrix} 1 \\ -\beta_Y^M \\ -\beta_Y^T \end{pmatrix} &= \sigma_{\epsilon_Y}^2.
\end{aligned} \tag{25}$$

The last two equalities can be extracted from (20)–(21) to identify the correlations ρ_{TM} and ρ_{MY}, which are described as follows:

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

With $\beta_T^Z, \beta_M^T, \beta_Y^M, \beta_Y^T$ identified by the four equalities in (22)–(24), and $\sigma_{\epsilon_Z}, \sigma_{\epsilon_T}, \sigma_{\epsilon_M}, \sigma_{\epsilon_Y}$ identified by the four equalities in (25), equation (26) is therefore just-identified for ρ_{TM} and 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 (22)–(24). Each identifying formula is associated with a well-known econometric estimator as follows:

1. Parameter β_T^Z is identified by (22) as the covariance between Z and T divided by the variance of Z. This formula implies that β_T^Z can be estimated by the OLS regression of T on Z.

$$\text{OLS:} \quad T = \beta_T^Z \cdot Z + \epsilon_T. \tag{27}$$

2. Parameter β_M^T is identified by (23) as the ratio of the covariance between Z and M divided by the covariance of Z and T. This formula implies that β_M^T can be estimated by a 2SLS estimation where Z is the IV, 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 as follows:

$$\text{First Stage:} \quad T = \beta_T^Z \cdot Z + \epsilon_T, \tag{28}$$

$$\text{Second Stage:} \quad M = \beta_M^T \cdot \hat{T} + \epsilon_M, \tag{29}$$

where $\hat{T}$ stands for the estimated values of T in the first stage.

3. Parameters β_Y^M, β_Y^T are jointly identified by the two remaining equations in (24). In Online Appendix F 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:

$$\text{First Stage:} \quad M = \gamma_M^Z \cdot Z + \gamma_M^T \cdot T + \epsilon_T, \tag{30}$$

$$\text{Second Stage:} \quad Y = \beta_Y^M \cdot \hat{M} + \beta_Y^T \cdot T + \epsilon_Y, \tag{31}$$

where $\hat{M}$ are the estimated values of M in the first stage.

The estimation procedure associated with identification formulas (28) and (29) is the standard IV approach, and commonly understood. By contrast, the estimation procedure associated with identification formulas (30) and (31) is novel. 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. For this, and the intuition for the first stage (30), we refer the reader back to the exclusion restriction of Lemma **L-2**, i.e. $Z \perp\!\!\!\perp Y(m)|T$, as well as Corollary **C-1**.

Having established our core result and the intuition for our model, we also want to make the link to Table 1 explicit. In it, *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 (28) and (29). *Model III* is the mediation model with IV. This model is estimated by *Model I* plus the 2SLS regression represented by the linear equations (30)–(31). *Model II* stands for the IV model that evaluates the total effect (TE) of T on Y directly. 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 (29) into the outcome equation in (31) as follows:

$$Y = \beta_Y^M \cdot M + \beta_Y^T \cdot T + \epsilon_Y \text{ and } M = \beta_M^T \cdot T + \epsilon_M \tag{32}$$

$$\Rightarrow Y = \beta_Y^M \cdot \left(\beta_M^T \cdot T + \epsilon_M\right) + \beta_Y^T \cdot T + \epsilon_Y \tag{33}$$

$$= \underbrace{\left(\beta_Y^M \cdot \beta_M^T + \beta_Y^T\right)}_{TE} \cdot T + \underbrace{\beta_Y^M \epsilon_M + \epsilon_Y}_{\eta_Y} \equiv g_Y(T, \eta_Y). \tag{34}$$

The error term η_Y of *Model II* is mapped into $\beta_Y^M \epsilon_M + \epsilon_Y$ in (34). 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**.[22]

Remark 2.5 In Online Appendix H, we extend the identification results and estimation methods of Section 2.3 to enable conditioning on additional covariates $\mathbf{K}$. The identification analysis amounts to simply replacing the covariance matrix $\widetilde{\Sigma}_{\mathbf{X}}$ of the observed variables $\mathbf{X}$ by the covariance matrix $\widetilde{\Sigma}_{\mathbf{X}|\mathbf{K}}$ of variables $\mathbf{X}$ conditioned on covariates $\mathbf{K}$. The identified parameters can be estimated by adding those covariates to the OLS regression (27) and the two 2SLS regressions defined in (28)–(29) and (30)–(31).

2.4 Allowing for General Error Dependency

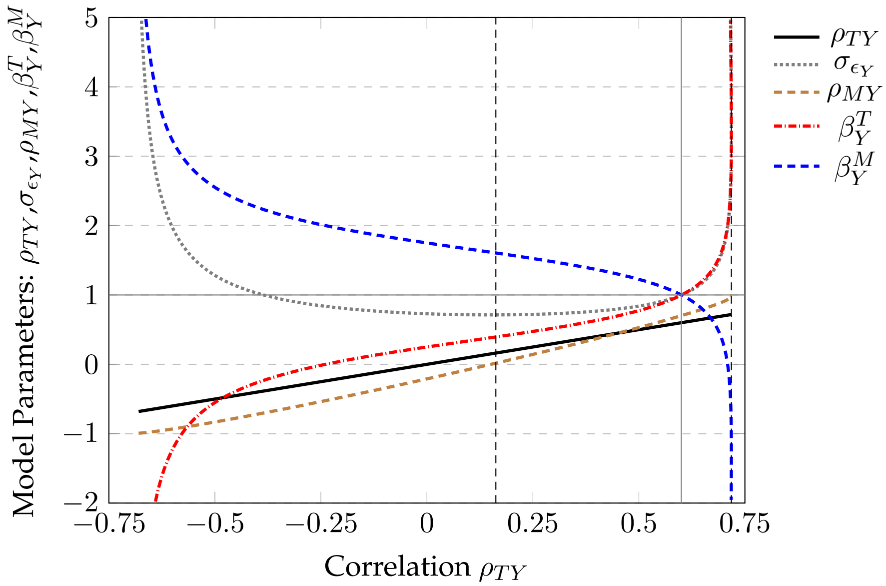
As previously noted, Assumption **A-2** is equivalent to assuming that T is endogenous in a regression of Y on T only because of confounders that affect T and M. Under linearity, Assumption **A-2** implies $\rho_{TY} = 0$, and generates a model in which all parameters are point-identified. We view Assumption **A-2** as quite plausible in our setting, but there are evidently many IV mediation settings (i.e. settings that can be described by equations (12)–(15)), where the identifying assumption is harder to defend. It is therefore helpful to extend the framework to relax Assumption **A-2**, i.e. allowing $\rho_{TY} \neq 0$. This is equivalent to leaving the statistical dependence among error terms entirely unrestricted. In Appendix C.1, we show how to generate a simulated dataset such that ρ_{TY}

[22] In Online Appendix G, 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 (28)–(29) and (30)–(31) 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.

can take any value.

Under $\rho_{TY} \neq 0$ we have a model with eleven instead of ten 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 (19), which yields ten identifying equalities. In Appendix C.2, 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 $\sigma^2_{\epsilon_Z}, \sigma^2_{\epsilon_T}, \sigma^2_{\epsilon_M}$. For the remaining five parameters, $\beta_Y^M, \beta_Y^T, \rho_{TY}, \rho_{MY}, \sigma^2_{\epsilon_Y}$, we are left with four equalities. Since these five parameters are thus not point-identified, neither the direct effect $DE = \beta_Y^T$ nor the indirect effect $IE = \beta_M^T \cdot \beta_Y^M$ are point-identified. We can still evaluate bounds for these effects. In Appendix C.2, we show a procedure for doing so in which we express parameters $\beta_Y^M, \beta_Y^T, \rho_{MY}$ and $\sigma^2_{\epsilon_Y}$ as functions of the values that parameter ρ_{TY} could take. In turn, we impose simple model restrictions on the covariance structure of the data to bound the possible range of ρ_{TY}. Figure 1 illustrates the bounds of the parameters when we generate a simulated dataset, as described in Appendix C.1, with a true value of the correlation $\rho_{TY} = 0.6$ (solid vertical line). Each possible value of ρ_{TY} within the lower and upper bound (dashed vertical lines around 0.2 and 0.7) implies values of the other four parameters $\beta_Y^M, \beta_Y^T, \rho_{MY}$, and $\sigma^2_{\epsilon_Y}$. Appendix C.1 generates the simulated data so that $\beta_T^Z = \beta_M^T = \beta_Y^T = \beta_Y^M = \sigma_{\epsilon_Y} = 1$. At the true value of $\rho_{TY} = 0.6$, the values of the β_Y^T, β_Y^M and σ_{ϵ_Y} also attain their true values, that is, $\sigma_{\epsilon_Y} = \beta_Y^T = \beta_Y^M = 1$. In Section 5, we will apply this bounding exercise to our empirical application.

Figure 1: Bounds for Parameters ρ_{TY},σ_{ϵ_Y},ρ_{MY},β_Y^T and β_Y^M.



Notes: This figure presents the bounds for model parameters ρ_{TY},σ_{ϵ_Y},ρ_{MY},β_Y^T,β_Y^M. The dashed vertical lines denote the bounds on ρ_{TY}. These in turn generate the bounds of all remaining parameters. The true value of ρ_{TY}, that is, $\rho_{TY} = 0.6$, is marked by a vertical solid line. Note that at the true value of $\rho_{TY} = 0.6$, the values of the β_Y^T, β_Y^M and σ_{ϵ_Y} also attain their true values, that is, $\sigma_{\epsilon_Y} = \beta_Y^T = \beta_Y^M = 1$.

3 Data

Our data is organized as a stacked panel of first differences between election dates, 1987 to 1998 and 1998 to 2009, staying as close as possible to the decadal changes usually studied in the literature. We study German regions' exposure to trade with Eastern Europe and China, which 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 from 1987 to 1998 (from about 20 billion to about 60 billion Euros each),[23] and again tripled between 1998 and 2009.

Our data is observed at the county (*Landkreis*) level.[24] 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. This imbalanced nature of the panel has no bearing on

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

[24] 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.

the estimation because we include period-specific region fixed effects for four broad regions in all estimations (North, West, South, and East Germany). Indeed, none of our results are affected at all by dropping East Germany altogether; we include it primarily to stay close to the existing literature. We observe 408 counties in our data, 86 of which are in East Germany. Over the two periods, we have 730 ($(408-86)+408$) observations in total. For reference, we represent the data as two separate *Landkreise*-maps for the two periods in Appendix D.

With a view toward 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 IV for T_{it}. We now explain how these variables are measured.[25]

3.1 Import Exposure (Treatment T)

Following Autor et al. (2013) and Dauth et al. (2014), we construct *net import exposure* as follows:[26]

$$T_{it} = \sum_j \frac{L_{ijt}}{L_{jt}} \frac{\Delta IM_{Gjt} - \Delta EX_{Gjt}}{L_{it}}. \tag{35}$$

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.[27] 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 manufacturing countries are not only more balanced in the aggregate than in the U.S. but also correlate positively at the industry level. As a result, when working with German data it is better to consider a local labor market's *net* import exposure, and this is what we do throughout the paper.

One concern with the measure of import exposure in equation (35) is that it is a composite effect

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

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

[27] 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 U.N. Comtrade trade data using the crosswalk described in Dauth et al. (2014), which covers 157 manufacturing industries.

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 well-known problem 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 database to glean information on workers' industry of employment, occupation, and place of work for all German workers subject to social insurance.[28] 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 I 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*).[29] Due to its at-large voting system, Germany, like most continental European countries, has consistently had a multiparty system that spans the full political spectrum from far-left to extreme-right parties. This fact allows us to contrast the effect of import exposure on populists parties' vote share with that for moderate parties. We label four parties as "established" in that they were persistently represented in parliament over the 22 years we study. There is also many small parties whose vote share is always far below the 5 percent threshold of party votes needed to enter the federal parliament.[30] We collected these data to cre-

[28] 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.

[29] 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 it 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.

[30] 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 PDS party

ate a novel dataset of party vote shares at the county level. We group the small parties into three categories: far-left, extreme-right, other small parties. Altogether, *Landkreis*-level voting outcomes are divided into changes in the vote-share of (i) four mainstream parties (the CDU, the SPD, the FDP and the Green party), (ii) extreme-right parties, (iii) far-left parties, (iv) other small parties, and (v) turnout; see Falck et al. (2014).[31]

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), instrumenting Germany's net imports from China and Eastern Europe, $\Delta IM_{Gjt} - EX_{Gjt}$, with the average imports from (exports to) China and Eastern Europe of a similar set of high-wage economies, ΔIM_{Ojt} (ΔEX_{Ojt}):[32]

$$Z_{it}^{IM} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta IM_{Ojt}}{L_{it-1}} , \qquad Z_{it}^{EX} = \sum_j \frac{L_{ijt-1}}{L_{jt-1}} \frac{\Delta EX_{Ojt}}{L_{it-1}}. \tag{36}$$

Finally, again 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.

won four seats by *Erststimme*. As a result, the party received 30 seats in total, according its 4.4 percent of party votes (*Zweitstimme*) received.

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

[32] 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.

4.1 Model I

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

$$T_{it} = \beta_T^Z \cdot Z_{it} + \beta_T^K \mathbf{K}_{it} + \eta_{it}, \tag{37}$$

$$M_{it} = \beta_M^T \cdot T_{it} + \beta_M^K \mathbf{K}_{it} + \epsilon_{it}. \tag{38}$$

Equations (37) and (38) are equations (28) and (29) 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 (38).

$\mathbf{K}_{it}$ includes i's start-of-period manufacturing employment share; the employment share in the largest sector;[33] employment share in the chemical industry;[34] the start-of-period employment share that is foreign-born; share of female population and share of population at retirement age; and the start-of-period unemployment rate; finally, start-of-period turnout and parties' vote-shares. Also included is a set of period-specific region fixed effects (North, West, and South); the regions are comparable to U.S. Census divisions (Dauth et al., 2014).[35] 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 control 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 observe (*ii*) manufacturing's employment

[33] 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.

[34] To account for the spatial clustering of this specific industry in Germany.

[35] Each of West Germany's 10 states (*Bundesländer*) is fully contained inside one of these three regions.

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*).[36]

Without additional separate dedicated sets of instruments for each potential mediator, our 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.[37] This approach is appealing as long as the mediating effects are sharply concentrated in a single principal component that has a clear interpretation. We label the principal components as "labor market components" (LMC).

Table 2: 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 the 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 2 reports on the LMCs' factor loadings. By construction, there are always as many LMCs as variables but to keep this section brief, we

[36] 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).

[37] 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).

report only the first three. The convention is to regard only the LMCs with an eigenvalue larger than 1, i.e. only the first two in our case. 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 explains only 10.2 percent, the fourth and fifth together explain the remaining 10 percent. LMC_1 is somewhat ambiguous: it has a weak positive correlation with overall employment but a negative one with manufacturing employment. Moreover, it is positively correlated with changes in wages, but also with unemployment.[38] 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. Thus, while LMC_1 appears to capture the divergence between manufacturing jobs and the rest of the labor market, LMC_2 captures common changes between manufacturing jobs and the overall labor market, i.e. the core labor market adjustments that are most affected by import exposure.

The estimation results of *Model I* are displayed in Table 3, 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.[39] 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 a control for the employment share in the chemical industry. In column 3, we add controls for the start-of-period employment share that is foreign-born, share of female population, population share at retirement age, and the start-of-period unemployment rate. In column 4, we add start-of-period turnout and party vote-shares. 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

[38] 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 past 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).

[39] 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 Section 3.4.

Table 3: *Model I*: Effect of Import Exposure T on Labor Markets M

	(1)	(2)	(3)	(4)
Panel A: Second Stage (38), for Individual Labor Market Outcomes				
Δ log employment	-0.023***	-0.024***	-0.024***	-0.022***
	[0.004]	[0.006]	[0.003]	[0.006]
Δ Share Manuf. Empl.	-0.440**	-0.486**	-0.684***	-0.732***
	[0.048]	[0.018]	[0.000]	[0.000]
Δ log Manuf. Wage	-0.006**	-0.006**	-0.005**	-0.005**
	[0.013]	[0.014]	[0.020]	[0.018]
Δ log Non-Manuf. Wage	-0.005***	-0.005***	-0.003**	-0.002
	[0.004]	[0.002]	[0.046]	[0.309]
Δ Share Unempl.	0.076	0.061	0.127*	0.082
	[0.271]	[0.426]	[0.053]	[0.232]
LMC_1	-0.032	-0.032	0.024	0.029
	[0.495]	[0.473]	[0.512]	[0.409]
LMC_2	-0.265***	-0.278***	-0.322***	-0.308***
	[0.003]	[0.002]	[0.000]	[0.001]
LMC_3	-0.028	-0.028	-0.060*	-0.056
	[0.523]	[0.564]	[0.097]	[0.156]
LMC_4	0.023	0.011	-0.001	-0.035
	[0.435]	[0.698]	[0.966]	[0.255]
LMC_5	0.017	0.011	0.007	-0.009
	[0.415]	[0.589]	[0.749]	[0.691]
Panel B: First Stage Equation (37)				
β_M^{IM}	0.225***	0.212***	0.214***	0.217***
	[0.000]	[0.000]	[0.000]	[0.000]
β_M^{EX}	-0.206***	-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 the two periods 1987–1998 and 1998–2009, and 86 *Landkreise* in East Germany 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 add controls: Column 1 controls only for start-of-period manufacturing and period-specific region fixed effects. Column 2 adds further controls for industrial structure listed in the text. Column 3 adds further socio-economic and demographic controls listed in the text. Column 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} (€1,372 per worker) decreased total employment by about 3 percent, ($e^{-0.022 \cdot 1.372} - 1 = -0.03$). (*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.

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} (€1,372 per worker) decreases total employment by about 3 percent ($e^{-0.022 \cdot 1.372} - 1 = -0.03$).[40] 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 PC analysis 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 0.001 or lower in all specifications. By contrast, the other four LMCs are largely unaffected by import competition.[41] 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 (37). These results are highly significant, have the expected sign, and are in line with estimates in the existing literature.[42] We do not study any individual-level labor market results; we 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 (37) with *Model I*, but replaces M_{it} with Y_{it} in the second-stage equation (38). 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 4 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-

[40] We report corresponding OLS results in Online Appendix L (Table Online Appendix Table 4).

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

[42] For added clarity, we break the instrument $\beta_T^Z \cdot Z_{it}$ in the first-stage equation (37) 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.

shares of moderate, small, extreme-right, and far-left parties, as well as turnout. Each row in Table 4 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 3. 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. Looking at the whole political spectrum, it is only the support of extreme-right parties that consistently and significantly responds to increasing import competition. We see no significant effects on turnout or on established, small, or far-left parties in our preferred specification in column 4 of table 4. Only the FDP vote share shows something close to a significant response to import exposure, but this response is much less significant than the response of the extreme-right parties' vote share and half the relative magnitude (in column 5), so that our focus is on the extreme-right parties.[43] In our preferred specification in column 4, a one-standard-deviation increase in T_{it} (€1,372 per worker) increases the extreme-right vote share by 0.13 ($0.096 \cdot 1.372$) percentage points, around 20 percent of the average per-decade increase of 0.64 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 other parties are also economically insignificant as compared to the effect on extreme-right parties.[44]

Interpreting the effects: Whether far-left or extreme-right populists capture protectionist sentiments is ultimately country-specific. As *The Economist* (July 30th, 2016) put it in a headline,"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 over the past 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

[43] The less precisely estimated effect on the vote share of the market-liberal FDP may be interpreted as weak evidence for polarization. One possible explanation for this effect 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. More generally, beneficiaries from import competition may demand increasing trade liberalization. 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.

[44] We present corresponding OLS estimates in Table Online Appendix Table 5 of Online Appendix L.

Table 4: *Model II*: Effect of Import Exposure (T) on Voting (Y)

	(1)	(2)	(3)	(4)	(5)
Δ Turnout	0.002	0.003	0.003	0.002	0.032
	[0.348]	[0.266]	[0.225]	[0.254]	[0.254]
<u>Established Parties:</u>					
Δ voteshare CDU,CSU	-0.128	-0.159	-0.174	-0.090	-0.021
	[0.457]	[0.369]	[0.310]	[0.471]	[0.471]
Δ voteshare SPD	-0.020	-0.004	0.024	0.026	0.002
	[0.897]	[0.978]	[0.872]	[0.810]	[0.810]
Δ voteshare FDP	0.215***	0.204***	0.202***	0.120	0.022
	[0.005]	[0.009]	[0.007]	[0.102]	[0.102]
Δ voteshare Greens	-0.132**	-0.152***	-0.093*	-0.056	-0.020
	[0.022]	[0.009]	[0.079]	[0.204]	[0.204]
<u>Non-Established Parties:</u>					
Δ voteshare Extreme-Right	0.118***	0.129***	0.122***	0.096**	0.047**
	[0.001]	[0.001]	[0.002]	[0.045]	[0.045]
Δ voteshare Far-Left	-0.037	-0.035	-0.087	-0.081	-0.021
	[0.773]	[0.793]	[0.440]	[0.434]	[0.434]
Δ voteshare Other Small	-0.015	0.018	0.006	-0.015	-0.012
	[0.696]	[0.640]	[0.874]	[0.713]	[0.713]
Controls	Baseline	+ Industry	+Socio	+ Voting	$\sim$ (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 the two periods 1987–1998 and 1998–2009, and 86 *Landkreise* in East Germany 1998–2009. We drop three city-states (Hamburg, Bremen, and Berlin). (*b*) All specifications include region-by-period fixed effects. Columns incrementally include added controls in identical fashion as Table 3. Column 4 is our preferred specification, which includes all 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 as column 5 shows also the most sizeable voting response, which occurs on the extreme right. (*d*) The first stage is identical to that in Table 3. *p-values* are reported in square brackets, standard errors are clustered at the level of 96 commuting zones. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

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."[45] In Appendix D, we provide anecdotal evidence linking local import exposure to increasing support for the extreme right in two German regions.

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 important, 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, however, not on the magnitude of the effect of import exposure on voting behavior but on the causal mechanisms underlying it.[46]

Robustness to Gravity Estimation: 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 a set of relevant destination markets.[47] The gravity results are reported in Appendix E and are in line with those in Table 4. Reporting these gravity results is standard in the related literature (Autor et al., 2013; Dauth et al., 2014), but our focus is naturally on the IV setting to which our identification framework applies.

Corroborating Evidence in an Individual-Level Analysis: One benefit of using German data is that Germany's Socio-Economic Panel (SOEP) has a long-run panel structure that is unique

[45] For illustration, we 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).

[46] 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 K.

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

among attitudinal socioeconomic surveys, starting in 1984 (GSOEP, 2007).[48] 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.[49] 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 4, e.g. one dummy if the individual would vote for the CDU, one if the individual would vote for the SPD, etc. The results, reported in Appendix F, mimic closely our main Table 4. 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 sectors, who are also most likely to experience adverse labor market effects from import exposure.

Without *Model III*, this is as far as we can go. Standard IV methods generate a causally identified effect of import exposure on total employment ($\beta_M^T = -0.027$) and a causally identified effect of import exposure on voting ($TE = 0.102$). We now estimate the mediation model to identify the extent to which 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.[50]

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

[49] 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.

[50] The total effect is estimated in *Model II* as 0.0963, reported in Table 4. The indirect effect is $\hat{\beta}_M^T$, reported in Table 3, multiplied by the effect of M on Y, which we now estimate.

Model III can be estimated by evaluating the following two-stage model:

$$M_{it} = \gamma_M^Z \cdot Z_{it} + \gamma_M^T \cdot T_{it} + \gamma_M^K \mathbf{K}_{it} + \epsilon_{it}, \tag{39}$$

$$Y_{it} = \beta_Y^M \cdot M_{it} + \beta_Y^T \cdot T_{it} + \beta_Y^K \mathbf{K}_{it} + \eta_{it}. \tag{40}$$

Equations (39) and (40) are exactly equations (30) and (31) in section 2.3, except that control variables are included and subscripts added to reflect the panel nature of the data.[51]

Which mediators should one consider in estimating equations (39) and (40)? Table 3 showed that two mediators in particular, i.e. the share of manufacturing employment as well as total employment, were most strongly impacted by import exposure. The principal component that summarizes the common variation in these two variables, LMC_2, was also very strongly impacted by import exposure. Since we cannot unpack the mediating effects of manufacturing employment and overall employment separately, we view LMC_2 as our key mechanism, as it is likely to capture all of the core labor market adjustments to import exposure. Nonetheless, for illustrative purposes, Table 5 reports the results of estimating equations (39) and (40) for all five labor market variables we have, as well as their first two principal components, across columns 1–7.

Panel A presents the results of the second-stage equation (40). As one may have expected, columns 1, 2 and 7 show that it is the labor market variables that most responded to import exposure in Table 3 that in turn appear to be the only ones to significantly impact extreme-right voting. The log of manufacturing wages is the only labor market outcome that responded to import exposure in Table 3 and is not estimated to have a significant mediation effect in Table 5. The point estimate $\hat{\beta}_Y^M$ in column 1 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. $6.181/100$, relative to an average per-decade increase of 0.64 percentage points during the 22 years we study.

The real importance of the point estimates $\hat{\beta}_Y^M$ lies in generating the indirect effect $IE = \hat{\beta}_Y^M \times \hat{\beta}_M^T$. This, together with other parameters of the mediation model is reported in Panel B of Table 5. The $\hat{\beta}_M^T$ of import exposure on log employment was estimated in Table 3 (as *Model I*), to be -0.0216. The $\hat{\beta}_M^T$ of import exposure on the share of manufacturing employment was estimated in Table 3 to be -0.7317. The $\hat{\beta}_M^T$ of import exposure on LMC_2 was estimated in Table 3 (as *Model I*) to be

[51] 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 5: Estimates of *Model III*, the Mediation Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mediating Variables:	Δ log Employm.	Δ Share Manuf. Employm.	Δ log Manuf. Wage	Δ log Non-Manuf. Wage	Δ Share Unempl.	LMC_1	LMC_2	LMC_2
Assumption A-2:	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Panel A: Second Stage (40)								
β_Y^M	-6.181***	-0.183**	-2.962	-63.684	-0.073	-2.345	-0.391***	$[-0.2227, -0.4193]$
	[0.000]	[0.031]	[0.188]	[0.231]	[0.805]	[0.192]	[0.004]	
DE: β_Y^T	-0.051	-0.099***	-0.019	-0.012	-0.002	0.120	-0.071**	$[0.0087, -0.0835]$
	[0.106]	[0.008]	[0.273]	[0.854]	[0.956]	[0.390]	[0.012]	
Panel B: *Model III* Parameters								
Model I: β_M^T	-0.0216	-0.7317	-0.005	-0.002	0.082	0.029	-0.308	-0.308
IE: $\beta_M^T \cdot \beta_Y^M$	0.134	0.134					0.121	$[0.069, 0.129]$
Model II: TE	0.0963	0.0963					0.0963	0.0963
$S = IE/TE$	0.139	0.139					0.125	$[0.711, 1.339]$
Panel C: First Stage Equation (39)								
γ_M^{IM}	-0.005**	-0.030	0.001**	-0.001	-0.020	-0.003	-0.032*	
	[0.011]	[0.473]	[0.024]	[0.107]	[0.192]	[0.698]	[0.094]	
γ_M^{EX}	0.006***	0.145***	0.003***	0.001	-0.017	0.013	0.082***	
	[0.000]	[0.003]	[0.000]	[0.306]	[0.264]	[0.156]	[0.000]	
γ_M^T	0.001	-0.354***	-0.002	0.001	0.130***	0.071***	-0.077	
	[0.864]	[0.007]	[0.281]	[0.477]	[0.001]	[0.002]	[0.152]	
Observations	730	730	730	730	730	730	730	
R-Squared	0.504	0.570	0.727	0.912	0.765	0.928	0.402	

Notes: (*a*) Columns 1–7 present the results of estimating *Model III* on the five labor-market outcomes, and the first two principal components Column 8 shows the results from the bounding exercise described in section 2.4. (*b*) Panel A presents the results of the second-stage equation (40). Panel B shows the main parameters of the mediation model. Panel C presents the results of the first-stage equation (39), where the estimated coefficients have the sign suggested by the logic discussed in section 2.2. (*c*) *p-values* are reported in square brackets in columns 1–7, with standard errors clustered at the level of 96 commuting zones. *** p<0.01, ** p<0.05, * p<0.1.

-0.308. The indirect effect IE is the product of $\hat{\beta}_Y^M \times \hat{\beta}_M^T$. The estimated IE of 0.134 happens to be identical to the third digit for log employment and the share of manufacturing employment, i.e. across columns 1–2. For LMC$_2$, our preferred summary measure of the impact of import exposure on all labor market outcomes, it is very similar at 0.121. The TE estimated in Table 4 (as *Model II*) is constant across columns. Finally, the implied share of the total effect that is mediated by labor market adjustments is 139 percent ($0.134/0.096$) when focusing on the log of total employment or the share of manufacturing employment, and it 125 when using our preferred summary measure LMC$_2$.

These results suggest that other channels linking import exposure to voting need to be politically moderating in the aggregate. Indeed, $\hat{\beta}_Y^T$ in Table 5 consistently has a negative sign.[52] Another way to state this is that if the only effect of import exposure was to decrease employment, the political response would stronger than the one actually observed.

Panel C presents the results of the first-stage equation (39). It is useful to develop a prior about the coefficients' expected signs before estimating equation (39), because the interpretation is not as straightforward as it is for the standard first-stage equation (37). The literature worries about domestic, industry-specific demand conditions as source of confounding bias. German industries that experience negative domestic demand shocks are expected to see fewer imports and declining employment. As discussed in section 2.2, this would suggests that, conditional on Germany's imports T, higher imports from China into other countries will "cause" additional reductions in German employment by virtue of acting as a proxy for negative German demand. The first-stage coefficients have signs that are in line with this argument. In column 1 for instance, other countries' imports from China worth €1,000 per worker reduce total employment by 0.5 percent ($\hat{\gamma}_M^{IM} = -0.005$), while the same amount in exports to China increases German employment by 0.6 percent ($\hat{\gamma}_M^{EX} = 0.006$).[53]

Columns 8 reports on the bounds we derive when we allow $\rho_{TY} \neq 0$. The parameter β_M^T in *Model I*, as well as the directly estimated TE in *Model II* are still point-identified when $\rho_{TY} \neq 0$. By contrast, the parameters β_Y^M, β_Y^T can only be bounded, using restrictions on the covariance in

[52] While we can only speculate, potentially moderating channels may follow from the association between import exposure and rich-countries' task-upgrading (Becker and Muendler, 2015), and switching production towards more differentiated and higher-markup varieties (Holmes and Stevens, 2014). Of course, broadly speaking one may interpret these as labor market adjustments as well but they are not part of the available labor market aggregates usually studied.

[53] As before, we use two instruments Z_{it}^{IM}, Z_{it}^{EX} as IVs for one endogenous variable, namely T.

the data; see section 2.4. The covariance structure in the data allows the mediated share IE/TE of the effect of import exposure on extreme-right voting that is explained by LMC_2 to lie between 71 percent and 134 percent.

In summary, under the identifying assumption $\rho_{TY} = 0$, the results reported here suggest that changes in labor markets explain about 125 percent of the observed polarizing effect of import exposure on voters. The results thus suggest that the core labor market adjustments to import exposure combined lead to a polarizing effect on voters that is one quarter larger than the actual observed effect, which has to imply that other channels linking import exposure to voting are both moderating and important. When we allow $\rho_{TY} \neq 0$, this leaving the statistical dependence among error terms entirely unrestricted, the bounds generated by the model restrictions suggest that labor market adjustments explain at least 71 percent of the polarizing effect of import exposure on voters. Therefore, even at the lower bound the evidence clearly suggests that any effective remedy for the populist response to trade integration needs to focus on labor markets first and foremost.

6 Conclusion

A substantial body of recent evidence shows that in high-wage manufacturing countries such as 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. Our focus here is to ask whether this 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 toward 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: We can use standard IV methods to causally identify the effect of a treatment (import competition T) on an outcome (labor market adjustments M), as well as causally identifying the effect of the same treatment on another outcome (voting for the extreme right Y). But if the effect of T on Y is likely to work through M, then standard IV methods cannot estimate the extend to which it does.

This matters if we care about the mechanisms linking T and Y. 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 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 researchers need to critically evaluate its validity before applying it to other mediation-type IV settings.

References

Arzheimer, K., "Contextual Factors and the Extreme Right Vote in Western Europe, 1980-2002," *American Journal of Political Science*, 2009, *53* (2), 259–275.

Autor, David, David Dorn, and Gordon H Hanson, "Untangling trade and technology: Evidence from local labour markets," *The Economic Journal*, 2015, *125* (584), 621–646.

_ , _ , and Gordon Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

_ , _ , _ , and Kaveh Majlesi, "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure," *NBER Working Paper*, 2016.

Bade, Franz-Josef, Claus-Friedrich Laaser, and Rüdiger Soltwedel, "Urban specialization in the internet ageempirical findings for Germany, Processed," *Kiel Institute for World Economics*, 2003.

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

Becker, Sascha O and Marc-Andreas Muendler, "Trade and tasks: an exploration over three decades in Germany," *Economic Policy*, 2015, *30* (84), 589–641.

Bender, Stefan, Anette Haas, and Christoph Klose, "IAB Employment Subsample 1975-1995 Opportunities for Analysis Provided by the Anonymised Subsample," *IZA Discussion Paper 117*, 2000.

Blundell, Richard and James Powell, "Endogeneity in Nonparametric and Semiparametric Regression Models," in L. P. Hansen M. Dewatripont and S. J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. 2, Cambridge, UK: Cambridge University Press, 2003.

_ **and** _ , "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, July 2004, *71* (3), 655–679.

Che, Yi, Yi Lu, Justin R Pierce, Peter K Schott, and Zhigang Tao, "Does Trade Liberalization with China Influence US Elections?," Technical Report, National Bureau of Economic Research 2016.

Christensen, Garret S and Edward Miguel, "Transparency, reproducibility, and the credibility of economics research," Technical Report, National Bureau of Economic Research 2016.

Dahl, Christian M., Martin Huber, and Giovanni Mellace, "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*, 2017.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum, "The Rise of the East and the Far East: German Labor Markets and Trade Integration," *Journal of European Economic Association*, 2014, *12* (6), 1643–1675.

Dippel, Christian, Robert Gold, and Stephan Heblich, "Globalization and its (Dis-) Content: Trade Shocks and Voting Behavior," *NBER Working Paper*, 2015, (w21812).

Dix-Carneiro, Rafael and Brian K Kovak, "Trade liberalization and regional dynamics," *American Economic Review*, 2017, *107* (10), 2908–46.

Duranton, Gilles and Diego Puga, "From sectoral to functional urban specialisation," *Journal of Urban Economics*, 2005, *57* (2), 343–370.

Falck, Oliver, Robert Gold, and Stephan Heblich, "E-lections: Voting Behavior and the Internet," *American Economic Review*, 2014, *104* (7), 2238–65.

Feigenbaum, James J and Andrew B Hall, "How Legislators Respond to Localized Economic Shocks: Evidence from Chinese Import Competition," *Journal of Politics*, 2015, *77* (4), 1012–30.

Frölich, Markus and Martin Huber, "Direct and indirect treatment effects: causal chains and mediation analysis with instrumental variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2017, pp. n/a–n/a.

Goos, Maarten, Alan Manning, and Anna Salomons, "Job polarization in Europe," *The American Economic Review*, 2009, *99* (2), 58–63.

_ **,** _ **, and** _ , "Explaining job polarization: Routine-biased technological change and offshoring," *The American Economic Review*, 2014, *104* (8), 2509–2526.

Grumke, Thomas, *The Extreme Right in Europe*, Vandenhoeck & Ruprecht, 2012.

GSOEP, "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements," Technical Report 1 2007.

Hafeneger, Benno and Sven Schönfelder, *Politische Strategien gegen die extreme Rechte in Parlamenten. Folgen für kommunale Politik und lokale Demokratie*, Friedrich-Ebert-Stiftung: Berlin, 2007.

Heckman, James and Rodrigo Pinto, "Unordered Monotonicity," *Forthcoming Econometrica*, 2017.

Heckman, James J. and Edward J. Vytlacil, "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, May 2005, *73* (3), 669–738.

__ **and Rodrigo Pinto**, "Causal Analysis after Haavelmo," *Econometric Theory*, 2015, *31* (1), 115–151.

Heckman, James J and Rodrigo Pinto, "Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs," *Econometric reviews*, 2015, *34* (1-2), 6–31.

Heckman, James J., Rodrigo Pinto, and Peter A. Savelyev, "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 2013, *103* (6), 2052–2086.

Holmes, Thomas J and John J Stevens, "An alternative theory of the plant size distribution, with geography and intra-and international trade," *Journal of Political Economy*, 2014, *122* (2), 369–421.

Imai, Kosuke, Luke Keele, and Dustin Tingley, "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 2010, *15* (4), 309–334.

__ **,** __ **, and Te Yamamoto**, "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 2010, *25* (1), 51–71.

__ **,** __ **, Dustin Tingley, and Teppei Yamamoto**, "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies," *American Political Science Review*, 2011, *105* (4), 765–789.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, March 1994, *62* (2), 467–475.

Jun, Sung Jae, Joris Pinkse, Haiqing Xu, and Neşe Yildiz, "Multiple Discrete Endogenous Variables in Weakly-Separable Triangular Models," *Econometrics*, 2016, *4* (1).

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental analysis of neighborhood effects," *Econometrica*, 2007, *75* (1), 83–119.

Krugman, Paul R, "Trade and Wages, Reconsidered," *Brookings Papers on Economic Activity*, 2008, *2008* (1), 103–154.

MacKinnon, David Peter, *Introduction to statistical mediation analysis*, Routledge, 2008.

Malgouyres, Clement, "The Impact of Exposure to Low-Wage Country Competition on Votes for the Far-Right: Evidence from French Presidential Elections," *working paper*, 2014.

Malgouyres, Clément, "The impact of chinese import competition on the local structure of employment and wages: Evidence from france," *Journal of Regional Science*, 2017, *57* (3), 411–441.

Mughan, Anthony and Dean Lacy, "Economic Performance, Job Insecurity and Electoral Choice," *British Journal of Political Science*, 2002, *32* (3), 513–533.

__ **, Clive Bean, and Ian McAllister**, "Economic globalization, job insecurity and the populist reaction," *Electoral Studies*, 2003, *22* (4), 617–633.

New York Times, "Ancient City's Nazi Past Seeps Out After Stabbing," 2009, *February 11th.*

Pearl, Judea, "The Mediation Formula: A Guide to the Assessment of Causal Pathways in Nonlinear Models," 2011. Forthcoming in *Causality: Statistical Perspectives and Applications*.

_ , "Interpretation and identification of causal mediation," *Psychological Methods, Special Section: Naturally Occurring Section on Causation Topics in Psychological Methods*, 2014, *19*, 459–481.

Pierce, Justin R and Peter K Schott, "The Surprisingly Swift Decline of US Manufacturing Employment," *American Economic Review*, 2016, *106* (7), 1632–62.

Robins, James M. and Sander Greenland, "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 1992, *3* (2), 143–155.

Rodrik, Dani, "Political economy of trade policy," *Handbook of international economics*, 1995, *3* (3), 1457–1494.

Sommer, Bernd, "Anti-capitalism in the name of ethno-nationalism: ideological shifts on the German extreme right," *Patterns of Prejudice*, 2008, *42* (3), 305–316.

The Economist , "The New Political Divide," July 30th, 2016.