INTERNATIONAL MONETARY FUND

Specialization, Market Access and Real Income

Dominick Bartelme, Ting Lan, Andrei A. Levchenko

WP/24/51

2024
MAR

WORKING PAPER

 WP/24/51

IMF Working Paper
Research Department

Specialization, Market Access and Real Income
Prepared by Dominick Bartelme, Ting Lan, and Andrei A. Levchenko*

Authorized for distribution by Jaewoo Lee
March 2024

ABSTRACT: This paper estimates the impact of external demand shocks on real income. We utilize a first order approximation to a wide class of small open economy models that feature sector-level gravity in trade flows, which allows us to measure foreign shocks and characterize their welfare impact in terms of reduced-form elasticities. We use machine learning techniques to group 4-digit manufacturing sectors into a smaller number of clusters, and show that the cluster-level elasticities of income with respect to foreign shocks can be estimated using high-dimensional statistical techniques. Foreign demand shocks in complex intermediate and capital goods have large positive impacts on real income, whereas impacts in other sectors are negligible. We showthat the estimates imply that countries that specialize in these sectors enjoy greater gains from increased openness, and that (small) export subsidies to these sectors are welfare-improving. Finally, a calibrated multi-sector production and trade model with input-output linkages and external economies of scale can match the empirical estimates.

RECOMMENDED CITATION: Bartelme, Dominick, Ting Lan, and Andrei A. Levchenko, 2024, "Specialization, Market Access and Real Income," IMF Working Paper 24/51.

JEL Classification Numbers:	F43, F62
Keywords:	trade specialization; real income; gravity; k-means clustering
Author's E-Mail Address:	dbartelm@umich.edu, tlan@imf.org, alve@umich.edu

* "The authors would like to thank Christian Hansen, Kalina Manova, Fernando Parro, John Ries, Andrés Rodríguez-Clare, Sebastian Sotelo, Dan Trefler, and workshop participants at AEA 2021, Berkeley, CEPR MG, CUHK, IADB, MSU, NBER ITI, the NYUAD-CEPR "The End of Globalization?" Conference, the Rocky Mountain Empirical Trade Conference, Sciences Po Summer Workshop, Texas, Toronto, the Trade Integration and Growth Conference (Bogota), Western Ontario, and Yale for helpful suggestions, and Junwei Tang for superb research assistance.

WORKING PAPERS

Specialization, Market Access and Real Income

Prepared by Dominick Bartelme, Ting Lan, Andrei A. Levchenko

1. Introduction

How much do countries benefit from access to the international market? Are the welfare effects of foreign market access in some sectors larger than in others? Do some countries, by virtue of their comparative advantage, benefit more from globalization? How should trade and industrial policy respond to these sectoral differences? These questions are at the heart of the study of international trade as well as recent policy debates. An important and active tradition, following in the footsteps of Eaton and Kortum (2002), tackles these questions using quantitative models that are calibrated to data and micro-level elasticities. While this approach has been tremendously and deservedly influential, its limitation is that the general equilibrium outcomes are often sensitive to model features (e.g. Costinot and Rodríguez-Clare, 2014; Kehoe et al., 2017). Because the general equilibrium predictions are typically neither disciplined by nor tested against the data, these quantifications provide little guidance as to which (if any) of the many possible models give reliable answers to these questions.

This paper takes an alternative approach: we directly estimate the general equilibrium impact of foreign shocks in different sectors. Our methods strike a balance between the clarity and rigor of the quantitative tradition and more model-robust statistical methods that "let the data speak." We begin by showing that in a large class of multi-sector small open economy models with sector-level gravity equations, all relevant information regarding external demand in a given country and sector is summarized in a single variable which we call *external firm market access*. This object can be estimated from trade data using standard techniques, and reflects the geography of sector-level trade costs and foreign demand. We then define an elasticity of welfare (equivalently, real per capita income) with respect to sector-specific external firm market access. This elasticity reflects the long run benefits of export market access in a sector, aggregating all the relevant microeconomic mechanisms and features of the initial state of the economy.

Expressing the impact of foreign market access in terms of welfare elasticities has three benefits. First, these general equilibrium elasticities can be estimated econometrically in a theory-consistent way. Second, these welfare elasticities are intimately connected with the welfare effects of trade and industrial policy. We derive sufficient statistics for the first-order welfare impacts of foreign tariffs, domestic export taxes and (under additional restrictions) domestic production subsidies as a function of the welfare elasticities. Third, since in fully specified models the welfare elasticities are functions of underlying model mechanisms and parameters, empirical estimates of these elasticities can be used to discriminate between different quantitative models. Our approach thus offers a theoretically consistent way to discipline the general equilibrium responses of structural models.

Econometric estimation of the welfare elasticities must overcome two primary challenges. The first is that there are hundreds of traded sectors and only a small sample of available real GDP per capita data. To reduce the number of parameters to be estimated, we employ a machine learning technique to group sectors into a small number of clusters based on their characteristics, such as their position in the production network and their factor intensities. We use the k-means clustering algorithm (MacQueen et al., 1967) along with 7 sectoral characteristics to group 233 manufacturing industries into 4 clusters: i) processing of raw materials, ii) complex intermediate inputs, iii) capital goods, and iv) consumer goods. We group agriculture and mining sectors into their own clusters and estimate the elasticity for each of the 6 clusters. The second challenge is common in cross-country regressions: omitted variables and endogeneity. We first provide formal conditions under which the welfare elasticities are identified by an OLS regression that fully conditions on the initial observables. The result exploits the typical invertibility properties of gravity models as well as an orthogonality assumption on unobserved contemporaneous domestic shocks. We rely on the fact that most countries are small in foreign markets, and measure the foreign shocks in such a way as to minimize any direct effect of domestic shocks on foreign variables. To deal with the high dimensionality of the control vector we employ the Post-Double-Selection method of Belloni et al. (2014b, 2017) to select a lower-dimensional set of "important" controls while maintaining consistency and uniformly valid inference.[1]

Our main empirical result is that the impact of external firm market access on real income varies significantly across sectors. The effects are largest in the capital goods and complex intermediate sectors, with the baseline point estimates indicating that a 1% increase in external firm market access causes a long run increase in real income of about 1.3% $\times$ export share in complex intermediates and 3% $\times$ export share in capital goods. All other sectors have estimated impacts that are close to zero, economically and statistically. The estimates directly imply that an increase in export market access in complex intermediates or capital goods is much more valuable (in the long run) compared to other tradable sectors, despite the initial partial equilibrium increase in export sales being the same by construction. There are thus quantitatively important sectoral differences in the general equilibrium impact of foreign shocks that are not captured by the foreign sales share.

We apply our estimates to three questions of substantive interest in international trade. First, we ask how much more valuable is it to have comparative advantage in some sectors compared to others. We compute the distribution of welfare impacts of a uniform increase in export market access in all countries and sectors (a "neutral globalization"). Countries that specialize in agricultural commodities, mining and simpler manufactured products experience little long-run welfare benefit from a neutral globalization, whereas open middle-income economies specializing in complex intermediate and capital goods (such as Hungary, South Korea, Taiwan and Malaysia) experience large increases in welfare. Thus, what you export matters.

Second, we examine the implications our our estimates for trade and industrial policy. We use the World Input-Output Database (WIOD) to calibrate an economy with global average initial state (i.e. trade, production, investment, factor and consumption shares) and, after calibrating a standard trade elasticity, apply our sufficient statistic formulas for the first-order gains from export

[1] In principle our approach could also be applied to estimate the impact of foreign supply shocks, but it turns out that there is much less variation across countries and time in our measured foreign supply shocks, and the estimated impacts are quite noisy. For these reasons, the paper focuses on estimating the effects of foreign demand shocks, while controlling for foreign supply shocks in the estimation. See Sections 3 and 4 as well as Appendix B.5 for more detail.

taxes/subsidies and production taxes/subsidies. We find that the average country clearly benefits from export *subsidies* to complex intermediates and capital goods relative to *laissez-faire*. Subsidies in other sectors are not welfare-improving; rather, taxes are. While export subsidies are generally not first-best policies (Bartelme et al., 2019) and have been controversial in both theory and practice, our results show that they indeed have a welfare rationale in some sectors. We stress that these results apply to all models in our class and do not rely on any particular mechanisms or functional form restrictions other than isoelastic foreign demand and a standard trade elasticity. Indeed, the main benefit of using the directly estimated welfare elasticities in this policy analysis is that we do not need to specify the exact economic mechanisms through which subsidies benefit the aggregate economy. The effects of sector-specific production subsidies are similar, with gains naturally even larger.[2] These results provide empirical evidence that countries benefit from actively reallocating economic activity toward some sectors rather than others.

In the third application we assess the implications of our estimates for quantitative trade models. We reverse the typical "micro to macro" approach in which partial equilibrium elasticities are inputs and general equilibrium responses are outputs. Instead, we use our estimates to go from "macro to micro," asking what model features and parameter values can match the empirically observed general equilibrium responses. We use a model of a small open economy with intermediate input linkages, endogenous capital accumulation and industry-level scale effects that are external to the firm and calibrate it to the global average initial state using WIOD data. It turns out that a parsimonious parameterization with only two structural parameters – the trade and scale elasticities – is quite successful in matching our welfare elasticity estimates. In particular, the best-fit model features an Armington elasticity of 3.2 (a common average estimate, see, e.g. Broda and Weinstein, 2006), and a scale elasticity of 0.29. The latter coincides almost exactly with the 0.30 mean sectoral scale elasticity estimated by Lashkaripour and Lugovskyy (2023). The model matches the heterogeneity in estimated coefficients purely through differing internal propagation of foreign demand shocks within the home economy, without the need for different trade or scale elasticities across sectors. Our main finding is that the combination of input-output linkages and scale effects is crucial for replicating the income responses to sector-specific foreign shocks; models that omit one or both of these two features are much less successful at matching the data. We conclude that these features (or similar mechanisms) are important for models to produce reliable long-run general equilibrium responses to shocks.

In summary, to reliably answer the question of how sectoral trade policies or other changes in market access affect real income, we must engage with the empirical relationships between trade variables and real income. While such an approach must clear several econometric hurdles, it is ultimately fruitful and informative for both our empirical understanding of the world, and as guidance for theoretical and quantitative modeling.

[2]With production subsidies, the caveat is we restrict attention to economies with Cobb-Douglas utility and production functions.

Related literature. Our paper contributes to the literature on trade and welfare, which would be impractical to review comprehensively here. A number of influential papers estimate the impact of overall openness on real income (e.g. Sachs and Warner, 1995; Frankel and Romer, 1999; Rodríguez and Rodrik, 2001; Redding and Venables, 2004; Rodrik et al., 2004; Wacziarg and Welch, 2008; Feyrer, 2009; Pascali, 2017; Feyrer, 2019). Our paper is closer to the work on export patterns and real income. Most of this literature considers only one characteristic of trade patterns at a time. Some examples include the natural resource curse (e.g. Sachs and Warner, 1999; Humphreys et al., eds, 2007; Sala-i-Martin and Subramanian, 2013), specialization in primary goods (Prebisch, 1959; Hadass and Williamson, 2003; Williamson, 2008), "high-income goods" (Hausmann et al., 2007; Jarreau and Poncet, 2012), the location in the product space (Hidalgo et al., 2007; Hidalgo and Hausmann, 2009), or skill intensity (Atkin, 2016; Blanchard and Olney, 2017).[3] We make three contributions to this literature. First, we consider multiple dimensions of trade patterns simultaneously, and let the data tell us which characteristics of exports matter. Second, we focus on exogenous foreign demand shocks, rather than the potentially endogenous specialization patterns themselves. And third, we use theory to link our estimates to policy-relevant objects.

In a sense, all of international trade theory is about the relationship between openness and welfare. Many mechanisms have been proposed for how the pattern of sectoral specialization can affect aggregate productivity, ranging from market failures (Haberler, 1950; Hagen, 1958; Bhagwati and Ramaswami, 1963; Krugman and Venables, 1995), to static (Graham, 1923; Chipman, 1970; Ethier, 1982; Kucheryavyy et al., 2020) and dynamic (Bardhan, 1971; Young, 1991) externalities, and to political economy (Tornell and Lane, 1999; Levchenko, 2013; Berman et al., 2017; Dippel et al., 2020). The wealth of potential theoretical mechanisms motivates the more data-driven approach in our paper.

Our paper also contributes to the quantitative and empirical literature on trade and industrial policy, particularly in environments involving scale economies. The quantitative literature (e.g. Bartelme et al., 2019; Lashkaripour and Lugovskyy, 2023) uses "micro-to-macro" to quantification. Empirical studies (e.g. Head, 1994; Fajgelbaum et al., 2020; Lane, 2021; Choi and Levchenko, 2021) have used partial equilibrium estimation of the effects of selected episodes of trade and industrial policy. Our paper is the first to use sufficient statistics of directly estimated general equilibrium responses of the economy to shocks. In addition to our focus on general equilibrium, the other novel feature of our method is that we do not study specific policy episodes but instead theoretically link policy outcomes to the response of the economy to foreign demand shocks. This indirect approach leverages the vastly more plentiful country and time variation in foreign demand shocks for estimation.

Our paper also makes contact with the general equilibrium quantitative trade tradition (e.g. Whalley, 1985; Deardorff and Stern, 1990; Eaton and Kortum, 2002, and the large literature that followed). Most closely related are quantifications of multi-sector models (e.g. Chor, 2010; Costinot

[3] The literature also considered variation on the import side, such as capital goods (Eaton and Kortum, 2001; Caselli and Wilson, 2004), skill-intensive goods (e.g. Nunn and Trefler, 2010), or intermediate inputs (e.g. Amiti and Konings, 2007; Kasahara and Rodrigue, 2008).

et al., 2012; Caliendo and Parro, 2015; Hsieh and Ossa, 2016; Levchenko and Zhang, 2016; Caliendo et al., 2017; Conte et al., 2021), as well as recent work on trade counterfactuals that apply across families of models (e.g. Arkolakis et al., 2012; Adão et al., 2017; Baqaee and Farhi, 2019; Allen et al., 2020). Relative to the latter group of papers, we focus more on understanding and quantifying sectoral heterogeneity in the responses to foreign shocks. Relative to the former, we invert the typical "micro-to-macro" approach to quantification. Adão et al. (2023) develop a test for validity of the general-equilibrium predictions of quantitative models, that is based on the exclusion restrictions employed in IV estimation of the partial-equilibrium responses of outcomes on the policy shocks. Our approach is complementary as we instead estimate the general-equilibrium responses directly. Section 2 discusses the relationship between our approach and the quantitative trade literature in further detail.

The rest of the paper is organized as follows. Section 2 lays out the model, while Section 3 discusses identification and estimation. Section 4 describes the data and Section 5 presents the results. Section 6 discusses the quantitative implications. The details of the derivations, data construction and manipulation, and additional empirical results are collected in the Appendices.

2. Theoretical Framework

We consider the steady state of a small open economy Home (H) in a world with N other countries indexed by i and n, K sectors indexed by k, and J factors of production indexed by j. Home is "small" in the sense that Home variables do not affect foreign aggregates, but it may be large in its own domestic market and faces downward sloping demand for its products in the international markets (as in Armington, 1969; Demidova et al., 2022). All results in this section are derived in Appendix A.

Technology and market structure. Each sector within each country produces a homogeneous good. Primary factors $L_{H,j}$ are in fixed supply and mobile across sectors. Input and output markets are competitive. Firms are infinitesimal and perceive a production technology that is constant returns to scale in their own inputs, but may feature external economies of scale both within and across sectors. We summarize the production technology in each sector by the unit cost function

$$c_{H,k} = c_{H,k}(\mathbf{w}_H, \mathbf{P}_H, \mathbf{L}_H; T_{H,k}), \tag{2.1}$$

where $\mathbf{w}_H$ and $\mathbf{P}_H$ are vectors of primary factor prices and intermediate goods prices, $\mathbf{L}_H$ is the matrix of primary factor allocations to sectors, and $T_{H,k}$ is an exogenous productivity shifter. We assume the unit cost function is continuously differentiable in all of its arguments. Trade across countries is subject to iceberg bilateral trade barriers $\tau_{in,k}$ to ship from from i to n in sector k.

Demand. The sector k composite good in country n is an Armington aggregate of varieties coming from different source countries,

$$Q_{n,k} = \left(z_{n,k}^{\frac{1}{\sigma_k}} \cdot q_{nn,k}^{\frac{\sigma_k - 1}{\sigma_k}} + \sum_{i \neq n} q_{in,k}^{\frac{\sigma_k - 1}{\sigma_k}} \right)^{\frac{\sigma_k}{\sigma_k - 1}}, \tag{2.2}$$

where $q_{in,k}$ is the quantity of sector k exported from country i to country n, and $z_{n,k}$ is an exogenous demand shifter that controls the degree of home bias in consumption. We assume that $Q_{n,k}$ can be used as both a final good and an intermediate input in country n. This assumption plus equation (2.2) implies that foreign demand for Home's exports in sector k takes the form

$$p_{Hn,k} \cdot q_{Hn,k} = (c_{H,k} \cdot \tau_{Hn,k})^{1-\sigma_k} \cdot \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}}, \tag{2.3}$$

where $p_{Hn,k} = c_{H,k} \cdot \tau_{Hn,k}$ is the price of the good in destination n, and $E_{n,k}$ and $P_{n,k}$ are total expenditure and the CES price index associated with equation (2.2) in country n.

All factor income in Home accrues to a representative consumer, who has homothetic preferences over sectoral quantity bundles $Q_{H,k}$.

Foreign shocks. We now define the key object underlying our analysis. By summing export revenues across foreign destinations, we get total foreign revenues as a function of Home costs and *External Firm Market Access* (FMA):[4]

$$\sum_{n \neq H} p_{Hn,k} q_{Hn,k} = c_{H,k}^{1-\sigma_k} \cdot \underbrace{\sum_{n \neq H} \tau_{Hn,k}^{1-\sigma_k} \cdot \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}}}_{FMA_{H,k}}. \tag{2.4}$$

External firm market access has three key features. First, $FMA_{H,k}$ is an exogenous sector k demand shifter from Home's perspective, since it depends only on foreign variables when Home is a small open economy. To interpret it, note that an x% change in $FMA_{H,k}$ implies an x% change in the quantity that foreigners demand when holding the price fixed. Second, any change in foreign demand affects the Home equilibrium only through its effects on $FMA_{H,k}$. $FMA_{H,k}$ has no bilateral dimension, and varies at the exporter and sector level. Third, $FMA_{H,k}$ can be estimated from trade data using conventional techniques, as detailed in Section 4.

To complete our description, we define *External Consumer Market Access* (CMA) by summing Home

[4]This concept differs from other definitions of market access (e.g. Redding and Venables, 2004) as it excludes domestic demand.

imports across source countries:

$$\sum_{n\neq H} p_{nH,k} \cdot q_{nH,k} = \frac{E_{H,k}}{P_{H,k}^{1-\sigma_k}} \cdot \underbrace{\sum_{n\neq H} (c_{n,k} \cdot \tau_{nH,k})^{1-\sigma_k}}_{CMA_{H,k}} . \tag{2.5}$$

From Home's perspective, $CMA_{H,k}$ is an exogenous supply shifter. An x% change in $CMA_{H,k}$ causes an x% change in Home's expenditure on foreign goods, holding total sectoral expenditure fixed. What drives this shift in expenditure is an exogenous change in import prices $c_{n,k} \cdot \tau_{nH,k}$. As with $FMA_{H,k}$, $CMA_{H,k}$ summarizes all relevant information about foreign supply.

Our paper focuses on estimating the general equilibrium impact of foreign demand shocks, but in principle the same techniques could also be used to estimate the impact of foreign supply shocks. In practice, limited statistical power due to the lower variability of the foreign component of $CMA_{H,k}$ in the trade data precludes reliable estimation of these effects.[5] As will become clear below, the $CMA_{H,k}$ will serve as important elements of the control set during estimation, but their impacts themselves will not be reported in the baseline analysis. Appendix B.5 provides a fuller discussion of the $CMA_{H,k}$, and reports estimates of their welfare elasticities.

Competitive equilibrium. A competitive equilibrium is a set of goods and factor prices and allocations such that firms and consumers optimize taking prices as given, factor and output markets clear and trade balances. We can characterize the equilibrium set as the set of solutions to a system of simultaneous equations in the unit cost and expenditure functions, factor prices and allocations, and trade balance. The equilibria are completely determined by the cost functions $c_{H,k}(\cdot)$, utility function $U(\cdot)$, the substitution elasticities σ_k and the exogenous variables (productivity and demand shifters, external firm and consumer market access, and primary factor supplies).

Our first order approach to estimation and counterfactual welfare analysis requires a locally unique and smooth mapping from the exogenous variables to equilibrium outcomes. There are no general results available on the equilibrium properties of this class of models, and we do not pursue them here.

First order welfare approximation. We now drop the H subscript. Using our assumption of homothetic preferences to equate real expenditure with welfare and the assumption of trade balance to equate nominal GDP with expenditure, we can write Home's welfare as

$$y = \frac{\sum_{k\in K} \mu_k c_k^{1-\sigma_k} \cdot \left(z_k \frac{E_k}{P_k^{1-\sigma_k}} + FMA_k \right)}{\mathbb{P}} . \tag{2.6}$$

[5]To be precise, the $\tau_{nH,k}$ component of $CMA_{H,k}$ may have elements that are controlled by the Home country, such as import tariffs or other inward trade barriers that may be endogenous. In principle this applies to $FMA_{H,k}$ as well, although most countries do not intentionally impose barriers to exports. While the component of $CMA_{H,k}$ that depends only on foreign variables can be extracted from the trade data, it has very low cross-country variability relative to the strictly foreign component of $FMA_{H,k}$.

Welfare thus corresponds to the real income of primary factors, computed as the nominal income divided by the aggregate consumption price index $\mathbb{P}$. Nominal primary factor income in sector k is in turn the value of sectoral gross output times the share of value added in gross output μ_k.

Equation (2.6) highlights the two ways the FMA_k's affect Home's welfare. There are direct effects through changes in foreign sales when FMA_k changes. There are also indirect effects on domestic prices and quantities as Home producers and factor owners alter their production plans and consumers alter their consumption patterns in response to these external shocks. A unique and smooth mapping from domestic and foreign shocks to equilibrium quantities implies that, to a first order, the total effect of a set of log changes in foreign demand on log welfare is

$$d \ln y \approx \sum_k \delta_k^x \cdot \left[\lambda_k^x d \ln FMA_k\right], \tag{2.7}$$

where λ_k^x is the initial share of total sales accounted for by exports in sector k.

Interpretation. The elasticities δ_k^x capture the first order general equilibrium response of real income to exogenous changes in foreign demand in different industries. As is evident from equation (2.4), foreign demand shocks in this environment can come from a variety of sources such as foreign taste or productivity shocks, changes in aggregate foreign expenditure, iceberg trade costs, or foreign trade policy.[6] To interpret these elasticities, consider the following thought experiment. Two small open economies, initially identical in every respect, experience a different pattern of foreign demand shocks. Specifically, suppose economy A sees a 1% increase in foreign demand in industry 1 while economy B sees a 1% increase in foreign demand in industry 2. Which economy will experience a greater change in real income? Assuming both industries have the same initial export sales shares, the answer is the economy that gets the shock to the industry with the highest δ_k^x.

The δ_k^x depend on the structural parameters of the model as well as the initial state of the economy. The variation in the δ_k^x's across sectors is determined by sector-level structural parameters, such as the trade and scale elasticities, as well as other sectoral characteristics such as openness to trade, the position in the input-output network, and the final use of the industry (consumption vs. investment). They capture the welfare impacts due to changes in the terms of trade as well as from changes in resource allocations across sectors, which have first order impacts in distorted economies. Appendix A.3 presents some simple examples and a fuller discussion, while Section 6 and Appendix A.4 detail a more realistic quantitative environment featuring scale economies, intermediate goods and endogenous capital accumulation, and present an analytical solution for δ_k^x. The formula in this more elaborate model makes it clear that the δ_k^x depend on the full structure of the economy in

[6]The response of welfare to changes in iceberg trade barriers relates to δ_k^x via the formula

$$\frac{\partial \ln y}{\partial \ln \tau_k^x} = (1 - \sigma_k)\lambda_k^x \cdot \delta_k^x.$$

complex ways. This very complexity provides one of the primary motivations for our more agnostic, data-driven approach to estimation and quantification.

Policy implications. There are close connections between the welfare effects of export demand shocks and the welfare effects of trade and industrial policies. Intuitively, the δ_k^x are informative about the consequences of changing the size of sector k, which is also the key objective of trade and industrial policy interventions. The effects of policies differ from δ_k^x because policies have an impact on government revenue or expenditure. Accounting for the fiscal effects and their incidence in the economy generally requires additional information. However, there are several cases of interest in which the δ_k^x, along with trade elasticities and observable shares from the data, are sufficient statistics for the welfare impact of policies.

The simplest case is that of a foreign tariff t_k^f imposed on the value of Home exports. Foreign tariffs are frequently the target of Home government policy during trade negotiations and WTO disputes. Changes in foreign tariffs do not result in changes in Home tariff revenue, and hence the welfare effect of a small change in $(1-t_k^f)$ for all trading partners is simply

$$\frac{d\ln y}{d\ln\left(1-t_k^f\right)} = \lambda_k^x \sigma_k \delta_k^x. \tag{2.8}$$

The effect of changes in foreign tariffs is identical to foreign demand shocks, with the foreign demand elasticity simply translating the change in tariff to the size of the demand shock.

A more interesting case is that of a Home export tax or subsidy. Home export taxes differ from foreign tariffs because they affect Home government revenue. We consider a change in the export tax t_k^x, defined so that the firm's revenue from each unit sold to foreigners at price p_k^f is $\tilde{p}_k = (1-t_k^x)p_k^f$, at an initially zero-tax equilibrium.[7] We assume the government uses lump-sum taxation to balance its budget. Let Y be aggregate gross output, V be aggregate final expenditure, e_k be sector k's share of final expenditure, and θ_k^d be the share of Home's expenditure on k that is sourced domestically. Then the welfare effect of a small export subsidy is

$$\frac{d\ln y}{d\ln\left(1-t_k^x\right)} = \lambda_k^x \left(\delta_k^x \sigma_k - \frac{Y}{V} - \sum_{k'} \delta_{k'}^x e_{k'} \theta_{k'}^d\right). \tag{2.9}$$

The first term is the positive GE effect of the demand shock, excluding the revenue effects of the subsidy, and is identical to the effect of a change in the foreign tariff (2.8). The second term, $-\lambda_k^x Y/V$, is the direct effect of the taxes necessary to fund the subsidy. The third term, $-\lambda_k^x \sum_{k'} \delta_{k'}^x e_{k'} \theta_{k'}^d$ is the terms of trade effect of the lump sum tax. The tax shrinks the size of the domestic market, inducing domestic firms to sell abroad which lowers the terms of trade. This negative TOT effect of lump sum

[7]Our approach can be modified to accommodate initial equilibria with non-zero taxes. The formula below will be accurate unless tariff and/or export tax revenue is very large.

taxation is stronger in sectors with a larger proportion of sales in the domestic market, θ_k^d. An export subsidy will be welfare improving only if the positive effects of the demand shock are large enough to cover the cost of the subsidy and its negative TOT effects.

These formulas can be implemented with estimates of σ_k and δ_k^x along with widely available data on trade, production and consumption. They offer a method to evaluate policies that is fully grounded in both theory and empirics while utilizing far fewer assumptions than more standard fully structural approaches. The general approach can also be extended to relate the δ_k^x to the welfare effects of import tariffs and production subsidies, but some of the appealing simplicity of equation (2.9) is lost. Unlike foreign tariffs or Home export taxes, Home import tariffs or production subsidies directly affect Home consumer and producer prices in a way that is not captured by the δ_k^x. Accounting for these effects requires knowledge of sectoral elasticities of substitution in consumption and production (when there are domestic intermediate goods). As a further illustration of our approach in a special case of interest in the quantitative trade literature, Appendix A derives a sufficient statistic formula for the welfare effects of production subsidies under the assumption of Cobb-Douglas production and upper-tier utility functions (equation A.17). We implement these formulas using our estimates in Section 6 below.

This analysis does not establish the microfoundations under which export and production subsidies/taxes are welfare-improving, as the production structure is not fully specified. The literature on trade and industrial and policies is vast, and it would not be practical to revisit all the possible rationales for policy interventions.[8] Rather, what we show is that estimates of δ_k^x can be used to infer the welfare benefits of certain trade and industrial policies, without fully specifying the model and microfounding the sources of departures from efficiency. However, it is interesting to note some recent theoretical results that provide a rationale for export subsidies in the presence of sector-level scale effects (Campolmi et al., 2014; Bagwell and Lee, 2020; Lashkaripour and Lugovskyy, 2023). Section 6 shows that such scale effects provide a simple and powerful structural rationale for our empirical findings.

Comparing our approach to the literature. The dominant approach to quantifying general equilibrium responses to shocks in the trade literature would be to complete the description of the model, which here would amount to specifying functional forms for $c_{H,k}$ and the utility function $U(\cdot)$. Disciplining the model with data takes the form of estimating structural parameters using the partial equilibrium relationships implied by the model.[9] General equilibrium responses to shocks are then computed using these estimated parameters, the initial shares and the model structure, but are not themselves directly disciplined by the data. A very incomplete list of recent examples includes Eaton

[8]See Harrison and Rodríguez-Clare (2010) and the Handbook of Commercial Policy (Bagwell and Staiger, eds, 2016).

[9]By "partial equilibrium" we mean estimation approaches that utilize a strict subset of the model equations to estimate any given parameter. An example would be the estimation of trade elasticities in a gravity model using only the implied relationship between relative trade costs and relative trade shares. See the Handbook Chapter by Head and Mayer (2014) for a discussion of why the trade elasticity estimates are partial equilibrium, and the usage of the term "partial" to refer to these estimates.

and Kortum (2002), Caliendo and Parro (2015), Bartelme et al. (2019) and Fajgelbaum et al. (2020). Alternatively, another strand of the literature recovers structural parameters by computing or simulating the general equilibrium response of the model to shocks, and comparing model moments with data. By construction, this method produces parameter estimates that give the best in-sample fit of the chosen model to the (targeted) moments of the endogenous variables in general equilibrium. Examples in the international trade literature include Yi (2003), Fieler et al. (2018), Allen et al. (2020), Adão et al. (2020a), and Adão et al. (2020b). Crucially, both approaches impose a fully specified model on the data.

Instead, this paper estimates the GE responses δ_k^x econometrically. As such, the δ_k^x are not generally structural parameters. While our strategy has more in common with the latter quantification approach than the former, we impose less structure than would be required to use a fully specified model for estimation. Rather than explicitly modeling and quantifying each aspect of the underlying structure of the economy, we recover the reduced form elasticities that are directly relevant to the relationship between foreign demand shocks and welfare. One clear advantage over methods that require a more complete specification of the model is that our estimates are robust to model uncertainty within the wide class of trade models encompassed by our framework. On the other hand, compared to the reduced-form empirical literature on trade patterns and income (summarized in, e.g. Lederman and Maloney, 2012) we provide enough structure to enable clear interpretation, precise conditions for identification in terms of model primitives, and local counterfactuals.

There are some costs to achieving robustness to model uncertainty. First, completely specifying a (correct) model permits more efficient estimation of the relevant parameters. Second, a structural model reveals the economic mechanisms that generate the results. Third, a fully specified model can be solved in its nonlinear form, which enables more accurate counterfactuals with respect to large shocks. We view our strategy as a complement to the fully structural approach. In particular, our estimates can be used as moments to be targeted by models, either for estimation or as out-of-sample validation. Section 6 implements one such exercise, by evaluating the ability of a series of quantitative trade models to match our estimates under different parameter values.

We briefly discuss some isomorphisms and extensions. We use the competitive Armington environment in the theoretical framework to maximize clarity. The truly crucial assumptions are gravity in trade flows, homothetic upper tier preferences and the unique and smooth equilibrium mapping. Models with alternative micro-foundations for gravity, such as those based on Eaton and Kortum (2002), Krugman (1980), or Melitz (2003) with a Pareto distribution for productivity, will be isomorphic to our model in the sense that they have a first order approximation of the same form as equation (2.7) and the same interpretation of the market access elasticities δ_k^x. In addition, while the model described above is static, equation (2.7) is also valid for small shocks in the steady state of a dynamic economy with some reproducible factors of production. Section 6 presents one example of a dynamic model with such a steady state representation.

3. Identification and Estimation

3.1 Identification

We now consider identification of the δ_k^x in equation (2.7). We will view data as a collection of small open economies (indexed by i) observed over a number of steady states indexed by t. For each economy and point in time, we observe real income $y_{i,t}$ and a set of additional equilibrium outcomes $\{x_{ik,t}\}$ (e.g. factor prices, trade shares, foreign demand and supply shifters, expenditure shares, etc). While foreign demand and supply shifters ($FMA_{ik,t}$ and $CMA_{ik,t}$) are not directly observable, they can be consistently estimated from trade data (see Section 4) and we will treat them as observable for the rest of this section.

For a given small economy, and suppressing i, t subscripts, the log change in real income between two steady states can be written as a function of the trade-share-weighted log changes in FMA_k, other changes in exogenous variables $\{d\ln a_k\}$, and the initial fundamentals of the economy $\{Z_k\}$:

$$\begin{aligned} d\ln y &= F\left(\{\lambda_k^x d\ln FMA_k\}, \{d\ln a_k\}, \{Z_k\}\right), \\ \{d\ln a_k\} &= \{\{\lambda_k^m d\ln CMA_k\}, \{d\ln T_k\}, \{d\ln z_k\}, \{d\ln L_j\}\}, \\ \{Z_k\} &= \{\{T_k\}, \{z_k\}, \{L\}, \{FMA_k\}, \{CMA_k\}\}. \end{aligned} \tag{3.1}$$

We proceed in several steps. First, we assume that the equilibrium mapping $G(\{Z_k\}) \to (\{x_k\})$ between all relevant exogenous variables and observable equilibrium outcomes is locally smoothly invertible (up to a normalization); this implies that the observables "reveal" all relevant aspects of the initial state of the economy. This is a typical property of quantitative trade models, that justifies the widespread use of the "exact hat algebra" (Dekle et al., 2008) to conduct counterfactual analysis. Not every exogenous variable needs to be identified, only the combinations of parameters that are sufficient to compute counterfactual changes. Using this assumption, we can rewrite (after redefining $F(.)$) the log change in real income as

$$d\ln y = F\left(\{\lambda_k^x d\ln FMA_k\}, \{d\ln a_k\}, \{x_k\}\right). \tag{3.2}$$

Second, we consider the joint distribution of the domestic shocks ($\{d\ln T_k\}$, $\{d\ln z_k\}$, and $\{d\ln L\}$), which may depend on the initial state of the economy. We decompose each domestic shock into its conditional expectation with respect to the initial state and a residual term that satisfies $E\left[\varepsilon_k | \{x_k\}\right] = 0, \forall k$, which allows us to write (again redefining $F(.)$)

$$d\ln y = F\left(\{\lambda_k^x d\ln FMA_k\}, \{d\ln \tilde{a}_k\}, \{x_k\}, \{\varepsilon_k\}\right), \tag{3.3}$$

where $\{\tilde{a}_k\}$ is a set of observable shocks (e.g. $\{\lambda_k^m d\ln CMA_k\}$). Finally, we apply Taylor's Theorem to all variables in equation (3.3) and re-introduce the i, t subscripts to derive our log-linear estimating equation

$$d\ln y_{i,t} \approx \kappa + \sum_k \delta_k^x \cdot \left[\lambda_{ik,t}^x d\ln FMA_{ik,t}\right] + \zeta d\ln \mathbf{a}_{it} + \eta \mathbf{x}_{i,t} + \varepsilon_{i,t}, \tag{3.4}$$

where κ reflects the initial point of approximation, $d\ln \mathbf{a}_{it}$ is a vector of observable shocks, $\mathbf{x}_{i,t}$ is the vector of initial observables and $\varepsilon_{i,t}$ combines the first-order effects of domestic shocks with the approximation error.

In order to interpret the OLS estimates $\hat{\delta}_k^x$ as the causal effect of foreign demand shocks on real income, we need the additional assumption that

$$E[\varepsilon|\{\lambda_k^x d\ln FMA_k\}, d\ln \mathbf{a}, \mathbf{x}] = 0. \tag{3.5}$$

Recalling that $E[\varepsilon|\mathbf{x}] = 0$ by construction, assumption (3.5) thus states that adding foreign demand shocks to the information set does not help predict the component of the unobserved domestic innovations that are orthogonal to the initial state.

Before considering potential violations of this assumption, it is useful to briefly discuss the sources of variation that will identify the δ^x in equation (3.4). The first source of identifying variation comes from comparing the growth rates of countries with similar initial export baskets (and other initial conditions) that experience different foreign shocks due to different geographic positions (e.g. South Korea vs. Germany). The second source of variation comes from comparing initially *dissimilar* countries that experience similar foreign shocks (e.g. South Korea vs. Taiwan). This comparison is incomplete because we must also account for the possibility that the initial state itself predicts growth. However, with sufficiently many different countries experiencing different foreign shocks we can separate the predictive power of the initial state itself from the way it affects growth by mediating the impact of foreign demand shocks.

3.1.1 Threats to Identification

The error term in (3.4) contains the components of domestic productivity growth $\{d\ln T_k\}$, factor supply growth $\{d\ln L\}$, and changes in domestic tastes $\{d\ln z_k\}$ that are orthogonal to initial observables. The identifying assumption (3.5) is that foreign demand shocks are uncorrelated with the orthogonalized domestic shocks. In practice, the foreign shocks will be extracted from gravity specifications estimated on the bilateral trade matrices (see Section 4.3 below). With this in mind, there are three principal ways assumption (3.5) could be violated.[10]

Domestic policies appearing in $d\ln FMA$. The first concern is that domestic policies correlated with domestic shocks may affect foreign market access. As written, $d\ln FMA$ in (2.4) includes iceberg trade

[10] A fourth issue, which we do not discuss in detail, is our reliance on a first-order approximation, which requires countries to be relatively similar and shocks to be relatively small. In principle our methods could be extended to incorporate higher-order terms or different points of approximation, although in practice we lack sufficient data to push very far in these directions. We do explore the heterogeneity in the δ_k^x across developed and developing countries in Appendix C.

costs of exporting from the Home economy $\tau_{Hn,k}$, which may be determined in part by the domestic policymakers (e.g. export taxes). We address this by estimating foreign shocks with a leave-one-out strategy that uses only foreign data. As such, the estimated $d \ln FMA$s reflect only the components of $\tau_{Hn,k}$ determined by foreign variables and exogenous geographic characteristics.

Small country assumption. The second concern is that domestic shocks may affect foreign incomes, prices and production in international general equilibrium, creating a structural correlation between domestic shocks and the estimated foreign demand shocks. Although in principle a variety of complex interactions and feedback mechanisms are possible, the most straightforward violation of the small country assumption induces a negative structural correlation between the regressors and the error term in equation (3.4). Following a positive domestic aggregate productivity shock, a large Home economy will lower the price of its goods on international markets and induce its trading partners to increase imports from Home at the expense of imports from third countries. Since our leave-one-out strategy drops Home's export sales from estimation of foreign demand, only the decline in imports from third countries will be reflected in Home's estimated $d \ln FMA$s. Thus the international general equilibrium impact of a positive Home productivity shock is to lower its estimated foreign demand shock, creating a negative correlation between the regressors and the error term that tends to bias the coefficients towards zero.[11]

The quantitative relevance of this mechanism is limited to country-sector combinations that have substantial international market share. Section 5 assesses the robustness of our results to international general equilibrium forces in two ways. First, we drop trade partners for whom the Home country is a large source of imports from the calculation of the foreign demand shocks. Second, we drop countries that have large world export shares in individual industry clusters. These robustness checks directly address the effect described above whereby Home supply shocks affect price indices in its export destinations and therefore measured foreign demand shocks, as well as other potential possibilities such as endogenous foreign supply responses.

Spatial correlation of shocks. The final concern is that domestic shocks may be spatially correlated, leading to a positive statistical association between the measured foreign demand shocks (which depend on domestic shocks in neighboring countries) and Home's unobserved domestic shocks. To give an example, if growing economies tend to have high investment rates then their demand for imports of capital goods will increase. We may then observe a geographic cluster of countries with both high growth and large increases in foreign demand for capital goods, and mistakenly infer that the foreign demand for capital goods causes high growth. Other examples leading to bias in either direction could also be constructed.

[11]This argument applies directly to sector-neutral aggregate productivity shocks. Suppose instead that productivity growth is biased towards a subset of sectors. The international substitution effect described above still leads to a negative correlation between the error term and foreign demand shocks in the faster-growing sectors, but in slower-growing sectors the correlation between the error terms and the foreign shocks is ambiguous and depends on the internal structure of the economy.

We undertake a number of checks to rule out the possibility that our results are driven by spatial correlation in domestic shocks. First, we drop neighboring countries from the calculation of $d \ln FMA$. Second, we control for neighboring countries' TFP growth directly in estimation. Third, we include the *unweighted* foreign demand shocks $d \ln FMA$ as additional controls. Recalling that the regressors of interest are the foreign demand shocks times the initial export revenue share ($\lambda^x d \ln FMA$), adding the unweighted $d \ln FMA$ controls directly for the possibility that the foreign demand shocks reveal information about domestic productivity growth via spatial correlation. The effect of these additional control is to strip out our first source of identifying variation described above (initially similar economies, different foreign shocks), leaving only the second (dissimilar countries experiencing the same foreign shocks). This second source of variation is robust to the most plausible ways in which spatial correlation in shocks could bias estimates.

3.2 Estimation

Estimation of equation (3.4) by OLS is consistent for the δ_k^x under the assumptions of smooth invertibility plus the exogeneity condition (3.5). However, in practice estimation must confront the scarcity of medium- or long-run country growth rates relative to the number of distinct industries that are observed in the trade data. This imbalance raises two related but distinct issues: i) the large number of parameters of interest $\{\delta_k^x\}$, and ii) the large number of controls. We discuss our methods for handling each of these challenges below.

We lack sufficient data to precisely estimate each δ_k^x separately for highly disaggregated industries. We reduce the number of parameters by grouping industries into a smaller number mutually exclusive clusters, and estimating a single elasticity per cluster. Formally, we group industries into G clusters and estimate the equation

$$d \ln y_{i,t} \approx \kappa + \sum_{g \in G} \delta_g^x \cdot \left[\lambda_{ig,t}^x d \ln FMA_{ig,t} \right] + \zeta d \ln \mathbf{a}_{it} + \eta \mathbf{x}_{i,t} + \varepsilon_{i,t}, \tag{3.6}$$

where

$$\lambda_{ig,t}^x d \ln FMA_{ig,t} \equiv \sum_{k \in g} \lambda_{ik,t}^x d \ln FMA_{ik,t}. \tag{3.7}$$

The cluster-level elasticities δ_g^x can be interpreted as weighted averages of the industry-level elasticities, with the weights reflecting the variance of the industry-level shocks and their covariance with one another. Note that we do not assume that the industries within each cluster are identical to one another; each industry maintains its separate foreign demand shock and initial export sales share.

While it is possible to estimate and interpret cluster-level elasticities for any grouping scheme, both estimation and interpretation are facilitated by choosing clusters of industries that share similar characteristics. As discussed in Section 2, theory predicts that the δ_k^x are determined by sectoral characteristics such as position in the production network, factor intensities, and so on. Clustering industries with similar characteristics thus has the benefit of minimizing intra-cluster heterogeneity in the δ_k^x, which increases efficiency in estimation. It also helps locate the ultimate sources of variation in the cluster-level elasticities in terms of the shared characteristics of industries in each cluster. The lower-digit groupings of conventional industrial classification schemes (e.g. SIC, NAICS) are not generally constructed based on the relevant industry features, so we construct our own groups based on a number of potentially relevant characteristics using machine learning techniques. Section 4 describes the clustering procedure in detail.

The second issue we need to address is the large set of control variables, which includes the variables from the initial state (e.g. trade shares) as well as contemporaneous foreign supply shocks. We deal with this problem by using the Post-Double-Selection estimator developed by Belloni et al. (2014b, 2017). This approach involves selecting a subset of "important" controls by regressing each dependent and independent variable on the full set of potential controls using an estimator that sets some or all of the coefficients to zero (e.g. LASSO). The selection is "double" in that the controls are selected based on their correlations with both the dependent and independent variables. The union of the sets of controls that are thus selected (i.e. have non-zero coefficients) in each regression then form the control set for an OLS regression of the dependent variable on the independent variables, including the selected controls.

Belloni et al. (2014b) show that this estimator is consistent and asymptotically normal, with the usual standard errors generating uniformly valid confidence intervals, under conditions that are quite plausible in our setting. The most important condition is that the true control vector admits an *approximately sparse* representation in the sense that the true control function can be well-approximated by a function of a subset of the controls.[12] This condition does not require that the control function exhibit true sparsity, only some combination of true sparsity, many small coefficients, and high correlation between controls. These conditions seem reasonable in our application. We discuss our implementation of the Post-Double-Selection estimator in detail in Section 5 and in Appendix B.4.

4. Data, Clustering and Foreign Shock Estimation

This section briefly summarizes our data sources and measurement strategy. Appendix B collects the detailed descriptions of all steps.

4.1 Data

Our empirical implementation requires data on (i) real income per capita, (ii) sectoral bilateral trade flows and trade barriers, and (iii) sectoral characteristics. Income per capita is sourced from the Penn World Tables 9.0, computed as the real GDP at constant national prices divided by population.[13] We

[12]We refer the reader to Belloni et al. (2014a), Belloni et al. (2014b) and Belloni et al. (2017) for additional details and regularity conditions.

[13]We acknowledge substantial empirical challenges in measuring real per capita income. We first note that our real per capita income is always in log-differences, sidestepping the many thorny issues in the computation of Purchasing

drop countries with population less than 2 million from our sample. Per capita income growth is computed at 10-year intervals for a maximum of 5 10-year growth rates per country (there are some missing values).

The bilateral trade flow data at the 4-digit SITC Rev 2 level come from the UN COMTRADE Database. We convert the trade data from the SITC to the 1997 NAICS classification. Appendix B.1 describes the construction of the concordance in detail. All in all, the 784 4-digit SITC items are matched to 268 NAICS sectors. Among them are 233 manufacturing, 26 agricultural, and 9 mining sectors. Geographic variables (bilateral distance and contiguity measures) come from CEPII. The final sample covers 127 countries listed in Appendix Table A4, 268 sectors and 5 decades from 1965 to 2015, with a total of 548 10-year GDP growth rate observations.[14]

A machine learning algorithm groups 233 manufacturing sectors into clusters based on their sectoral characteristics. While our set of sectoral characteristics is to some extent dictated by data availability, we assemble a collection of indicators tied to mechanisms prominent in the economic growth literature, such as physical and human capital, position in the input network, and contract intensity. We use data from the United States to measure the sectoral characteristics, since data at a comparable 4-digit level of disaggregation are not available for a large sample of countries. We collect data on 7 features: investment sales shares, intermediates using shares, intermediates sales shares, 4-firm concentration ratios, skilled worker shares, physical capital intensities, and the contract intensity of inputs. Sectoral characteristic variables are collected from various data sources with similar but not always identical industry classifications. We convert all of them to the 1997 NAICS classification.

Our measures of the investment sales shares, intermediates sales shares and intermediate using shares are based on data from the 1997 Benchmark Detailed Make and Use Tables. The investment sales share is computed as the ratio of spending on sector k for investment purposes to the the total gross output of sector k. Thus, this variable captures in a continuous way the extent to which sector k produces capital goods. Similarly, intermediates sales and using shares of gross output capture the extent to which sector k is a large producer or user of intermediate goods, respectively. The four-firm concentration ratios are sourced from the 2002 Economic Census. The skilled worker shares are calculated as the share of workers in sector k that have a bachelor degree or higher, and are computed based on data from the 2000 American Community Survey. The capital intensity variable is measured as 1 minus the labor share of value added (payroll), based on the NBER-CES Manufacturing Industry Database. The contract intensity of a sector is measured as the fraction of a sector's inputs that need

Power Parities. So we never have to compare the theoretical object $\mathbb{P}$ in our model to how the PWT constructs price *levels*. All we need is that the *changes* in the consumption price indices over time are computed correctly by the national statistical agencies. Here, we believe we are on somewhat firmer ground, as, for example, the Laspeyres price index is a first-order approximation to the growth rate of the ideal CES price index. Chain-weighted price indices provide even better approximations to the ideal price index change, as they (partly) account for the second-order substitution terms. Burstein and Cravino (2015) show that the procedures used by the national statistical agencies to measure real GDP will correctly pick up welfare changes due to trade shocks to first order.

[14]Conceptually, the shares $\lambda^{x}_{ik,t}$ have aggregate gross output in country i in the denominator. As gross output is not available for our sample of countries and decades, we proxy for gross output by total nominal GDP in US dollars (the units in which the trade data come) times 2, since the share of value added in gross output is commonly calibrated to 0.5.

relationship-specific investments, and comes from Nunn (2007). We use the version of this variable that measures the fraction of inputs not sold on organized exchanges and not reference priced to capture the importance of relationship-specific investments in a sector.

4.2 K-means Clustering

We use the k-means clustering algorithm (MacQueen et al., 1967) to group sectors into clusters based on the 7 characteristics described above. Sectors are assigned to clusters based on their characteristics so as to minimize the within-cluster sum of squared deviations from the cluster mean. The k-means algorithm works as follows: given M manufacturing sectors, each with a vector of $\mathcal{N}$ different sectoral characteristics, $x(k) \in \mathbb{R}^{\mathcal{N}}$, $k = 1, \ldots, M$, assign the M sectors into G clusters. The G clusters are indexed by $g = 1, 2, \ldots, G$.

1. Initialize cluster centroids $m_1, m_2, \ldots, m_G \in \mathbb{R}^{\mathcal{N}}$ for each cluster.

2. Assign each sector k to the cluster whose centroid is closest to $x(k)$. The cluster assignment is $c(k) \in \{1, 2, \ldots, G\}$,
$$c(k) = \underset{g \in \{1,\ldots,G\}}{\operatorname{argmin}} ||x(k) - m_g||^2.$$

3. Replace cluster centroid m_g by the coordinate-wise average of all points (sectors) in the gth cluster,
$$\hat{m}_g = \frac{\sum_{k=1}^{M} \mathbf{1}(c(k) = g) \cdot x(k)}{\sum_{k=1}^{M} \mathbf{1}(c(k) = g)}.$$

4. Iterate on steps 2 and 3 until convergence.

We use the "k-means ++" algorithm proposed by Arthur and Vassilvitskii (2007) to choose the initial values for the k-means clustering algorithm, and do extensive checks using alternative starting points. Following standard practice, we normalize the values of each characteristic to have zero mean and unit variance.[15]

The algorithm above requires a choice of the number of clusters. There is no unambiguously optimal method for choosing the number of clusters, although there are a number of conceptually similar approaches based on maximizing various measures of cluster fit. We use the silhouette width (Rousseeuw, 1987) as our measure of cluster fit. Loosely speaking, the silhouette width measures the similarity of industries within a cluster relative to industries in the nearest cluster. A good clustering scheme will maximize the average silhouette width while minimizing the number of sectors near the boundaries. The silhouette analysis suggests that either 4 or 5 are good values for the number of clusters. Appendix B.2 reports the results of the silhouette analysis along with a fuller discussion. In the interest of parsimony we choose to group the 233 manufacturing industries into 4 clusters in our baseline analysis, and show that our results are insensitive to this choice.

[15]This step is prudent because k-means clustering is not invariant to the scale used to measure the characteristics.

Table 1: Summary Statistics of Clusters in Manufacturing

	Cluster					
	1	2	3	4	Mean	Std. Dev.
Investment Share	0.00	0.05	0.52	0.04	0.13	0.22
Intermediates, Using	0.78	0.58	0.65	0.66	0.66	0.16
Intermediates, Sales	0.84	0.70	0.27	0.28	0.57	0.31
Concentration Ratio	0.47	0.27	0.38	0.56	0.40	0.21
Skill Intensity	0.32	0.28	0.35	0.36	0.32	0.13
Capital Intensity	0.68	0.55	0.54	0.70	0.61	0.10
Contract Intensity	0.26	0.56	0.73	0.52	0.51	0.22
Number of industries	60	84	47	42		
Trade share	0.33	0.26	0.23	0.11		
Label	Raw Materials Processing	Complex Intermediates	Capital Goods	Consumer Goods		
Abbreviation	RAW	INT	CAP	CONS		

Notes: This table reports the summary statistics of the sectoral characteristics among the sectors selected into each cluster. The last two columns report the mean and standard deviations of those characteristics among all manufacturing sectors. The row "Number of industries" reports the number of sectors in each cluster, and "Trade share" reports the fraction of world trade accounted for by sectors in that cluster. The bottom panel lists the intuitive labels of the clusters, as well as 3-letter abbreviations. Both are heuristic and assigned by the authors.

Table 1 summarizes the characteristics of the 4 clusters. Since it turns out that each cluster has some salient features that distinguish it from others, we name the clusters based on these key features. The sectors in cluster 1 have the highest intermediate sales and using shares, and lowest contract intensity. We label these sectors "raw materials processing" sectors. These sectors typically involve the first stage of turning raw materials into manufactured goods. Cluster 2 has the second-highest intermediate sales shares (after cluster 1), but considerably higher contract intensity than cluster 1. We thus label it "complex intermediates." Cluster 3 stands out most clearly as capital goods, with an average investment sales share of 0.52 compared to investment shares ranging from 0.00 to 0.05 in the other clusters. Cluster 4 has a low intermediate sales share and a negligible average investment sales share. Thus we label it "consumer goods." Appendix Table A3 lists the 3 most representative sectors in each cluster, defined as those closest to the cluster centroid. As we do not have information on these characteristics for non-manufacturing sectors, we group all agricultural sectors into cluster 5, and all mining sectors into cluster 6. In total, the 268 sectors are grouped into 6 clusters.

4.3 Estimation Strategy for $FMA_{ik,t}$

To obtain $FMA_{ik,t}$ for country i sector k at time t, we estimate structural sector-specific gravity equations using the matrix of sectoral bilateral trade flows at decadal intervals.[16] For a given sector k

[16]To reduce measurement error, we use three-year averages of the trade flows. For instance, to estimate the vector of $FMA_{ik,t}$ for $t = 1965$, we use the average trade flows for 1964, 1965, and 1966.

at time t, the gravity equation (2.3) can be rewritten as

$$\lambda_{ink,t} = c_{ik,t}^{1-\sigma_k} \cdot P_{nk,t}^{\sigma_k - 1} \cdot \tau_{ink,t}^{1-\sigma_k}, \tag{4.1}$$

where $\lambda_{ink,t}$ denotes the share of n's expenditure on sector k that is sourced from country i. Since we do not observe domestic trade flows, we calculate $\lambda_{ink,t}$ as the share of import expenditure. We model the bilateral resistance term $\tau_{ink,t}^{1-\sigma_k}$ as a function of geographic distance and contiguity with sector-time-specific coefficients, leading to our empirical specification

$$\lambda_{ink,t} = \kappa_{ik,t}^{x} \cdot \kappa_{nk,t}^{m} \cdot Distance_{in}^{\zeta_{kt}} \cdot \exp\left(\xi_{kt} \cdot Contig_{in}\right) \cdot \varepsilon_{ink,t}, \tag{4.2}$$

where $\kappa_{ik,t}^{x}$ is the exporter fixed effect, $\kappa_{nk,t}^{m}$ is the importer fixed effect, and ζ_{kt} and ξ_{kt} are the distance and common border coefficients. We estimate the non-linear equation (4.2) using the Poisson Pseudo-Maximum Likelihood (PPML) method proposed by Silva and Tenreyro (2006) and Eaton et al. (2012), separately for 268 sectors and each of the 5 decades spanning 1965-2015.

We use our estimates from equation (4.2) to construct the external market access terms as follows:

$$FMA_{ik,t} = \sum_{n \neq i} E_{nk,t}(i) \cdot \kappa_{nk,t}^{m} \cdot Distance_{in}^{\zeta_{kt}} \cdot \exp\left(\xi_{kt} \cdot Contig_{in}\right) \tag{4.3}$$

where $E_{nk,t}(i) \equiv \sum_{i' \neq n,i} E_{i'nk,t}$ is total importer n expenditure in k at time t when leaving country i out.

In practice, we add two wrinkles to the method described above. First, we employ the leave-one-out strategy to remove any direct effect of a country's exports and imports on the fixed effects of their trading partners. That is, we estimate equation (4.2) N times for each sector and time period, each time leaving out the trade flows from a particular country i. We then construct each country i's foreign shocks using the estimates from the regression that omitted its data. Second, as is well known, $\kappa_{ik,t}^{x}$ and $\kappa_{nk,t}^{m}$ are identified only up to a sector-time-specific multiplicative constant and require normalization. Rather than the usual practice of designating a particular numéraire country, we restrict the sum of the logged importer effects to be zero. This normalization ensures that the relative growth rates of the foreign shocks across industries are not driven by fluctuations in the trade flows of the numéraire country, minimizing measurement error.

This procedure uses only foreign data to construct external market access and projects bilateral flows onto a small number of variables (distance and contiguity). By construction, it excludes domestic factors that act as country-specific average export taxes that apply to all destinations. It also excludes idiosyncratic bilateral factors that affect trade flows. This tends to minimize concerns about domestic policies or shocks influencing measured market access.

5. Empirical Results

5.1 Summary of Empirical Procedure

Because the estimation strategy involves several distinct components, before reporting the main estimation results we provide a compact summary of the estimation steps:

1. Leave-one-out gravity equation estimation with PPML to recover the foreign component of $FMA_{ik,t}$ and $CMA_{ik,t}$ by country and decade for 268 sectors.

2. K-means clustering algorithm to group manufacturing sectors into 4 clusters. Agriculture and mining are separate clusters.

3. Construct cluster-level $d\ln FMA_{ig,t}$ and $d\ln CMA_{ig,t}$.

4. LASSO of $d\ln y_{i,t}$ and $d\ln FMA_{ig,t}$ on $d\ln \mathbf{a}_{it}$ and $\mathbf{x}_{i,t}$ to select the set of controls.

5. OLS regression of $d\ln y_{i,t}$ on $d\ln FMA_{ig,t}$ and selected controls to obtain estimates of δ_g^x.

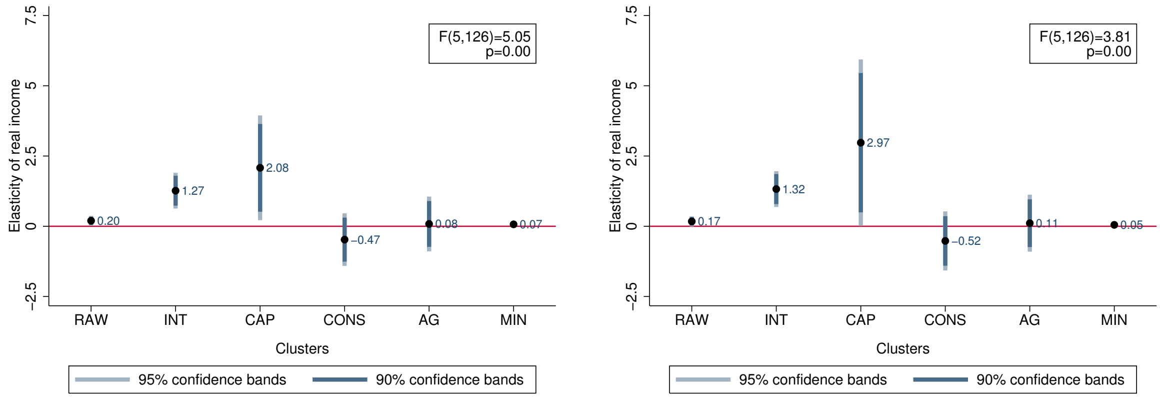
5.2 Baseline Estimates

Figure 1 presents the estimation results graphically by displaying the coefficients on the foreign demand shocks for each cluster. All specifications include (i) time effects; and (ii) the natural log of initial GDP per capita, to control for conditional convergence (Barro and Sala-i-Martin, 1992). Clusters 1-4 are manufacturing clusters obtained by the k-means algorithm, cluster 5 is agriculture, and cluster 6 mining and quarrying. The darker/lighter bars depict 90%/95% confidence intervals obtained with standard errors clustered at the country level.

The coefficients in the left panel come from OLS estimation. The right panel displays the Post-Double-Selection estimation results (Belloni et al., 2014b). The Post-Double-Selection model augments the OLS specification with the controls that were selected by the procedure described in detail in Appendix B.4. The first apparent feature of the results is the considerable heterogeneity in the coefficients. Indeed, the F-tests reject the equality of these coefficients at the 1% level of significance. Foreign demand shocks in the complex intermediates (INT), and the capital goods (CAP) clusters have positive estimated real income effects that are notably larger than the other clusters, although the confidence interval on CAP is fairly wide. In contrast, all other clusters have estimated elasticities that are close to zero (although mostly positive) and that are relatively precisely estimated.

The LASSO model includes a full set of potential controls, namely the full vector of $d\ln CMA_g$'s, the industry-level initial equilibrium variables (initial import and export shares, and weighted initial firm and consumer market access levels), initial population, initial capital, and initial per capita income squared. In total, 1106 potential control variables are included and 6 of them are selected in the double-selection procedure via LASSO. Appendix Table A6 lists the potential and the selected

Figure 1: Cluster-Specific Coefficients and Confidence Intervals



(a) OLS Estimates

(b) LASSO Estimates

Notes: This figure reports estimates of the δ_g^x coefficients in equation (3.6). All specifications control for (i) time effects and (ii) log initial GDP per capita. The left panel displays the baseline OLS estimates. The right panel displays the Post Double-LASSO estimates. 6 control variables are selected in the double-selection step. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display the results of an F-test for equality of the coefficients in each plot.

controls in the Post-Double-Selection estimation.[17] Substantively the results are quite similar to the OLS specification, although some confidence intervals widen.

5.3 Robustness

Small country assumption. We assess the sensitivity of the results to possible violations of the small country assumption. Country i can be a large trading partner of country n, such that the fixed effects estimated for country n are affected by the shocks to country i itself. Note that this concern is mitigated by the fact that the fixed effects are extracted from the gravity equations using the leave-one-out approach, whereby country i is dropped from the gravity sample when estimating the fixed effects that go into building country i's $d \ln FMA$'s. As argued in Section 3.1.1, domestic supply shocks in a large country tend induce a negative correlation between the regressors of interest and the error term and bias the coefficients downwards. It therefore seems unlikely that the general equilibrium effects of large country shocks can explain the large coefficients in the complex intermediate and capital goods clusters. Nonetheless, we check the robustness of the results to the inclusion of countries with substantial international market shares in two ways.

First, we drop the countries for whom i is a large trading partner from the computation of the foreign demand shocks. Specifically, when constructing the country i's $d \ln FMA$ in sector k, we drop

[17] We follow Belloni et al. (2014a) and choose the tuning parameter by K-fold cross validation: see Appendix B.4.3. The statistics literature often chooses the tuning parameter to be one standard deviation above the minimizing value in order to select a more parsimonious model. Our baseline specification uses the minimizing value, which results in *more* controls being selected. The results using a smaller tuning parameter are available upon request.

importer n from the summation in equation (4.3) if more than 25% of its imports in sector k are from country i, i.e. $\lambda_{ink,t} > 0.25$. The results are reported in panel (a) of Appendix Figure A6. The results are broadly similar to the baseline specification. Second, we drop countries that account for large world export shares in individual clusters and decades from the estimation sample. Specifically, we isolate the top 100 world export shares at the country-decade-cluster level. These 100 observations represent 2.3% of the 4304 available country-cluster-decade observations. The smallest of these top 100 world export shares is 7.7%. We then drop the country-decade instances in which the country was in the top 100 world export shares in any cluster. Panel (b) of Appendix Figure A6 depicts the results, which are again quite similar to the baseline.

Spatial correlation in shocks. Our identification strategy relies on the assumption that country i's unobserved shocks are uncorrelated with the foreign demand shocks. This assumption could be violated if productivity shocks are spatially correlated, so that nearby countries are subject to similar shocks (see Section 3.1.1 for a discussion). To address this concern, Appendix Figure A7 reports 3 robustness checks. First, in panel (a) we omit neighboring countries from the calculation of the $d \ln FMA$ terms and re-estimate the model. We define the set of neighboring countries as the union of the contiguous countries and the 5 closest countries by geographic distance. Second, in panel (b) we control directly for the average TFP growth of the neighboring countries in the post-LASSO OLS. TFP is sourced from the Penn World Tables. Third, in panel (c) we control for unweighted $d \ln FMA_{ig,t}$ in the post-LASSO OLS. The unweighted $d \ln FMA_{ig,t}$ captures the growth of foreign demand in each cluster, absorbing any relevant correlation between domestic and foreign shocks. The coefficients of interest are then identified solely from the interaction between foreign shocks and initial export revenue shares. All three of these checks reveal very little change relative to the baseline.

We also examined whether the δ_k^x vary across time, or across countries. Appendix B.6 presents the results of splitting the sample by developed vs. developing countries. We found similar qualitative results for both groups. In unreported results, we estimated the coefficients of interest on the first 3, and the last 3, decadal growth rates our sample. We could not reject equality of the coefficients for the early vs. late time periods. Finally, we conducted robustness checks on the number of clusters by allowing 5 manufacturing clusters and re-estimating. Table A5 reports the summary statistics for the resulting clusters. The characteristics of the original 4 manufacturing clusters are similar to the baseline. When given an opportunity to isolate a fifth cluster, the procedure picks out skill-intensive industries. Figure A11 reports the results with 5 manufacturing clusters. The new cluster has a δ_k^x close to zero, whereas the other clusters' δ_k^x's are unaffected similar to the baseline.

6. Quantitative Implications

This section evaluates the quantitative implications of our estimates along three different dimensions. We first compute the welfare impacts of uniformly decreasing external trade barriers. Second, we assess the potential gains from export and production subsidies. Third, we explore the implications of our estimates for quantitative trade models. Since our welfare elasticities are estimated for the 6 tradeable sector clusters, in this section all expenditure and sales shares are computed at the cluster level, and the model is calibrated to data at the cluster level. Thus, with some abuse of notation in this section k indexes clusters, to keep the subscripting consistent with Section 2.

6.1 Specialization Patterns and the Effects of Globalization

Reductions in trade costs lead to reallocation towards sectors with comparative advantage. Our estimates show that these reallocations do not have the same welfare effects; trade-induced reallocations to some sectors are much more valuable than to others. In order to gauge the quantitative magnitude of these effects, we compute the elasticity of each country's real income to the same uniform log-change in FMA in every sector. This change can be interpreted from a country's perspective as a reduction in trade costs to every foreign destination.[18] A simple transformation of our estimating equation leads to the following expression for this elasticity:

$$\frac{d \ln y_{i,t}}{d \ln FMA} = \sum_k \widehat{\delta}_k^x \lambda_{ik,t}^x. \tag{6.1}$$

This counterfactual allows us to focus purely on the role of export specialization, as reflected in the $\lambda_{ik,t}^x$'s, since these are the only variables that differ across countries.

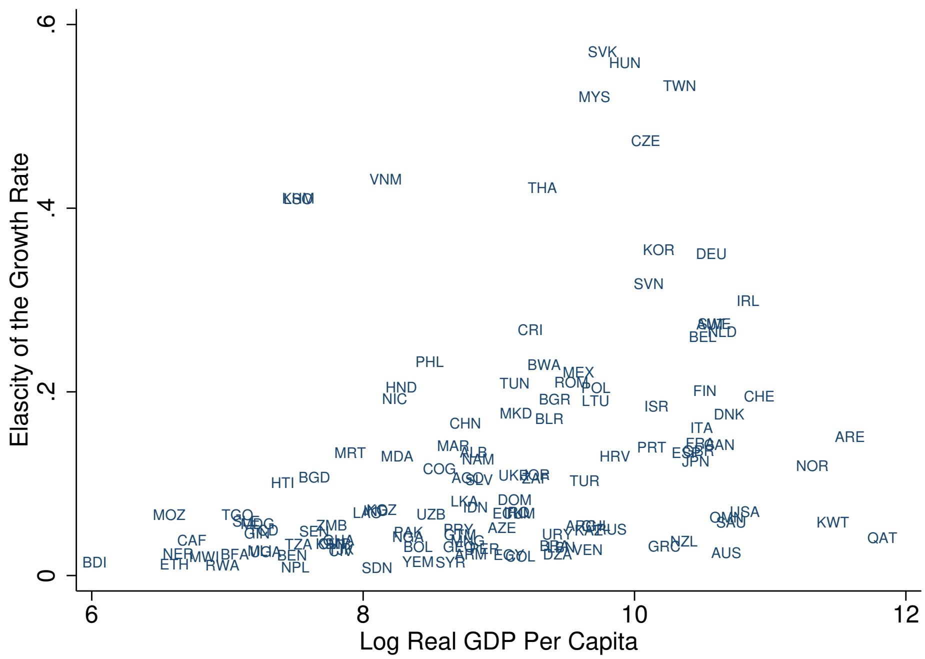
We compute the elasticities (6.1) based on the 2015 trade shares and the double-LASSO estimates from the right panel of Figure 1. Figure 2 plots them against log PPP-adjusted income per capita. There is indeed a great deal of heterogeneity in the country impact of foreign demand shocks. The real income elasticity with respect to a uniform foreign demand shocks ranges from essentially zero for countries chiefly in Sub-Saharan Africa, to around 0.5 for some Central European and East Asian countries such as Hungary, Slovakia, Malaysia, and Taiwan. The elasticities are positively correlated with real GDP per capita, but there is still substantial heterogeneity for middle and high-income countries depending on export specialization as well as openness to trade. Thus, what you export matters: countries that specialize in complex intermediate and capital goods benefit much more from declining trade costs than do agricultural and commodity producers.

6.2 Policy Implications

In Section 2, we derived the formula (2.9) for the welfare gains from small export subsidies in terms of the δ_k^x, the foreign demand elasticities σ_k, and shares from the initial state of the economy. Appendix A derives an analogous formula (A.17) for the welfare gains from production subsidies when upper-tier utility and production functions are Cobb-Douglas. We now implement these formulas on data

[18] This should not be interpreted as the global GE effect of a worldwide reduction in trade costs, since we are maintaining the small country assumption.

Figure 2: Elasticity of Real Income with Respect to a Uniform Foreign Shock



Notes: This figure scatters the elasticity of real income with respect to a uniform foreign demand shock (equation (6.1)) against real GDP per capita, calculated using the estimated δ^x and trade shares in 2015.

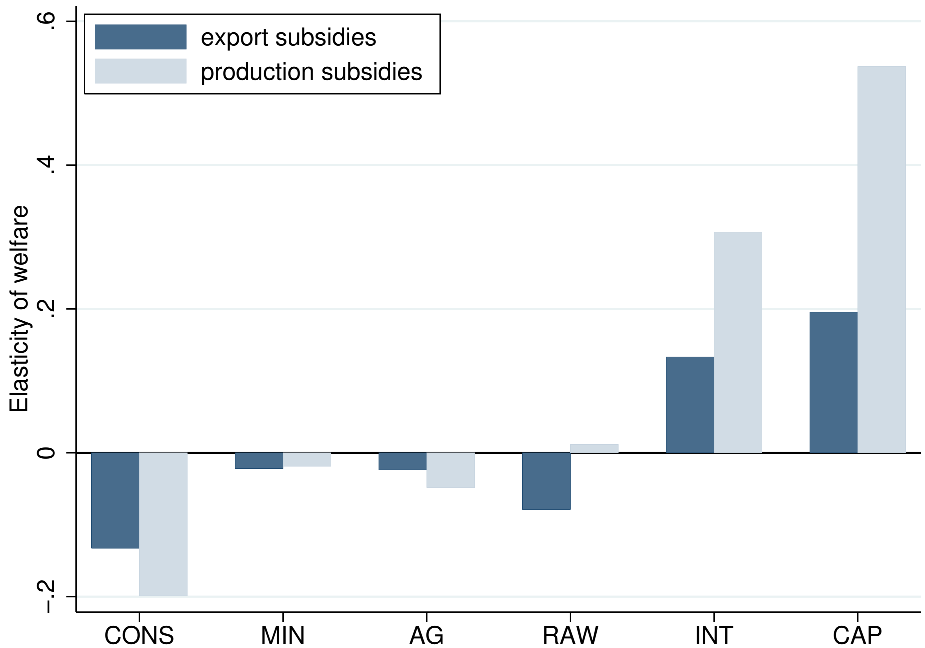
from a calibrated "representative country."

Calculating the gains from small export subsidies requires the export revenue shares λ_k^x, sectoral final expenditure shares e_k, and domestic sourcing shares θ_k^d as well as the ratio of gross output to value added, Y/V. We set the λ_k^x to equal the simple average from the countries in our sample in 2015, and set $Y/V = 2$. Since COMTRADE does not have information on domestic shares, we obtain e_k and θ_k^d from the WIOD (Timmer et al., 2015) as simple averages of the values for each country (see Appendix A.4 for details). We use $\sigma_k = 6 \; \forall k$, which implies a standard value of 5 for the trade elasticity.

The dark bars of Figure 3 plot the results for our 6 tradable sectors. Export subsidies in the complex intermediates and capital goods sectors produce welfare gains, whereas in all other clusters they lead to welfare losses (or equivalently, gains from taxing exports instead). The first-order effects are quantitatively substantial: a 10% export subsidy produces a long run welfare gain of about 1.3% for complex intermediates and nearly 2% for capital goods. The light bars in Figure 3 plot the gains from production subsidies, assuming Cobb-Douglas expenditure and production functions (see equation A.17). The qualitative conclusions are similar to export subsidies, although the magnitudes are naturally larger given the larger base to which the subsidy is applied.

The formula (2.9) shows the first-order elasticity of welfare with respect to these policies at the margin. As such, it cannot be used to find the optimal industrial policies, as those often involve

Figure 3: First Order Gains from Export and Production Subsidies



Notes: Gains from policies computed using the formulas (2.9) and (A.17).

important non-linearities. What is robust is the finding that some positive level of export subsidies is welfare-improving in the capital and complex intermediate sectors.[19] Another way to look at these results is that they tell us which sectors in the economy are inefficiently small, and which inefficiently large. Welfare increases when factors are reallocated towards sectors in which $\delta_k^x > \frac{1}{\sigma_k}\left(\frac{Y}{V} + \sum_{k'} \delta_{k'}^x e_{k'} \theta_{k'}^d\right)$ (see 2.9), and vice versa.

These results thus imply substantial gains from reallocating export activity towards complex intermediate and capital goods sectors at the margin. Such policies have been widely used, most famously in some fast-growing East Asian economies, but their efficacy has been questioned. Our evidence shows that export subsidies can indeed play a positive role in development, echoing the recent findings of Lashkaripour and Lugovskyy (2023) who use a more structural approach . Although our inference is indirect in that we do not use variation in actual export subsidies for estimation, it is justified under fairly mild theoretical restrictions that encompass the vast majority of quantitative trade models in the literature.

[19]The gains from export subsidies are increasing in σ_k. Choosing a low enough σ_k can, in principle, make the gains from export subsidies negative in these sectors. In practice, the gains remain positive for $\sigma > 2$, which would imply a trade elasticity of 1. Lower values than 1 for the long run trade elasticity seem implausible. Figure 3 reports the welfare changes under the assumption that σ does not vary across clusters, to focus most squarely on the variation induced by the δ_k^x's. Appendix Figure A12 presents the results under cluster-specific σ_k's, obtained after concording the estimates by Caliendo and Parro (2015) to our clusters. Doing so leaves the basic results unchanged, though the welfare impact is somewhat higher for complex intermediates, and somewhat lower for the capital goods.

6.3 Structural Model

This section sets up a quantitative small open economy framework of production and trade that embodies mechanisms that have been explored in the previous literature, such as an input-output matrix, endogenous capital accumulation, and sector-level scale economies. We return to the small open economy model of Section 2, specify mechanisms and functional forms, calibrate it, and compare the implied real income elasticities with respect to foreign demand shocks inside the model to our estimates. Our goal is to uncover what model features and parameter values are consistent with the empirical evidence on the general equilibrium response to foreign shocks.

To complete the description of the model, we specify the unit cost functions $c_k(\cdot)$ and the upper-tier utility function $U(\cdot)$. The representative consumer supplies a constant quantity of labor L inelastically, owns the capital stock $\mathcal{K}_t$, and chooses a sequence of consumption and investment to maximize the present discounted value of utility:

$$
\begin{array}{rcl}
\max_{\{C_t, I_t\}} & \sum_{t=0}^{\infty} & \rho^t \dfrac{C_t^{1-\psi}}{1-\psi} \\
s.t. & & \\
\mathbb{P}_t^C C_t + \mathbb{P}_t^I I_t & \leq & w_t L + r_t \mathcal{K}_t \quad \forall t \\
\mathcal{K}_{t+1} & = & I_t + (1-\chi)\mathcal{K}_t,
\end{array}
$$

where I_t is investment, w_t is the wage, r_t is the price of capital, χ is the depreciation rate, and $\mathbb{P}_t^C$ and $\mathbb{P}_t^I$ are the consumption and investment price indices, respectively. Note that the sequence of budget constraints incorporates the assumption of no international borrowing and lending.

Total consumption and investment are aggregates of goods coming from different sectors:

$$
C_t = \prod_k C_{kt}^{e_k} \qquad\qquad I_t = \prod_k I_{kt}^{\nu_k},
$$

where C_{kt} and I_{kt} are quantities of sector k good used for consumption and investment, respectively. The sectoral compositions of consumption and investment may differ. The total quantity of sector k good available for consumption and investment is an Armington aggregate of domestic and foreign varieties (equation 2.2). As described in Section 2, the gravity relationship holds within each sector.

Production in sector k uses labor, capital, and intermediates from other sectors. The unit cost function in sector k is

$$
c_{kt} = T_k L_{kt}^{-\gamma_k} \left(w_t^{\beta} r_t^{1-\beta} \right)^{\mu} \prod_l P_{lt}^{\widetilde{\alpha}_{l,k}},
$$

where P_{lt} is the ideal price index of sector l goods associated with aggregation (2.2), L_{kt} is the amount of labor employed in sector k, and $\mu + \sum_l \widetilde{\alpha}_{l,k} = 1, \forall k$. The two most important features of this cost function are that sectors use output from other sectors as intermediate inputs, and the existence of scale effects: the unit cost is decreasing in total sectoral employment. The strength of the scale effect is governed by the parameter γ_k.

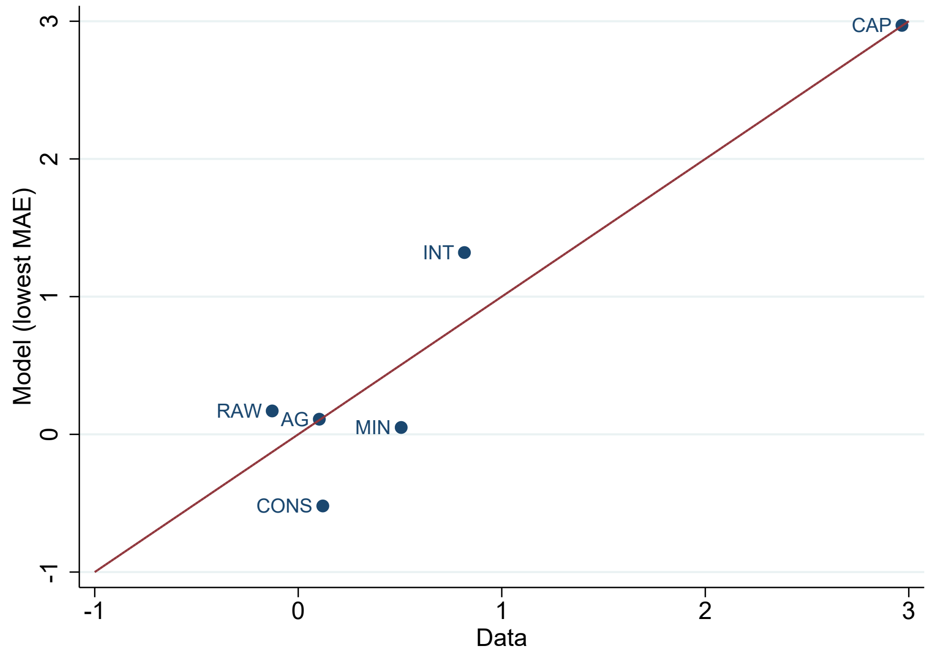
We analyze the steady state of this economy in which all the prices and quantities are constant over time. The steady state has a representation as a solution to a static model in which intermediate input shares reflect the fact that capital is also a produced input, with the steady state demand for capital governed by the rate of depreciation. To the first order, this model admits an analytical solution for the changes in output and real GDP following a shock to FMA. We introduce a non-tradeable service sector, and calibrate its size and role in production to the data. We use the WIOD to obtain the factor, production, consumption, investment and trade shares. Appendix A.4 details the model solution and calibration.

Our objective is to assess whether a simple model economy characterized by the typical distribution of sector sizes, trade shares, and the typical shape of the input-output matrix can produce the income elasticities to foreign shocks estimated in the data. We treat the elasticities of substitution and of scale as free parameters, and select them to best match the vector of δ^x_k's across clusters estimated in the data. Since there are 6 δ^x_k coefficients and potentially 12 different σ_k's and γ_k's, there are potentially infinitely many parameter combinations that will deliver a perfect fit to δ^x. To make the exercise non-trivial, we suppress heterogeneity in elasticities across sectors so that there is a single σ and a single γ that apply to all sectors of the economy (including nontradeables). We then select a pair (σ, γ) to minimize the Mean Absolute Error (MAE) between the vector of cluster-level δ^x from the data and the same objects in the model. The δ^x_k will generically differ across sectors in this environment even if σ and γ do not (and indeed, even if $\gamma = 0$) due to cross-sector differences in trade, intermediate input, expenditure and final use shares, as illustrated by the analytical solution in Appendix A.4.

Figure 4 displays the result. It plots the δ^x_k implied by the model against those estimated from the data, along with the 45-degree line. The model is quite successful at replicating the estimated coefficients. The correlation between the δ^x_k implied by the model and those estimated from the data is 0.94, and the average value across clusters produced by the model, 0.73, is also quite close to the data average of 0.68. Importantly, the model generates the variation in δ^x_k observed in the data purely through internal propagation mechanisms, without appealing to heterogeneity in the free parameters (σ and γ) across clusters. It is also reassuring to see that an important subset of the sectoral characteristics upon which our clustering scheme is based, such as position in the input-output network and final use (consumption vs capital goods), seem to generate large differences in the elasticities within the model. We explore this point further below.

In order to achieve this performance, the MAE-minimization procedure selects an elasticity of substitution $\sigma = 3.2$ and a scale elasticity of $\gamma = 0.29$. The substitution elasticity is reasonable in light of existing estimates (e.g. Broda and Weinstein, 2006). There are fewer estimates of γ in the literature. Remarkably, our best-fit scale elasticity coincides almost exactly with the 0.30 mean sectoral scale

Figure 4: Income Elasticities of Foreign Demand Shocks: Model vs. Data



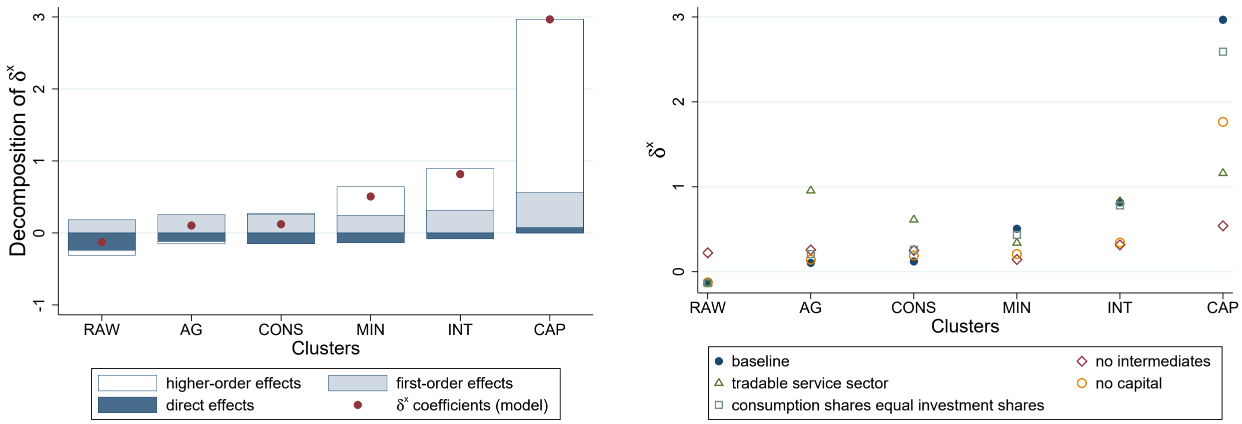
Notes: This figure plots the δ^x coefficients as estimated in the data and those generated by the model when selecting σ and γ to minimize the MAE between the data and model δ^x. The line through the data is the 45-degree line.

elasticity estimated by Lashkaripour and Lugovskyy (2023). Bartelme et al. (2019) find a somewhat lower average value of about 0.13. Appendix Figure A1 plots the MAE against γ.[20] While strictly speaking the minimum MAE criterion selects $\gamma = 0.29$, the MAE is actually quite flat from about $\gamma = 0.17$ (with associated $\sigma \approx 4.9$) to $\gamma = 0.30$. This suggests that the variation in the δ^x coefficients can actually be accounted for fairly well by a wide range of reasonable parameter values. The dashed line displays the average δ^x across clusters (right axis) against γ, with the horizontal line for data average. While the variation is about equally well-explained by a variety of γ's, one needs relatively higher values of γ to get the average δ^x right. Interestingly, the model matches the δ^x for the capital goods cluster – by far the highest δ^x in the data – almost exactly for all γ between 0.13 and 0.3. Thus, the sensitivity of the average δ^x to γ in the model is driven by other clusters. Appendix Table A1 presents some additional diagnostics on the model performance.

Input-output linkages and factor intensities. To better understand the mechanisms driving the results, we first separate the overall impact of foreign shocks into direct, first-order, and higher-order effects. Here, "first-" and "higher-order" are used in the input-output sense of intermediates being used directly vs. indirectly (i.e. the Neumann series of the input-output matrix), not to be confused

[20]Note that this is the lowest MAE across all possible values of σ conditional on the value of γ on the x-axis. Figure A2 shows that as γ increases, the σ that minimizes the MAE tends to decrease.

Figure 5: Model Performance: Mechanisms



(a) Direct, First-, and Higher-Order Effects

(b) Alternative Models

Notes: The left panel displays the decomposition of the overall model δ^x into direct, first-order, and higher-order effects. The right panel displays model δ^x under alternative production structures.

with the first-order Taylor approximation to the solution that is used throughout. Intuitively, the direct effect only applies to the sector experiencing the foreign demand shock and reflects how an increase in foreign sales translates into higher aggregate sales in partial equilibrium. The first-order effect reflects that the sector experiencing a foreign demand shock changes its purchases of intermediates, and the change in its value added affects final demand inside the home economy. It also captures the changes in unit costs, both through wages and scale effects (see equation A.27). Finally, the higher-order effects propagate these shocks further, as sectors affected by the initially shocked sector in turn change their demand for other sectors' output as well as the relative costs.

The left panel of Figure 5 decomposes the model-implied coefficients into the three effects. Both the levels and the variation across clusters are driven by higher-order effects. For the two sectors with the highest GDP impact – capital and complex intermediates – the higher-order effects account for the large majority of the total. It is also telling that higher-order effects are important in magnitude in only half of the clusters, even though both σ and γ are the same across clusters. This suggests that the entire matrix of sectoral interconnections matters quantitatively for the heterogeneity in the income elasticities to foreign shocks.[21]

To highlight which determinants of higher-order propagation are key, we examine a set of alternative economies that feature different internal propagation mechanisms. In the first alternative, we suppress intermediate good usage by setting $\widetilde{\alpha}_{k,l} = 0 \ \forall k, l$, $\mu = 1$. In the second, we abstract from capital – setting $\beta = 1$ – and thus from the responses of capital accumulation to shocks. The third

[21]It is sensible that the first- and higher-order effects are often much larger than the direct effects, because they include general equilibrium adjustments to unit costs, driven in part by the scale effects on productivity. As clarified by the Appendix A.4 equations, direct effects are not exactly the same across clusters because they differ in average size. Figure 5 shows that those differences in the direct effect are fairly minor.

alternative assumes that the composition of investment is the same as that of consumption: $e_k = \nu_k$. Finally, the fourth alternative assumes there is no non-tradeable sector, and assigns to services the level of trade openness in both imports and exports (θ^f and π^x) equal to the average of the traded sectors. This alternative economy is interesting because most of the GDP impact of the shocks to the capital and complex intermediates sectors on GDP is accounted for by the resulting expansion of the service sector. That is, the proximate reason for the high GDP impact of foreign demand shocks in these tradeable sectors is that service sector output goes up, mostly through higher-order effects. The service sector is special in the baseline model because of its non-tradeability (as well as its large size), which implies that an expansion in the service sector output does not lead – at least directly – to negative terms-of-trade effects. As a result, changes in service sector output have the largest impact on real GDP. Importantly, all 4 alternative models keep both exports and imports as a fraction of sectoral gross output the same as in the baseline. Thus, all 4 models feature the same level of external "openness" in the tradeable sector (and the first 3 models, economywide).[22] Only internal propagation mechanisms inside the economy differ between the alternative models and the baseline.

The right panel of Figure 5 displays the model-implied δ^x coefficients in the baseline and the alternative models.[23] Removing the input-output linkages has the largest impact on the model-implied δ^x. The average falls by some 60% relative to the baseline, and variation across clusters all but disappears. The capital cluster still has the highest coefficient, but at 0.5 it is one-sixth of the value in the data and baseline model. A model with no capital is somewhat more successful at matching the data than the model with no intermediates. It generates larger average δ^x and a coefficient of 1.8 in the capital sector, much closer to the data. Nonetheless, its average δ^x still falls about 30% short of both the baseline model and data. By contrast, the differences in the composition between investment and consumption goods do not matter as much quantitatively. What is important is the existence of capital as an input, rather than the relative composition of capital investment. The existence of a non-tradeable service sector ends up mattering quite a bit as well. If we make the service sector as tradeable as the other sectors, the average δ^x falls by more than 50%, and the model does not generate coefficients that closely match the observed variation across sectors.

Scale economies. We next evaluate how well this model can match estimated elasticities without appealing to scale economies. Table 2 considers a range of models with constant returns to scale ($\gamma = 0$). We first report the model-implied δ^x under a range of σ from 1 to 10. The average δ^x is decreasing in the Armington elasticity. It is not difficult to get average δ_k^x to be the same as estimated in the data, by simply lowering σ. However, lowering σ leads to δ^x that are much too high in 4 out of 6 clusters, where the data $\widehat{\delta}^x$ are near zero or negative. So just varying σ can get the average level right

[22] There are multiple notions of "keeping trade openness constant" when going from data with intermediate inputs to a model with no intermediates, because one needs to decide whether to keep trade flows constant as a share of gross expenditure or of value added. In these experiments, when we change the input-output structure we keep trade constant as a share of gross expenditure. This is the cleanest procedure in our context, as it involves changing only one scalar parameter (μ).

[23] Appendix Table A1 report additional details on these exercises.

at the expense of degrading the overall fit, with MAEs that are 2-3 times higher than our baseline MAE of 0.32. In addition, the σ needed to match the average estimated $\widehat{\delta^x}$, 2.2, is quite low relative to conventional wisdom.

Perhaps we can do better by appealing to variation in σ across sectors. We select 6 sector-specific σ_k to minimize the MAE with respect to the data. Selecting cluster-specific σ_k's to minimize the MAE with respect to the data yields σ_k's in the range of 4−7, but implies average δ^x about one-third of the data value, low dispersion across sectors, and the complex intermediate and capital goods coefficients that are much too small. Despite featuring 3 times as many parameters as our baseline calibration with a single σ and γ, the resulting fit is much worse, with an MAE is over twice as large as the baseline. Absent scale effects, the calibrated model lacks sufficient internal propagation mechanisms to generate the observed amount of dispersion in δ^x across sectors for any values of σ. More generally, cross-sector variation in international market power cannot generate the observed variation in δ^x in the absence of domestic distortions.

Thus, the constant-returns to scale version of this particular model cannot match the sectoral differences in the δ^x that we estimate in the data. The model with a positive but common scale elasticity generates sectoral differences in the δ^x_k because the endogenous productivity increase that results from sectoral expansion is amplified for strongly connected sectors, which (all else equal) are too small in the *laissez-faire* equilibrium. These effects are further amplified when the sector is relatively upstream from the non-traded sector, which is also too small due to the positive terms-of-trade effects associated with its expansion.[24]

Our main conclusions from the quantification exercise are as follows. First, a relatively standard and parsimonious model calibrated to a representative sectoral production and trade structure can successfully reproduce the estimated real income responses to foreign demand shocks. Importantly, the quantitative model achieves this via internal propagation within the home economy, without appealing to sectoral heterogeneity in substitution and scale elasticities. Furthermore, the model succeeds under reasonable substitution and scale elasticities, and in fact it performs well under a range of those rather than strongly preferring a narrow set of values.

Second, two features of the model are crucial for quantitative success: scale effects and input-output linkages. Substantial scale effects appear important for the current crop of quantitative trade models to match the variation in the long run general equilibrium response of economies to foreign demand shocks across sectors. Similarly, the entire structure of sectoral linkages inside the economy is important for the success of this particular model. Most of the overall effect of foreign shocks is due to higher-order propagation, rather than direct or first-order effects. Intermediate input linkages, capital accumulation, and service sector non-tradeability all matter individually, in the sense that the model becomes less successful (under the same structural elasticities) at replicating the data when one of these features is suppressed.

[24]When the non-traded sector expands the traded sectors contract, yielding improvements in the terms of trade.

Table 2: δ^x Coefficients: Data and Model, No Scale Effects

	$\widehat{\delta^x}$ data	δ^x model ($\gamma = 0$)			
		$\sigma = 1.01$	$\sigma = 3$	$\sigma = 10$	σ_k min MAE
RAW	0.17	1.27	0.36	0.10	0.17
INT	1.32	1.91	0.54	0.15	0.26
CAP	2.97	2.86	0.80	0.23	0.38
CONS	-0.52	1.64	0.46	0.13	0.22
AG	0.11	1.38	0.39	0.11	0.19
MIN	0.05	1.39	0.39	0.11	0.19
Average δ^x	0.68	1.74	0.49	0.14	0.23
MAE model vs. data		1.10	0.79	0.78	0.77

Notes: This table reports the estimates of $\widehat{\delta^x}$ (first column), and the model δ^x under the alternative values of σ, in a constant returns to scale model ($\gamma = 0$) throughout. In columns 2 through 4, σ equals 1.01, 3 and 10 respectively. Column 5 selects sector-specific σ_k's that minimize the Mean Absolute Error (MAE) between model and data. The bottom panel reports the average δ^x for each case, and for the theoretical models reports the MAE.

7. Conclusion

Using a theoretically grounded approach and employing new empirical techniques, we have shown that positive foreign demand shocks in sectors producing complex intermediate and capital goods lead to significantly larger increases in long-run welfare than shocks in other sectors. Our estimates, along with our theoretical results, imply that countries benefit from reallocating economic activity toward these sectors. Our quantification shows that trade models with scale effects, intermediate goods and endogenous capital accumulation can match the empirical estimates.

Questions surrounding the effect of the external environment on economic development, as well as the appropriate policy response to the international market, have been central in the great policy debates of the past 60 years, from import-substituting industrialization to the Washington Consensus to the "Washington Confusion" (Rodrik, 2006). Our results affirm the importance of the external environment for economic development and validate the renewed interest in the role of sectoral trade and industrial policy in development.

References

Adão, Rodrigo, Arnaud Costinot, and Dave Donaldson, "Nonparametric counterfactual predictions in neoclassical models of international trade," *American Economic Review*, 2017, *107* (3), 633–89.

_, _, and _, "Putting Quantitative Models to the Test: An Application to Trump's Trade War," May 2023. Mimeo, Chicago Booth and MIT.

_, Costas Arkolakis, and Federico Esposito, "General equilibrium effects in space: Theory and measurement," 2020. mimeo, Chicago Booth, Yale, and Tufts University.

_, Paul Carrillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz, "Exports, Imports, and Earnings Inequality: Micro-Data and Macro-Lessons from Ecuador," July 2020. mimeo, Chicago Booth, GWU, MIT, University of Zurich.

Allen, Treb, Costas Arkolakis, and Yuta Takahashi, "Universal gravity," *Journal of Political Economy*, 2020, *128* (2), 393–433.

Amiti, Mary and Jozef Konings, "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, December 2007, *97* (5), 1611–1638.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New trade models, same old gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Armington, Paul S, "A theory of demand for products distinguished by place of production," *IMF Staff Papers*, 1969, *16* (1), 159–178.

Arthur, David and Sergei Vassilvitskii, "k-means++: The advantages of careful seeding," in "Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms" Society for Industrial and Applied Mathematics 2007, pp. 1027–1035.

Atkin, David, "Endogenous Skill Acquisition and Export Manufacturing in Mexico," *American Economic Review*, 2016, *106* (8), 2046–2085.

Bagwell, Kyle and Robert W. Staiger, eds, *Handbook of Commercial Policy*, North-Holland, 2016.

_ and Seung Hoon Lee, "Trade policy under monopolistic competition with firm selection," *Journal of International Economics*, 2020, *127*, 103379.

Baqaee, David Rezza and Emmanuel Farhi, "Networks, Barriers, and Trade," August 2019. mimeo, UCLA and Harvard.

Bardhan, Pranab K, "On Optimum Subsidy to a Learning Industry: An Aspect of the Theory of Infant-Industry Protection," *International Economic Review*, February 1971, *12* (1), 54–70.

Barro, Robert J. and Xavier Sala-i-Martin, "Convergence," *Journal of Political Economy*, April 1992, *100* (2), 223–251.

Bartelme, Dominick, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The Textbook Case for Industrial Policy: Theory Meets Data," August 2019. NBER WP 26193.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen, "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 2012, *80* (6), 2369–2429.

—, Victor Chernozhukov, and Christian Hansen, "High-dimensional methods and inference on structural and treatment effects," *Journal of Economic Perspectives*, 2014, *28* (2), 29–50.

—, —, and —, "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 2014, *81* (2), 608–650.

—, —, Ivan Fernández-Val, and Christian Hansen, "Program evaluation and causal inference with high-dimensional data," *Econometrica*, 2017, *85* (1), 233–298.

Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig, "This Mine Is Mine! How Minerals Fuel Conflicts in Africa," *American Economic Review*, June 2017, *107* (6), 1564–1610.

Bhagwati, Jagdish and V.K. Ramaswami, "Domestic Distortions, Tariffs and the Theory of Optimum Subsidy," *Journal of Political Economy*, 1963, *71* (44-50).

Blanchard, Emily J. and William W. Olney, "Globalization and human capital investment: Export composition drives educational attainment," *Journal of International Economics*, 2017, *106*, 165 – 183.

Broda, Christian and David Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, May 2006, *121* (2), 541–85.

Burstein, Ariel and Javier Cravino, "Measured Aggregate Gains from International Trade," *American Economic Journal: Macroeconomics*, April 2015, *7* (2), 181–218.

Caliendo, Lorenzo and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 2015, *82* (1), 1–44.

—, —, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," *The Review of Economic Studies*, December 2017, *85* (4), 2042–2096.

Campolmi, Alessia, Harald Fadinger, and Chiara Forlati, "Trade policy: Home market effect versus terms-of-trade externality," *Journal of International Economics*, 2014, *93* (1), 92–107.

Caselli, Francesco and Daniel J. Wilson, "Importing Technology," *Journal of Monetary Economics*, 2004, *51* (1), 1–32.

Chipman, John S., "External Economies of Scale and Competitive Equilibrium," *Quarterly Journal of Economics*, 1970, *84* (3), 347–385.

Choi, Jaedo and Andrei A Levchenko, "The long-term effects of industrial policy," Technical Report, National Bureau of Economic Research 2021.

Chor, Davin, "Unpacking Sources of Comparative Advantage: A Quantitative Approach," *Journal of International Economics*, November 2010, *82* (2), 152–167.

Conte, Bruno, Klaus Desmet, DÃ!'vid KrisztiÃ!'n Nagy, and Esteban Rossi-Hansberg, "Local sectoral specialization in a warming world," *Journal of Economic Geography*, September 2021, *21* (4), 493–530.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," *Handbook of International Economics*, 2014, *4*, 197.

—, Dave Donaldson, and Ivana Komunjer, "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 2012, *79* (2), 581–608.

Deardorff, Alan V. and Robert M. Stern, *Computational Analysis of Global Trading Arrangements*, Ann Arbor: The University of Michigan Press, 1990.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, July 2008, *55* (3), 511–540.

Demidova, Svetlana, Takumi Naito, and Andrés Rodríguez-Clare, "The Small Open Economy in a Generalized Gravity Model," Technical Report, National Bureau of Economic Research 2022.

Dippel, Christian, Avner Greif, and Daniel Trefler, "Outside Options, Coercion, and Wages: Removing the Sugar Coating," *Economic Journal*, August 2020, *130* (630), 1678–1714.

Eaton, Jonathan and Samuel S. Kortum, "Trade in capital goods," *European Economic Review*, 2001, *45* (7), 1195 – 1235.

_ and _ , "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

_ , _ , and Sebastian Sotelo, "International Trade: Linking Micro and Macro," February 2012. NBER WP 17864.

Ethier, Wilfred J., "Decreasing Costs in International Trade and Frank Graham's Argument for Protection," *Econometrica*, 1982, *50* (5), 1243–1268.

Fajgelbaum, Pablo D, Pinelopi K Goldberg, Patrick J Kennedy, and Amit K Khandelwal, "The return to protectionism," *Quarterly Journal of Economics*, 2020, *135* (1), 1–55.

Feyrer, James, "Distance, Trade, and Income - The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment," December 2009. NBER Working Paper 15557.

_ , "Trade and Income – Exploiting Time Series in Geography," *American Economic Journal: Applied Economics*, October 2019, *11* (4), 1–35.

Fieler, Ana Cecília, Marcela Eslava, and Daniel Yi Xu, "Trade, quality upgrading, and input linkages: Theory and evidence from Colombia," *American Economic Review*, 2018, *108* (1), 109–46.

Frankel, Jeffrey A. and David H. Romer, "Does Trade Cause Growth?," *American Economic Review*, June 1999, *89* (3), 379–399.

Graham, Frank D., "Some Aspects of Protection Further Considered," *Quarterly Journal of Economics*, 1923, *37* (2), 199–227.

Haberler, Gottfried, "Some Problems in the Pure Theory of International Trade," *Economic Journal*, 1950, *60*, 223–240.

Hadass, Yael S and Jeffrey G Williamson, "Terms-of-Trade Shocks and Economic Performance, 1870-1940: Prebisch and Singer Revisited," *Economic Development and Cultural Change*, April 2003, *51* (3), 629–656.

Hagen, Everett, "An Economic Justification of Protectionism," *Quarterly Journal of Economics*, 1958, *72*, 496–514.

Harrison, Ann and Andrés Rodríguez-Clare, "Trade, Foreign Investment, and Industrial Policy for Developing Countries," in Dani Rodrik and Mark Rosenzweig, eds., *Handbook of Development Economics*, New York: North- Holland, 2010, chapter 63.

Hausmann, Ricardo, Jason Hwang, and Dani Rodrik, "What You Export Matters," *Journal of Economic Growth*, 2007, *12*, 1–25.

Head, Keith, "Infant industry protection in the steel rail industry," *Journal of International Economics*, 1994, *37* (3-4), 141–165.

_ and Thierry Mayer, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195.

Hidalgo, Cesar and Ricardo Hausmann, "The building blocks of economic complexity," *Proceedings of the national academy of sciences*, June 2009, *106* (26), 10570–10575.

_ , Bailey Klinger, Albert-Laszlo Barabasi, and Ricardo Hausmann, "The Product Space Conditions the Development of Nations," *Science*, 2007, *371*, 482–487.

Hsieh, Chang-Tai and Ralph Ossa, "A global view of productivity growth in China," *Journal of international Economics*, 2016, *102*, 209–224.

Humphreys, Macartan, Jeffrey D. Sachs, and Joseph E. Stiglitz, eds, *Escaping the Resource Curse*, Columbia University Press, 2007.

Jarreau, Joachim and Sandra Poncet, "Export sophistication and economic growth: Evidence from China," *Journal of Development Economics*, March 2012, *97* (2), 281 – 292.

Kasahara, Hiroyuki and Joel Rodrigue, "Does the use of imported intermediates increase productivity? Plant-level evidence," *Journal of Development Economics*, August 2008, *87* (1), 106–118.

Kehoe, Timothy J, Pau S Pujolas, and Jack Rossbach, "Quantitative trade models: Developments and challenges," *Annual Review of Economics*, 2017, *9*, 295–325.

Krugman, Paul, "Scale economies, product differentiation, and the pattern of trade," *The American Economic Review*, 1980, *70* (5), 950–959.

_ and Anthony Venables, "Globalization and the Inequality of Nations," *Quarterly Journal of Economics*, 1995, *110* (4), 857–880.

Kucheryavyy, Konstantin, Gary Lyn, and Andrés Rodríguez-Clare, "Grounded by gravity: A well-behaved trade model with industry-level economies of scale," March 2020. Mimeo.

Lane, N, "Manufacturing revolutions: industrial policy and industrialization in South Korea," 2021.

Lashkaripour, Ahmad and Volodymyr Lugovskyy, "Profits, Scale Economies, and the Gains from Trade and Industrial Policy," *American Economic Review*, October 2023, *113* (10), 2759–2808.

Lederman, Daniel and William F. Maloney, *Does What You Export Matter? In Search of Empirical Guidance for Industrial Policies*, Washington, DC: The World Bank, 2012.

Levchenko, Andrei A., "International Trade and Institutional Change," *Journal of Law, Economics, and Organization*, October 2013, *29* (5), 1145–1181.

Levchenko, Andrei A and Jing Zhang, "The evolution of comparative advantage: Measurement and welfare implications," *Journal of Monetary Economics*, 2016, *78*, 96–111.

MacQueen, James et al., "Some methods for classification and analysis of multivariate observations," in "Proceedings of the fifth Berkeley symposium on mathematical statistics and probability," Vol. 1(14) 1967, pp. 281–297.

Melitz, Marc J, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Nunn, Nathan, "Relationship-specificity, incomplete contracts, and the pattern of trade," *Quarterly Journal of Economics*, 2007, *122* (2), 569–600.

_ and Daniel Trefler, "The Structure of Tariffs and Long-Term Growth," *American Economic Journal: Macroeconomics*, 2010, *2* (4), 158–194.

Pascali, Luigi, "The Wind of Change: Maritime Technology, Trade and Economic Development," *American Economic Review*, September 2017, *107* (9), 2821–54.

Prebisch, Raúl, "Commercial Policy in the Underdeveloped Countries," *American Economic Review*, 1959, *49*, 251–273.

Redding, Stephen and Anthony J Venables, "Economic geography and international inequality," *Journal of International Economics*, 2004, *62* (1), 53–82.

Rodríguez, Francisco and Dani Rodrik, "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence," in "NBER Macroeconomics Annual 2000, Volume 15" NBER Chapters, National Bureau of Economic Research, Inc, November 2001, pp. 261–338.

Rodrik, Dani, "Goodbye Washington consensus, hello Washington confusion? A review of the World Bank's economic growth in the 1990s: learning from a decade of reform," *Journal of Economic Literature*, 2006, *44* (4), 973–987.

_ , Arvind Subramanian, and Francesco Trebbi, "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," *Journal of Economic Growth*, 2004, *9*, 131–165.

Rousseeuw, Peter J, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, 1987, *20*, 53–65.

Sachs, Jeffrey D. and Andrew M. Warner, "The Big Rush, Natural Resource Booms And Growth," *Journal of Development Economics*, June 1999, *59* (1), 43–76.

_ and Andrew Warner, "Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity*, 1995, *1*, 1–118.

Sala-i-Martin, Xavier and Arvind Subramanian, "Addressing the Natural Resource Curse: An Illustration from Nigeria," *Journal of African Economies*, 2013, *22* (4), 570–615.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries, "An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production," *Review of International Economics*, 2015, *23* (3), 575–605.

Tornell, Aaron and Philip R. Lane, "The Voracity Effect," *American Economic Review*, March 1999, *89* (1), 22–46.

Wacziarg, Romain and Karen Horn Welch, "Trade Liberalization and Growth: New Evidence," *World Bank Economic Review*, June 2008, *22* (2), 187–231.

Whalley, John, *Trade liberalization among major world trading areas*, MIT Press, 1985.

Williamson, Jeffrey G., "Globalization and the Great Divergence: terms of trade booms, volatility and the poor periphery, 1782-1913," *European Review of Economic History*, December 2008, *12* (3), 355–391.

Witten, Daniela M and Robert Tibshirani, "A framework for feature selection in clustering," *Journal of the American Statistical Association*, 2010, *105* (490), 713–726.

Yi, Kei-Mu, "Can vertical specialization explain the growth of world trade?," *Journal of Political Economy*, 2003, *111* (1), 52–102.

Young, Alwyn, "Learning by Doing and the Dynamic Effects of International Trade," *Quarterly Journal of Economics*, 1991, *106* (2), 369–405.

A. Theoretical Appendix

A.1 Competitive equilibrium

The competitive equilibrium of the economy can be represented as the set of solutions to the following system of simultaneous equations:

$$w_j L_{j,k} = \mu_{j,k} \cdot Y_k, \forall j \in J,\ k \in K \tag{A.1}$$

$$\sum_{k \in K} L_{j,k} = \bar{L}_j,\ \ \forall j \in J \tag{A.2}$$

$$V = \sum_k \sum_j w_j \cdot L_{j,k} \tag{A.3}$$

$$P_k^{1-\sigma_k} = z_k c_k^{1-\sigma_k} + CMA_k,\ \ \forall k \in K \tag{A.4}$$

$$Y_k = c_k^{1-\sigma_k} \left(z_k \frac{e_k \cdot V + \sum_{k' \in K} \tilde{\alpha}_{k,k'} Y_{k'}}{P_k^{1-\sigma_k}} + FMA_k \right)\ \ \forall k \in K. \tag{A.5}$$

Here e_k is the fraction of consumer expenditure devoted to industry k, $\mu_{j,k}$ is the fraction of industry k's gross output devoted to purchasing factor input j, and $\tilde{\alpha}_{k,k'}$ is the fraction of industry k''s gross revenue ($Y_{k'}$) used to purchase intermediate inputs from sector k. By Shephard's lemma, these shares equal the elasticities of the expenditure or cost functions with respect to price. These elasticities in principle depend on relative prices, of goods and/or factors; however, homotheticity and (perceived) constant returns imply that they do not depend on total expenditure (V) or industry gross output.

The first set of conditions (A.1) are the industry factor demand equations. The second set of conditions (A.2) equate factor demand with factor supply. The third condition equates total factor income and total expenditure, which also ensures (along with the other conditions) that trade balance holds. The fourth set of conditions (A.4) defines the price index, while the fifth set of equations (A.5) defines gross industry revenues as equal to total industry sales.

Notice that the last set of equations can be solved for Y_k as a function of the factor prices and factor allocations (as well as the exogenous market access terms) using matrix algebra. We can then plug this solution into the other equations, and also plug in the definitions of total expenditure and the price indices. We are then left with a set of equations in factor prices and factor allocations. If there is a unique solution for factor allocations given factor prices, i.e. a unique solution $\mathbf{L}$ for the factor demand equations (A.1) given a set of factor prices $\mathbf{w}$, then clearly we can reduce this system to a system of J equations setting factor demand equal to factor supply.

In a closed economy, these J equilibrium conditions equating factor supply and demand are homogeneous of degree 1, and hence a normalization is required. In the open economy these equations are not homogeneous of degree 1 in factor prices due to the presence of fixed foreign prices, and no normalization is required.

A.2 First order welfare effects of policies

Setup. We now add a government to the economy. The government imposes 2 types of taxes/subsidies, plus lump sum transfers to balance its budget:

1. Sector-level export taxes t_k^x on the value of exports, measured at the destination. That is, if the foreign consumer pays a tax-inclusive price of p_k^f for each unit shipped, then the firm's revenue from shipping that good is $\tilde{p}_k = (1 - t_k^x)p_k^f$.
2. Sector-level taxes $-s_k$ on production costs (i.e. positive s_k denote a subsidy). That is, if a firm pays its factors of production c_k to produce a single unit of the good, the government collects revenue equal to $-s_k c_k$ from the firm. Therefore, the effective unit cost of production faced by the firm is $c_k(1 - s_k)$.

The implications of each tax for quantities demanded, firm and government revenue, and worker/consumer income are as follows:

1. Foreign demand is isoelastic, and the shipped quantity demanded satisfies $q_k^f \propto (p_k^f)^{-\sigma_k}$, so we have $q_k^f \propto (\tilde{p}_k/(1 - t_k^x))^{-\sigma_k}$. Total firm revenue from foreign sales is therefore

$$X_k^f = \tilde{p}_k q_k^f = \tilde{p}_k^{1-\sigma_k}(1 - t_k)^{\sigma_k} \cdot FMA_k,$$

   whereas the government revenue from export taxes in sector k is $\frac{t_k^x}{1-t_k^x} X_k^f$. The total revenue from foreign sales is $\frac{X_k^f}{1-t_k^x}$.

2. The subsidy revenue is equal to $\sum_k -s_k c_k(q_k^d + q_k^f)$, where q_k^d is the domestic quantity shipped. Domestic demand is also isoelastic, and perfect competition implies that $\tilde{p}_k = c_k(1 - s_k)$, so government revenue can be expressed as $\sum_k -\frac{s_k}{1-s_k}(X_k^d + X_k^f)$, where X_k^d is total firm revenue from domestic sales.

The total firm revenue Y_k, from both sales and subsidy payments, can be written as

$$Y_k = \frac{1}{1 - s_k}(X_k^d + X_k^f).$$

Putting it all together, we can define a competitive equilibrium with taxes as a solution to the following set of equations:

$$w_j L_{j,k} = \mu_{j,k} \cdot Y_k, \forall j \in J,\ k \in K \tag{A.6}$$

$$\sum_{k \in K} L_{j,k} = \bar{L}_j,\ \ \forall j \in J \tag{A.7}$$

$$V = \sum_k \sum_j w_j \cdot L_{j,k} \tag{A.8}$$

$$P_k^{1-\sigma_k} = z_k \left(c_k(1-s_k)\right)^{1-\sigma_k} + CMA_k,\ \ \forall k \in K \tag{A.9}$$

$$Y_k(1-s_k) = \left(c_k(1-s_k)\right)^{1-\sigma_k} \left(z_k \left(\frac{e_k \cdot (V+T) + \sum_{k' \in K} \tilde{\alpha}_{k,k'} Y_{k'}}{P_k^{1-\sigma_k}} \right) + (1-t_k)^{\sigma_k} FMA_k \right)\ \ \forall k \in K, \tag{A.10}$$

$$T = \sum_k \frac{t_k^x}{1-t_k^x} X_k^f - \sum_k s_k Y_k. \tag{A.11}$$

Welfare is

$$U(E, \mathbb{P}) = \frac{V+T}{\mathbb{P}}, \tag{A.12}$$

where $\mathbb{P}$ is the tax-inclusive price index and V is the transfer-exclusive expenditures.

The elasticity of welfare with respect to export taxes. Fix an initial equilibrium with zero taxes. First consider a set of small changes to the $\{FMA_k\}_{k \in K}$, holding all other exogenous variables (including taxes) constant. Starting from the same zero tax equilibrium, now consider a small increase in the export tax t_l^x, with negative taxes indicating subsidies and with lump sum taxation to balance to government budget. The solution to the system (A.6)-(A.11) will be identical in the two experiments if the following conditions are satisfied:

$$\lambda_l^x d\ln FMA_l = \lambda_l^x \left(\sigma_l - e_l \cdot \theta_l^d\right) d\ln\left(1-t_l^x\right), \tag{A.13}$$

$$\lambda_k^x d\ln FMA_k = -\lambda_l^x \cdot e_k \cdot \theta_k^d \cdot d\ln\left(1-t_l^x\right),\ \forall k \neq l, \tag{A.14}$$

where θ_k^d is the share of Home's expenditure in k that is sourced domestically. The proof is straightforward: notice that FMA_k and t_k^x only enter the system directly through equation (A.10), make a change of variables from t_k^x to $(1-t_k^x)$, apply Taylor's theorem to the system in natural logs and equate the coefficients in each equation.

A set of foreign demand shocks that obey these equations will have the same effect on all endogenous variables as the export tax/subsidy in sector l. Thus the welfare effect will be the same *up to the subsidy revenue*, that is

$$\frac{d\ln U}{d\ln\left(1-t_l^x\right)} = \lambda_l^x \left(\delta_l^x \sigma_l - \frac{Y}{V} - \sum_k \delta_k^x e_k \theta_k^d \right). \tag{A.15}$$

A small export subsidy to sector l is welfare improving if and only if

$$\delta_l^x > \frac{1}{\sigma_l}\left(\frac{Y}{V} + \sum_k \delta_k^x e_k \theta_k^d\right). \tag{A.16}$$

The elasticity of welfare with respect to production subsidies . We assume that production functions and the upper tier of the utility function are Cobb-Douglas, so that $\mu_{j,k}$ and e_k are parameters rather than equilibrium objects. We have

$$c_k = C_k \cdot \prod_l P_{k'}^{\tilde{\alpha}_{k',k}}$$

where C_k is an endogenous equilibrium object that does not contain c_k or $(1-s_k)$ directly. Therefore

$$P_k^{1-\sigma_k} = z_k\left((1-s_k)c_k\right)^{1-\sigma_k} + CMA_k.$$

Taking a log-linear approximation to these two equations in response to a set of small changes in $(1-s_{k'})$ and solving, we get

$$d\ln c_k = g_k + \sum_{k'} \hat{a}_{k,k'} d\ln(1-s_{k'}),$$

where g_k is an endogenous object not directly dependent on c_k or $(1-s_k)$, and

$$\hat{a}_{k,k'} = \left[(I - A^T D\{\theta^d\})^{-1} A^T D\{\theta^d\}\right]_{k,k'},$$

$$A = \left[\tilde{\alpha}_{k',k}\right],$$

and $D\{\theta^d\} = Diag\{\theta_k^d\}$. We have

$$d\ln\theta_k^d = (1-\sigma_k)(1-\theta_k^d)(d\ln(1-s_k) + d\ln c_k), \;\; \forall k,$$

$$Y_k(1-s_k) = \theta_k^d(e_k V + \sum_{k'\in K} \tilde{\alpha}_{k,k'} Y_{k'}) - \theta_k^d e_k \sum_{k'} s_{k'} Y_{k'} + X_k^f, \;\; \forall k.$$

As with the case of export taxes, we take logs and differentiate with respect to $1-s_j$ and evaluate at $s_k = 0$ for each k, then equate coefficients. The result is that the set of shocks to FMA_k that deliver the same positive implications as a small change in $(1-slj)$ satisfy

$$\lambda_l^x d\ln FMA_l = \left(-\frac{Y_l}{Y} + \left(\frac{X_l^d}{Y}(1-\theta_l^d) + \lambda_l^x\right)(1-\sigma_l)(1+\hat{a}_{ll}) + \theta_l^d e_l \frac{Y_l}{Y}\right) d\ln(1-s_l),$$

and

$$\lambda_k^x d\ln FMA_k = \left(\left(\frac{X_k^d}{Y}(1-\theta_k^d) + \lambda_k^x\right)(1-\sigma_k)\hat{a}_{k,l} + \theta_k^d e_k \frac{Y_l}{Y}\right) d\ln(1-s_l), \;\; k \neq l.$$

As in the case without intermediate goods, the welfare effects will also be equalized, up to the subsidy revenue and (new term) the direct effect of the subsidy on the domestic consumer price index. Therefore, so

$$-\frac{d\ln U}{d\ln(1-s_l)} = \delta_l^x\left(\frac{Y_l}{Y} + (\sigma_l - 1)\left(\frac{X_l^d}{Y}(1-\theta_l^d) + \lambda_l^x\right)\right) - \frac{Y_l}{V} - \sum_k \delta_k^x\theta_k^d e_k\frac{Y_l}{Y} \tag{A.17}$$

$$+ \sum_k \delta_k^x(\sigma_k - 1)\hat{a}_{k,l}\left(\frac{X_k^d}{Y}(1-\theta_k^d) + \lambda_k^x\right) + e_l\theta_l^d.$$

The interpretation of the first three terms is the same as in equation (2.9) in the main text, adjusting for the different bases to which the subsidies are applied in the two cases. The fourth term captures the effect of the subsidy to sector l on downstream sectors through domestic production networks. The final term is the direct effect of the subsidy on the consumer price index in sector l.

A.3 The determinants of δ_k^x in simple examples

This section provides some simple examples of models in which the closed form solutions for the δ_k^x are easy to understand.

Planner's problem

Our starting point is an efficient economy in which a social planner directly chooses quantities and factor allocations to maximize domestic welfare, taking the production technology, factor supplies and the trade balance constraint as given. Denote by $q_k^{c,d}$ the quantity of final Home consumption of domestic goods, and by $q_{n,k}^{c,f}$ the quantity of final consumption of foreign goods from country n, and use an m superscript to indicate the corresponding intermediate use. We denote the quantity exported to n by $q_{n,k}^x$, and the production function in each sector by F_k. Define $D_{n,k} \equiv \tau_{n,k}^{1-\sigma_k}E_{n,k}/P_{n,k}^{1-\sigma_k}$.[25] Using

[25]Note that the iceberg trade cost assumption implies that the price received by the exporter is

$$p_{n,k}^x = (q_{n,k}^x)^{-\frac{1}{\sigma_k}} \cdot D_{n,k}^{\frac{1}{\sigma_k}}.$$

this notation, we can write the planner's problem as

$$\max_{\left\{q_k^{c,d}, q_{n,k}^{c,f}, q_k^{m,d}, q_{n,k}^{m,f}, q_{n,k}^{x}, L_{j,k}\right\}} \quad \ln U(\{q_k^{c,d}\}, \{q_{n,k}^{c,f}\})$$

$$s.t. \;\; F_k\left(\{L_{j,k}\}, \{q_k^{m,d}\}, \{q_{n,k}^{m,f}\}\right) = q_k^{c,d} + q_{n,k}^{m,d} + \sum_{n\in N} q_{n,k}^{x}, \;\; \forall k$$

$$\sum_k L_{j,k} = \bar{L}_j, \;\; \forall j$$

$$\sum_k \sum_n p_{n,k}^{f}\left(q_{n,k}^{c,f} + q_{n,k}^{m,f}\right) = \sum_k \sum_n (q_{n,k}^{x})^{\frac{\sigma_k - 1}{\sigma_k}} \cdot D_{n,k}^{\frac{1}{\sigma_k}}.$$

We first need to transform this into an expression involving FMA and CMA. Using the first order conditions, it is easy to show that at the optimum for any two export markets n and i

$$\frac{q_{n,k}^{x}}{q_{i,k}^{x}} = \frac{D_{n,k}}{D_{i,k}} \;\; \forall i, n \in N, \; k \in K,$$

Likewise, from the first order conditions and our CES aggregator for both consumption and intermediate goods, we have

$$\frac{q_{n,k}^{c,f}}{q_{i,k}^{c,f}} = \frac{q_{n,k}^{m,f}}{q_{i,k}^{m,f}} = \left(\frac{p_{n,k}^{f}}{p_{i,k}^{f}}\right)^{-\sigma_k} \quad \forall i, n \in N, \; k \in K.$$

This implies that we can define new variables $q_k^x = \sum_{n\in N} q_{n,k}^x$, $q_k^{c,f} = (\sum_{n\in N}(q_{n,k}^{c,f})^{\frac{\sigma_k-1}{\sigma_k}})^{\frac{\sigma_k}{\sigma_k-1}}$ and $q_k^{m,f} = (\sum_{n\in N}(q_{n,k}^{m,f})^{\frac{\sigma_k-1}{\sigma_k}})^{\frac{\sigma_k}{\sigma_k-1}}$ such that the problem above is equivalent to

$$\max_{\left\{q_k^{c,d}, q_k^{c,f}, q_k^{m,d}, q_k^{m,f}, q_k^{x}, L_{j,k}\right\}} \quad \ln U(\{q_k^{c,d}\}, \{q_k^{c,f}\})$$

$$s.t. \;\; F_k\left(\{L_{j,k}\}, \{q_k^{m,d}\}, \{q_k^{m,f}\}\right) = q_k^{c,d} + q_{n,k}^{m,d} + q_k^{x}, \;\; \forall k$$

$$\sum_k L_{j,k} = \bar{L}_j, \;\; \forall j$$

$$\sum_k \left(q_k^{c,f} + q_k^{m,f}\right) CMA_k^{\frac{1}{1-\sigma_k}} = \sum_{k\in K} (q_k^x)^{\frac{\sigma_k-1}{\sigma_k}} FMA_k^{\frac{1}{\sigma_k}}.$$

A simple application of the Envelope Theorem gives

$$\delta_k^x = \vartheta \cdot \frac{1}{\sigma_k},$$

where ϑ is the multiplier on the trade balance constraint. This multiplier equals the ratio of gross output to final expenditure (or GDP), $\vartheta = \frac{Y}{V}$ (Baqaee and Farhi, 2019).

This result follows directly from our definition of $d \ln FMA_k$ and the fact that, in an efficient economy, reallocation has no first order effect on welfare. A percentage increase in FMA_k causes a horizontal displacement of the foreign demand curve by the same percentage, and the welfare effect is given by the implied price increase (i.e. the vertical displacement) when quantity is held fixed. An alternative intuition is available by defining the export price index as $\sum_{k\in K} \lambda_k^x \ln p_k$; this result then implies that the welfare effect of a foreign shock is captured entirely by its effect on the terms of trade, with the factor $1/\sigma_k$ translating the demand shock into its implied effect on export prices.

Example without domestic distortions

The competitive equilibrium of an Armington economy is not generally welfare-maximizing from an individual country perspective, even when the economy is small and there are no domestic distortions. The economy faces downward-sloping demand for its products on international markets whenever $\sigma_k < \infty$. A welfare-maximizing planner would export in each sector to the point at which marginal revenue from exports equals marginal cost, while in the *laissez-faire* equilibrium the economy exports at the point for which price equals marginal cost. In contrast to the welfare-maximizing production allocation, the direct effect of a percentage increase in FMA_k under *laissez-faire* is an equal percentage increase in export quantity at fixed price, for any industry. This generates an increase in factor demand, leading to general equilibrium effects through changes in factor prices, goods prices and reallocation across industries that have first order welfare effects.

In the special case of a single-factor economy, the percentage increase in labor demand is the same regardless of the industry receiving the shock, and hence the general equilibrium impact on wages and domestic prices is the same. Since both the direct and indirect effects of shocks to FMA_k are identical for any two industries with the same initial export revenue share λ_k^x, the market access elasticities δ_k^x are also common across industries.

To see this most simply, assume upper tier Cobb-Douglas preferences with constant expenditure share e_k. The equilibrium conditions in this case specialize to the single equation

$$w\bar{L} = \sum_{k\in K} \left(\frac{w}{T_k}\right)^{1-\sigma_k} \cdot \left(z_k \frac{e_k \cdot w\bar{L}}{z_k \left(\frac{w}{T_k}\right)^{1-\sigma_k} + CMA_k} + FMA_k \right).$$

Taking natural logs of both sides and applying Taylor's theorem with respect to FMA_k, we get

$$d\ln w \approx \sum_{k\in K} \left(\lambda_k^d + (1-\sigma_k)\left(\lambda_k^d \theta_k^f + \lambda_k^x\right)\right) d\ln w + \sum_{k\in K} \lambda_k^x d\ln FMA_k.$$

The first term captures the effect of changes in wages on both foreign and domestic sales, accounting for both income and substitution effect, while the second term is the direct effect of changes in export market access.

Collecting terms and solving for $d\ln w$, we get

$$d\ln w \approx \sum_{k\in K} \frac{\lambda_k^x d\ln FMA_k}{1-\sum_{k'\in K}\left(\lambda_{k'}^d + (1-\sigma_{k'})\left(\lambda_{k'}^d\theta_{k'}^f + \lambda_{k'}^x\right)\right)}.$$

Using the Cobb-Douglas assumption and the results above, we can write

$$d\ln \mathbb{P} \approx \sum_{k\in K} e_k \left(\theta_k^d d\ln w + \frac{\theta_k^f}{1-\sigma_k} d\ln CMA_k\right).$$

Putting the two results together and solving, we get

$$d\ln y \approx \delta^x \cdot \sum_{k\in K} \lambda_k^x d\ln FMA_k,$$

with

$$\delta^x = \frac{\lambda^m}{1-\sum_{k'\in K}\left(\lambda_{k'}^d + (1-\sigma_{k'})\left(\lambda_{k'}^d\theta_{k'}^f + \lambda_{k'}^x\right)\right)},$$

where $\lambda^m = \sum_{k\in K}\lambda_k^m$.[26]

An interesting and instructive special case of this model is when $\lambda_k^m = 1,\ \forall k$ In that case, we have

$$\delta^x = \frac{1}{1+\sum_{k'\in K}\lambda_{k'}^x(\sigma_{k'}-1)},$$

and the welfare effect of a small export subsidy in sector k is positive if and only if

$$\sigma_k - 1 > \sum_{k'\in K}\lambda_{k'}^x(\sigma_{k'}-1).$$

The planner wishes to reallocate export activity toward the sectors with more elastic international demand, in order to exploit the country's monopoly power on international markets.

Example with external economies

Another reason our economy might deviate from efficiency is the presence of domestic distortions. These can take many forms in principle; we focus our discussion on external economies of scale in production at the sector level, a feature of many quantitative trade models (Kucheryavyy et al., 2020). The presence of external economies of scale implies that the *laissez-faire* equilibrium has some sectors smaller and some larger than socially optimal, and the effect of foreign demand shocks differs across sectors depending on which sectors ultimately expand or contract as a result.

To illustrate, consider a single factor economy with upper tier Cobb-Douglas preferences (as

[26]For a general homothetic upper tier, the formula would have to be modified to account for changes in industry expenditure shares, although the δ_k^x would still be common across industries.

above), but with external economies of scale as in Kucheryavyy et al. (2020). The cost function in each industry is $c_k = \frac{w}{T_k L_k^{\gamma_k}}$, with the parameter γ_k governing the scale economies in the sector. We specialize their model to the case with zero domestic sales in any industry. The equilibrium conditions can be expressed as

$$w\bar{L} = \sum_{k \in K} \left(\frac{w}{T_k L_k^{\gamma_k}} \right)^{1-\sigma_k} \cdot FMA_k$$

$$wL_k = \left(\frac{w}{T_k L_k^{\gamma_k}} \right)^{1-\sigma_k} \cdot FMA_k, \ \ \forall k \in K.$$

We assume that $\gamma_k(\sigma_k - 1) < 1$ for all industries to ensure a unique equilibrium that will be interior (and hence exhibit smooth comparative statics). Due to the zero domestic sales assumption, production and consumption are entirely distinct in this economy.

Solving the individual factor demand equations for L_k in terms of w and plugging them into the aggregate factor market clearing equation, we get

$$w\bar{L} = \sum_{k \in K} w^{\frac{(1+\gamma_k)(1-\sigma_k)}{1-\gamma_k(\sigma_k-1)}} \cdot FMA_k^{\frac{1}{1-\gamma_k(\sigma_k-1)}} \cdot T_k^{\frac{\sigma_k-1}{1-\gamma_k(\sigma_k-1)}}.$$

Using this expression, it is easy to see that

$$d\ln w \approx \kappa \sum_{k \in K} \left(\frac{1}{1-\gamma_k(\sigma_k-1)} \right) \lambda_k^x d\ln FMA_k$$

where

$$\kappa = \frac{1}{1 - \sum_{k' \in K} \frac{(1+\gamma_{k'})(1-\sigma_{k'})}{1-\gamma_{k'}(\sigma_{k'}-1)} \lambda_{k'}^x}.$$

For a stable interior equilibrium (ensured if $\gamma_k(\sigma_k - 1) < 1, \ \forall k$), the income elasticities to foreign shocks are given by

$$\delta_k^x = \frac{1}{1-\gamma_k(\sigma_k-1)} \cdot \frac{1}{1 - \sum_{k' \in K} \frac{(1+\gamma_{k'})(1-\sigma_{k'})}{1-\gamma_{k'}(\sigma_{k'}-1)} \lambda_{k'}^x} \ \ \forall k \in K.$$

All else equal, foreign demand shocks in sectors with larger external economies generate larger welfare effects. The intuition for this result is simple: holding factor prices fixed, the supply curve is downward sloping with elasticity γ_k. An expansion of foreign demand results in a movement down the supply curve, with the benefits of higher quantity sold moderated by the associated terms of trade losses. Scale economies are more valuable in sectors with more elastic international demand; with less elastic demand, achieving higher productivity comes at the expense of lower export prices.[27]

[27] The same fundamental intuition applies when there are positive domestic sales, although the formula must be modified to account for the heterogeneous impact of foreign demand shocks on domestic prices.

A.4 Quantitative model details

Equilibrium. A competitive equilibrium in this economy is a sequence of goods prices $\{P_{kt}\}\ \forall k,t$, factor prices $\{w_t, r_t\}\ \forall t$, factor allocations $\{L_{kt}, \mathcal{K}_{kt}\}\ \forall k,t$, and goods market allocations such that (i) consumers maximize utility; (ii) firms maximize profits; (iii) markets clear.

Denote by Y_{kt} the gross revenue of sector k. The market clearing condition for output of sector k at time t is:

$$Y_{kt} = \frac{z_k c_{kt}^{1-\sigma_k}}{P_{kt}^{1-\sigma_k}} \left(P_{kt} C_{kt} + P_{kt} I_{kt} + \sum_{l \in K} \widetilde{\alpha}_{k,l} Y_{lt} \right) + c_{kt}^{1-\sigma_k} FMA_{kt}.$$

The second term is the sector's exports at time t. The first term is domestic sales. The domestic sales are a product of all final and intermediate expenditures on sector k products and the share of the total sector k domestic absorption that is spent on domestically-produced goods, $z_k \left(c_{kt}/P_{kt}\right)^{1-\sigma_k}$.

Steady state. We drop the time subscripts to denote steady state values. The price of installed capital and the investment price index are proportional:

$$r = \mathbb{P}^I(\rho^{-1} + \chi - 1).$$

Let $Y = \sum_k Y_k$ denote the steady state aggregate gross revenue in this economy. The steady state capital stock is:

$$\mathcal{K} = \mu(1-\beta)\frac{Y}{\mathbb{P}^I(\rho^{-1} + \chi - 1)}.$$

Since the capital stock is constant in steady state, investment is simply: $I = \chi\mathcal{K}$. Hence, investment expenditure is a constant fraction of aggregate gross revenue:

$$\mathbb{P}^I I = \frac{\mu\chi(1-\beta)}{(\rho^{-1} + \chi - 1)} Y.$$

Since GDP is also a constant fraction of gross revenue ($\mathbb{P}^C C + \mathbb{P}^I I = \mu Y$) it follows that consumption expenditure is as well:

$$\mathbb{P}^C C = \mu \left(1 - \frac{(1-\beta)\chi}{(\rho^{-1} + \chi - 1)} \right) Y.$$

The combined consumption and investment expenditure on sector k goods can then be expressed as:

$$P_k C_k + P_k I_k = f_k \mu Y,$$

where $f_k \equiv e_k \left(1 - \frac{(1-\beta)\chi}{(\rho^{-1}+\chi-1)}\right) + \nu_k \frac{(1-\beta)\chi}{(\rho^{-1}+\chi-1)}$ is the constant steady state share of total final expenditure going to sector k. Thus, the steady state of this economy is characterized by the following system of equations:

$$Y_k = \frac{z_k c_{kt}^{1-\sigma_k}}{P_{kt}^{1-\sigma_k}} \sum_l \left(f_k \mu + \widetilde{\alpha}_{k,l}\right) Y_l + c_k^{1-\sigma_k} FMA_k \;\; \forall k \tag{A.18}$$

$$P_k^{1-\sigma_k} = z_k c_k^{1-\sigma_k} + CMA_k \;\; \forall k \tag{A.19}$$

$$c_k = T_k L_k^{-\gamma_k} w^{\beta\mu} \prod_l P_l^{\widetilde{\alpha}_{l,k} + \mu(1-\beta)\nu_l} \;\; \forall k \tag{A.20}$$

$$w L_k = \mu \beta Y_k \;\; \forall k \tag{A.21}$$

$$\sum_k L_k = L. \tag{A.22}$$

Mapping to regression coefficients. Note that while in the statement of equilibrium conditions (A.18), FMA_k enters by itself, in actual empirical estimation the regressor is weighted by the export share: $\lambda_k^x FMA_k$, see (2.7). Thus, we state the model solution directly in terms of export-share-weighted firm market access: $FMA_k^W \equiv \lambda_k^x FMA_k$. This way, the model solution is directly comparable to the regression coefficients.

Analytical solution. To first order, the vectors of log changes in revenues and prices following a vector of export-share-weighted firm market access shocks $d \ln \mathbf{FMA^W}$ are given by:

$$\begin{aligned} d\ln \mathbf{Y} &= \left\{ \mathbf{I} - \left[\mathbf{\Pi}^d + (\mathbf{I} - \boldsymbol{\sigma}) \left((\mathbf{I} - \boldsymbol{\pi}^x) \left(\mathbf{I} - \boldsymbol{\theta}^d \right) + \boldsymbol{\pi}^x \right) \left(\mathbf{I} - \mathbf{A}\boldsymbol{\theta}^d \right)^{-1} \left(\gamma \left(\boldsymbol{\lambda} \otimes \mathbf{1} - \mathbf{I} \right) + \mu\beta \boldsymbol{\lambda} \otimes \mathbf{1} \right) \right] \right\}^{-1} \\ &\quad \times diag\left(\boldsymbol{\lambda}\right)^{-1} d \ln \mathbf{FMA^W} \end{aligned} \tag{A.23}$$

$$d \ln \mathbf{P} = \boldsymbol{\theta}^d \left(\mathbf{I} - \mathbf{A}\boldsymbol{\theta}^d \right)^{-1} \left(\gamma \left(\boldsymbol{\lambda} \otimes \mathbf{1} - \mathbf{I} \right) + \mu\beta \boldsymbol{\lambda} \otimes \mathbf{1} \right) d \ln \mathbf{Y}. \tag{A.24}$$

In these expressions, the matrices are defined as follows:

- In $\mathbf{\Pi}^d$ each *row* represents the domestic absorption shares by sectors in the column of the sector in the row:

$$\mathbf{\Pi}^d \equiv \begin{bmatrix} \pi_{1,1}^d & \pi_{1,2}^d & \cdots & \pi_{1,K}^d \\ \vdots & & \ddots & \vdots \\ \pi_{K,1}^d & \cdots & & \pi_{K,K}^d \end{bmatrix},$$

  where $\pi_{k,l}^d \equiv \frac{\theta_k^d \left(f_k \mu + \widetilde{\alpha}_{k,l}\right) Y_l}{Y_k}$.

- A diagonal matrix of export absorption shares

$$\boldsymbol{\pi}^x \equiv \begin{bmatrix} \pi_1^x & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \pi_K^x \end{bmatrix},$$

where $\pi_k^x \equiv \frac{c_k^{1-\sigma_k} FMA_k}{Y_k}$.

- Matrix **A** where each *column* represents the use of the sector in the column as an intermediate input by the sector in the row:

$$\mathbf{A} \equiv \begin{bmatrix} \widetilde{\alpha}_{1,1} + (1-\beta)\,\mu\nu_1 & \widetilde{\alpha}_{2,1} + (1-\beta)\,\mu\nu_2 & \cdots & \widetilde{\alpha}_{K,1} + (1-\beta)\,\mu\nu_K \\ \widetilde{\alpha}_{1,2} + (1-\beta)\,\mu\nu_1 & \widetilde{\alpha}_{2,2} + (1-\beta)\,\mu\nu_2 & \cdots & \widetilde{\alpha}_{K,2} + (1-\beta)\,\mu\nu_K \\ \vdots & & \ddots & \vdots \\ \widetilde{\alpha}_{1,K} + (1-\beta)\,\mu\nu_1 & \widetilde{\alpha}_{2,K} + (1-\beta)\,\mu\nu_2 & \cdots & \widetilde{\alpha}_{K,K} + (1-\beta)\,\mu\nu_K \end{bmatrix}.$$

- A diagonal matrix of expenditure shares in each sector sourced from domestic producers:

$$\boldsymbol{\theta^d} \equiv \begin{bmatrix} \theta_1^d & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_K^d \end{bmatrix},$$

where $\theta_k^d \equiv \frac{z_k c_k^{1-\sigma_k}}{P_k^{1-\sigma_k}}$.

- Row vector of gross revenue shares:

$$\boldsymbol{\lambda} \equiv \begin{bmatrix} \lambda_1 & \cdots & \lambda_K \end{bmatrix},$$

where $\lambda_k \equiv \frac{Y_k}{\sum_l Y_l}$ is the gross revenue share of sector k, and $diag\,(\boldsymbol{\lambda})$ is a diagonal matrix with entries of $\boldsymbol{\lambda}$.

- Diagonal matrices collecting substitution and scale elasticities:

$$\boldsymbol{\sigma} \equiv \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_K \end{bmatrix}$$

$$\boldsymbol{\gamma} \equiv \begin{bmatrix} \gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_K \end{bmatrix}.$$

Real GDP. Since in the empirical estimation our independent variable is real GDP, we need to translate the changes in nominal revenue and prices (A.23)-(A.24) into changes in real GDP, which we define as:

$$y = \frac{wL + r\mathcal{K}}{\mathbb{P}},$$

where the price index $\mathbb{P} \equiv \left(1 - \frac{(1-\beta)\chi}{(\rho^{-1}+\chi-1)}\right)\mathbb{P}^C + \frac{(1-\beta)\chi}{(\rho^{-1}+\chi-1)}\mathbb{P}^I$ is the share-weighted average of the consumption and investment price indices. It is immediate that the log change in this price index is:

$$d \ln \mathbb{P} \quad = \quad \mathbf{f} \cdot d \ln \mathbf{P},$$

where the $1 \times K$ row vector $\mathbf{f}$ collects final shares. The real GDP change is thus

$$d \ln y = \boldsymbol{\lambda} \cdot d \ln \mathbf{Y} - \mathbf{f} \cdot d \ln \mathbf{P}. \tag{A.25}$$

Plugging (A.23)-(A.24) into (A.25), we obtain the following model elasticities with respect to foreign market access shocks:

$$\frac{d \ln y}{d \ln \mathbf{FMA^W}} \quad = \quad \left(\boldsymbol{\lambda} - \mathbf{f} \cdot \boldsymbol{\theta}^d \left(\mathbf{I} - \mathbf{A}\boldsymbol{\theta}^d\right)^{-1} \left(\gamma\left(\boldsymbol{\lambda} \otimes \mathbf{1} - \mathbf{I}\right) + \mu\beta\boldsymbol{\lambda} \otimes \mathbf{1}\right)\right) \frac{d \ln \mathbf{Y}}{d \ln \mathbf{FMA^W}}, \tag{A.26}$$

where $\frac{d \ln \mathbf{Y}}{d \ln \mathbf{FMA^W}}$ is given by (A.23).

The term in parentheses translates gross revenue changes into real income changes, since gross revenues affect both aggregate nominal value added (weighted according to sector size $\boldsymbol{\lambda}$), and the price index, captured by the second term. The vector of elasticities in (A.26) is the theoretical counterpart of the econometrically estimated elasticities of GDP with respect to foreign demand shocks, δ^x.

Calibration. We set the value added share in gross output $\mu = 0.5$ and the labor share in value added to $\beta = 2/3$. To calibrate the model, we need to parameterize the matrices and vectors $\boldsymbol{\lambda}$, $\mathbf{f}$, $\boldsymbol{\theta}^d$, $\mathbf{A}$, $\boldsymbol{\pi}^x$, and the vector $\boldsymbol{\nu}$ that collects investment expenditure shares ν_k. All other objects comprising the model solution are transformations of these. Since the coefficient estimates of the growth impacts of foreign shocks are at the cluster level, we parameterize our model for the 6 tradeable sector clusters from the econometric estimation, plus a seventh non-tradeable services sector. We use the COMTRADE data to construct the representative export and import shares by taking a simple average for each cluster across our sample. The matrices $\mathbf{f}$, $\boldsymbol{\theta}^d$, $\boldsymbol{\nu}$ and $\mathbf{A}$ describe domestic sectoral expenditure shares. Since this information is not available in the COMTRADE and Penn World Tables datasets used in the econometric estimation, we obtain these from the World Input-Output Database (Timmer et al., 2015, henceforth WIOD) as simple averages across the values for the 40 countries available in that database, after mapping the WIOD sectors to our clusters. The matrices $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}^x$ are constructed from the two datasets so as to ensure that all adding up constraints are satisfied, i.e. they are completely determined by the export and import shares plus $\mathbf{f}$, $\boldsymbol{\theta}^d$, $\boldsymbol{\nu}$ and $\mathbf{A}$. Overall, since this calibration is for one "typical" country, it is not very data-intensive and the shares we feed into the model are straightforward.

Direct, first-, and higher-order decomposition. To decompose the overall GDP elasticity to FMA into different-order effects, define an "impact matrix":

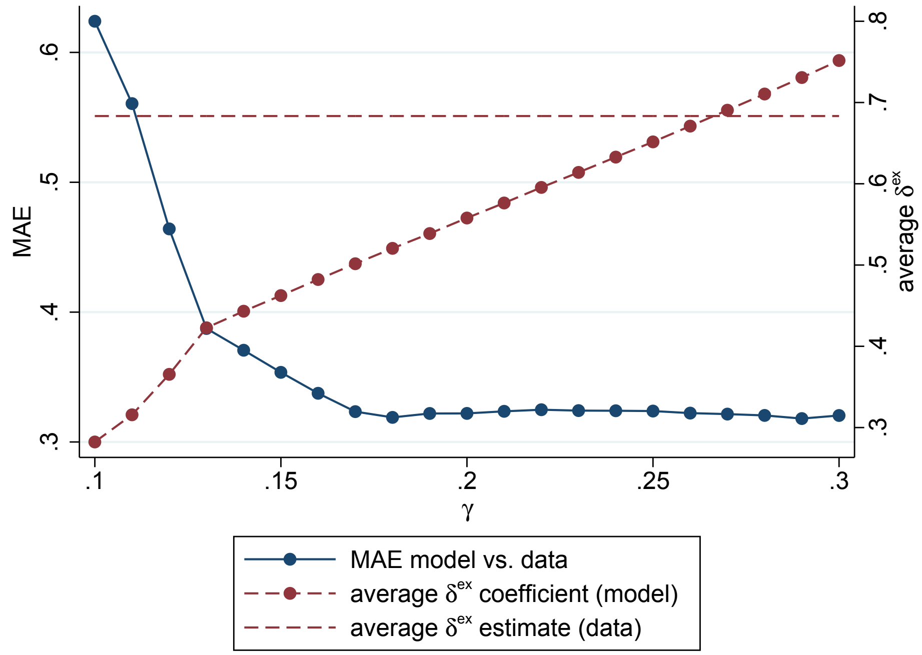
$$\boldsymbol{\Omega} \equiv \boldsymbol{\Pi}^{d} + (\mathbf{I}-\boldsymbol{\sigma})\left((\mathbf{I}-\boldsymbol{\pi}^{x})\left(\mathbf{I}-\boldsymbol{\theta}^{d}\right)+\boldsymbol{\pi}^{x}\right)\left(\mathbf{I}-\mathbf{A}\boldsymbol{\theta}^{d}\right)^{-1}\left(\gamma\left(\boldsymbol{\lambda}\otimes\mathbf{1}-\mathbf{I}\right)+\mu\beta\boldsymbol{\lambda}\otimes\mathbf{1}\right). \quad \text{(A.27)}$$

Then, the model solution can be stated as:

$$\begin{aligned}\frac{d\ln\mathbf{Y}}{d\ln\mathbf{FMA^{W}}} &= \{\mathbf{I}-\boldsymbol{\Omega}\}^{-1}\times diag\left(\boldsymbol{\lambda}\right)^{-1} \\ &= \left(\underbrace{\mathbf{I}}_{\text{direct}}+\underbrace{\boldsymbol{\Omega}}_{\text{first-order}}+\underbrace{\boldsymbol{\Omega}^{2}+\boldsymbol{\Omega}^{3}+\ldots}_{\text{higher-order}}\right)\times diag\left(\boldsymbol{\lambda}\right)^{-1},\end{aligned}$$

where the second line writes the Leontief inverse as an infinite expansion. The first term is the direct effect of a foreign demand in a sector. The matrix is diagonal, and thus the direct effect only applies to the sector experiencing the foreign demand shock. The first-order effect is given by the impact matrix $\boldsymbol{\Omega}$. Examining (A.27), the first-order effect is in turn comprised of two terms. The first, $\boldsymbol{\Pi}^{d}$, reflects the fact that the sector experiencing a foreign demand shock changes its purchases of intermediates, and the change in its value added affects final demand inside the home economy. The second term captures the change in unit costs that follows the change in foreign demand. It can be written more compactly as $(\mathbf{I}-\boldsymbol{\sigma})\left((\mathbf{I}-\boldsymbol{\pi}^{x})\left(\mathbf{I}-\boldsymbol{\theta}^{d}\right)+\boldsymbol{\pi}^{x}\right)\frac{d\ln\mathbf{c}}{d\ln\mathbf{FMA^{W}}}$. The unit costs will change both because of the fact that factor reallocation affects production scale (captured by $\gamma\left(\boldsymbol{\lambda}\otimes\mathbf{1}-\mathbf{I}\right)$), and because of general equilibrium impacts on economywide wages (captured by $\mu\beta\boldsymbol{\lambda}\otimes\mathbf{1}$). The change in costs will in turn change foreign sales (by $(\mathbf{I}-\boldsymbol{\sigma})\boldsymbol{\pi}^{x}$), as well as domestic sales (by $(\mathbf{I}-\boldsymbol{\sigma})(\mathbf{I}-\boldsymbol{\pi}^{x})\left(\mathbf{I}-\boldsymbol{\theta}^{d}\right)$). Finally, the higher-order effects propagate these shocks further, as sectors affected by the initially shocked sector in turn change their demand for other sectors' output as well as the relative costs. The Leontief inversion of the impact matrix captures these infinite-order effects.

Figure A1: Theory Diagnostics: MAE and Average δ^{ex}



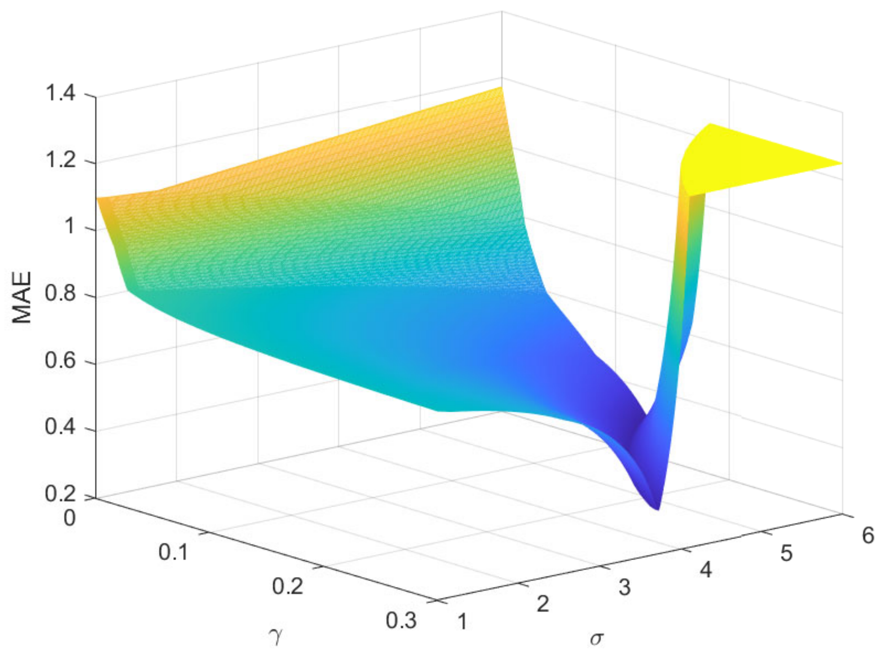
Notes: This figure plots the Mean Absolute Error (MAE) between model and data δ^{ex} (left axis) against the value of γ. For each value of γ displayed, σ is selected to minimize MAE. The figure also plots the average δ^{ex} in the model and the data (left axis).

Table A1: δ^x Coefficients: Data and Model

	$\widehat{\delta^x}$ data	δ^x model ($\gamma = 0.29 \quad \sigma = 3.19$)				
		lowest MAE	no inter-mediates	no capital	$v_k = e_k$	tradeable services
RAW	0.17	-0.13	0.22	-0.13	-0.13	-0.20
INT	1.32	0.82	0.32	0.34	0.78	0.43
CAP	2.97	2.97	0.54	1.76	2.59	1.73
CONS	-0.52	0.12	0.25	0.19	0.26	0.04
AG	0.11	0.10	0.26	0.14	0.21	0.05
MIN	0.05	0.51	0.14	0.21	0.43	0.16
Average δ^x	0.68	0.73	0.29	0.42	0.69	0.42
MAE model vs. data		0.32	0.75	0.56	0.41	0.56
Corr. model vs. data		0.94	0.91	0.90	0.93	0.90

Notes: This table reports the econometric estimates of $\widehat{\delta^x}$ (first column), and the model δ^x under the values of γ and σ that minimize Mean Absolute Error (MAE) between model and data (second column), and under the alternative model structures (last four columns). The bottom panel reports the average δ^x for each case, and for the theoretical models reports the MAE and the correlation of δ^x implied by the model and data.

Figure A2: Theory Diagnostics: MAE as a Function of σ and γ



Notes: This figure plots the Mean Absolute Error (MAE) between model and data δ^x (left axis) against the parameters γ and σ. The values in the right corner of the figure (high γ and high σ are truncated at 1.2 for increased readability).

B. Data and Estimation Appendix

B.1 Matching the Trade Data to Industries

The international trade data from 1965 to 2015 are from the UN COMTRADE Database, which reports bilateral trade flows at the 4-digit SITC Revision 2 level. To concord the trade data to the 1997 NAICS industry classification, we proceed as follows. First, we assign each 4-digit SITC item to its corresponding 6-digit NAICS industries. For instance, 7511 *Typewriters cheque-writing machines* are matched to 333313 *Office machinery manufacturing*. Second, for those items that are matched to more than one 6-digit NAICS industry, we check whether it could be assigned to the upper-level 5-digit industry. For example, 8510 *Footwear* is matched to 316211 *Rubber and plastics footwear manufacturing*, 316212 *House slipper manufacturing* and some other 6-digit NAICS industries with the first 5 digits "31612." In this case, we aggregate these 6-digit NAICS industries to the 5-digit one ("31612"), and concord the 4-digit SITC items to the 5-digit NAICS industry. Third, the same is done for the items that are assigned to more than one 5-digit NAICS industry. We matched them to the corresponding 4-digit NAICS industries.

Overall, the 784 4-digit SITC items are matched to 268 NAICS industries. Among them, 233 industries are in the manufacturing sector, 26 in agriculture, and 9 in mining.

B.2 K-means Clustering

B.2.1 Selecting the Number of Clusters with Silhouette Analysis

There is no unambigously optimal method for selecting a number of clusters. We follow Rousseeuw (1987), who introduces the silhouette plot for this purpose. With this method, each cluster is represented by a silhouette displaying which points lie well within the cluster and which ones are marginal to the cluster. The silhouette plot is based on the silhouette width measure, which compares the similarity (cohesion) of a point to points in its own cluster with the ones in neighboring clusters (separation).

The silhouette width s_k is measured as follows:

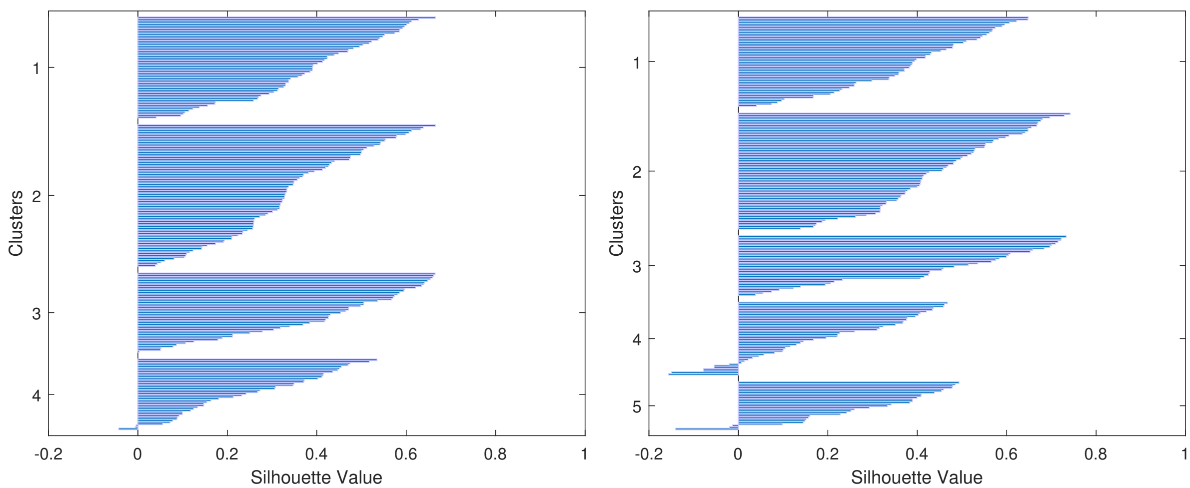
1. (*Measuring the cohesion*) Denote by a_k the average distance between point k and all other points in the same cluster.

2. (*Measuring the separation*) Denote by b_k the average distance between k and all points in the nearest cluster.

3. The silhouette width of the observation k is measured as $s_k = \frac{b_k - a_k}{max(a_k, b_k)}$.

The silhouette ranges from -1 to 1, where a high value indicates that the point is well assigned to its own cluster and dissimilar to neighboring clusters. A value of 0 indicates that the point is on

Figure A3: Silhouette Analysis



(a) Number of Clusters = 4 (b) Number of Clusters = 5

Notes: This figure plots the silhouette values for each industry, when there are 4 clusters (left panel), and 5 clusters (left panel).

or very close to the cluster boundary between two neighboring clusters and negative values indicate that those points might have been assigned to the wrong cluster.

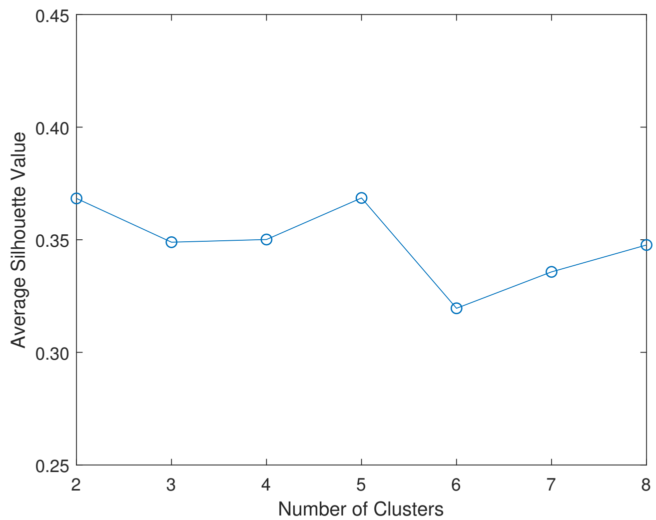
The average silhouette width provides one evaluation of clustering validity, and can be used as an input into selecting an appropriate number of clusters. A high average silhouette width indicates a strong clustering. The average silhouette method computes the average silhouette of observations for different numbers of clusters G. All else equal, one should prefer values of G with larger values of the average silhouette. Another desirable characteristic of a good clustering scheme is that it minimizes potential misclassificatin; that is, it should have few observations with zero or negative silhouette.

Figure A3 plots the silhouette width for industries in each cluster when there are 4 and 5 clusters, and Figure A4 plots the average silhouette value over the range of cluster numbers from 2 to 8. The silhouette analysis suggests that either 4 or 5 are good values for the number of clusters. While the average silhouette value is slighly higher for 5 clusters than 4, the silhouette analysis suggests that with 4 clusters fewer industries are potentially miscalssified.

B.2.2 K-means Clustering Using a Subset of Characteristic Variables

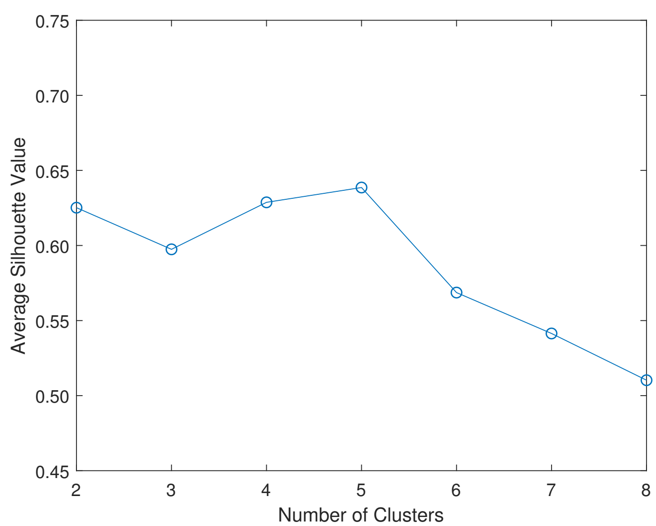
The average silhouette value under 4 clusters is about 0.35, which indicates that the cluster structure is somewhat weak. However, this could be due to the inclusion of irrelevant sectoral characteristics, which tend to drag down the average silhouette value. We investigate this hypothesis by implementing the algorithm on a subset of important characteristic variables: the investment sales share, intermediates sales shares and contract intensity. These variables are identified as especially impor-

Figure A4: Average Silhouette Value



Notes: This figure plots the average silhouette values across industries for the number of clusters on the x-axis.

Figure A5: Average Silhouette Value, Using a Subset of Sector Characteristics



Notes: This figure plots the average silhouette values across industries for the number of clusters on the x-axis, when using only a subset of sector characteristics for the clustering procedure.

tant through inspection of the cluster structure as well as more formally using methods developed in Witten and Tibshirani (2010). The average silhouette value is now about 0.65 (Figure A5), suggesting a strong cluster structure. Table A2 reports the summary statistics for sectoral characteristics of each cluster. The 4 clusters based on these three characteristics closely replicate the baseline cluster structure reported in Table 1.

Table A2: Summary Statistics of Clusters: K-means Clustering Using a Subset of Sector Characteristics

	cluster					
	1	2	3	4	Mean	Std. Dev.
Investment Share	0.01	0.07	0.56	0.05	0.13	0.22
Intermediates, Using	0.70	0.63	0.65	0.63	0.66	0.16
Intermediates, Sales	0.83	0.78	0.28	0.25	0.57	0.31
Concentration Ratio	0.41	0.30	0.34	0.48	0.40	0.21
Skill Intensity	0.30	0.31	0.35	0.33	0.32	0.13
Capital Intensity	0.64	0.55	0.55	0.64	0.61	0.10
Contract Intensity	0.29	0.65	0.72	0.57	0.51	0.22
Number of industries	87	45	42	59		
Trade share	0.38	0.16	0.20	0.19		

Notes: This table reports the summary statistics of the sectoral characteristics among the sectors selected into each cluster, when only a subset of sectoral characteristics is used in the clustering procedure. The last two columns report the mean and standard deviations of those characteristics among all manufacturing sectors. The row "Number of industries" reports the number of sectors in each cluster, and "Trade share" reports the fraction of world trade accounted for by sectors in that cluster.

Table A3: The 3 Most Representative Sectors in Each Cluster

Cluster	Label	Representative Sectors	
		Naics	Description
1	Raw Materials Processing	324199	All Other Petroleum and Coal Products Manufacturing
		31131	Sugar Manufacturing
		32419	Other Petroleum and Coal Products Manufacturing
2	Complex Intermediates	33512	Lighting Fixture Manufacturing
		33531	Electrical Equipment Manufacturing
		339994	Broom, Brush, and Mop Manufacturing
3	Capital Goods	333911	Pump and Pumping Equipment Manufacturing
		333994	Industrial Process Furnace and Oven Manufacturing
		333992	Welding and Soldering Equipment Manufacturing
4	Consumer Goods	312130	Wineries
		335211	Electric Housewares and Household Fan Manufacturing
		33521	Small Electrical Appliance Manufacturing

Notes: This table lists the 3 sectors closest to the cluster centroid for each cluster.

B.3 Estimation of $FMA_{ik,t}$

The foreign demand shocks are estimated by using sectoral bilateral trade flow data and a structural gravity equation. Equation (2.4) relates external Firm Market Access to the gravity equation. The

$FMA_{ik,t}$ for exporter i are expressed as follows:

$$FMA_{ik,t} = \sum_{n \neq i} \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}} \cdot \tau_{in,k}^{1-\sigma_k}.$$

The gravity equation (2.3) can be rewritten as

$$E_{ink,t} \equiv p_{ink,t} \cdot q_{ink,t} = c_{ik,t}^{1-\sigma_k} \cdot \frac{E_{nk,t}}{P_{nk,t}^{1-\sigma_k}} \cdot \tau_{ink,t}^{1-\sigma_k}, \tag{B.1}$$

where $E_{ink,t}$ denotes country n's total sector k expenditure on goods from country i. We do not observe the domestic trade flows. We estimate the share version of this equation *à la* Eaton et al. (2012). Dividing both sides by the total imports of country n, we get

$$\frac{E_{ink,t}}{\sum_{i' \neq n} E_{i'nk,t}} = c_{ik,t}^{1-\sigma_k} \cdot \frac{E_{nk,t}}{P_{nk,t}^{1-\sigma_k} \cdot \sum_{i' \neq n} E_{i'nk,t}} \cdot \tau_{ink,t}^{1-\sigma_k}.$$

It can be estimated by regressing bilateral trade flows on exporter and importer fixed effects and bilateral geographic distance measures. The estimating equation is (4.2) in the main text.

Shocks to large countries may affect their trading partners' estimated importer and exporter effects. In that case, those estimated fixed effects would not be pure measures of foreign shocks affecting the large country, as they would pick up in part the large country's domestic shocks. To address this potential endogeneity, we carry out the above gravity estimation using the leave-one-out approach. For each country ω, we estimate a set $\{\kappa_{nk,t}^{m}(\omega) \ \kappa_{ik,t}^{x}(\omega) \ \zeta_{kt}(\omega) \ \xi_{kt}(\omega)\}$ of country ω-specific importer and exporter fixed effects and distance/contiguity coefficients by dropping country ω from the gravity sample on both the exporter and importer side. In this notation, indexing by ω denotes estimates when country ω is left out of the sample. In practice this does not affect any of our conclusions. The results are very similar if we extract the importer and exporter fixed effects from the simple gravity regression with all countries included. This reflects the fundamental fact that most countries are small in foreign markets.

The fixed effects of log trade flows are identified only up to a sector-time-specific additive constant, and thus we renormalize them by restricting the sum of the log importer fixed effects to be zero:

$$\overline{\ln \kappa}_{nk,t}^{m}(\omega) = \ln \kappa_{nk,t}^{m}(\omega) - \frac{\sum_z \ln \kappa_{zk,t}^{m}(\omega)}{N_{kt}(\omega)}$$

where $N_{kt}(\omega)$ is the total number of countries with positive imports for industry k and time t when ω is left out. In this way, what matters is the share of each country in the total imports across industries, not the total imports of the numéraire country in the fixed effects estimation. The estimated $FMA_{ik,t}$ is then be computed as in (4.3), where, with some abuse of notation, $\kappa_{nk,t}^{m}$ denote the renormalized importer fixed effects when country ω is omitted. These importer fixed effects are estimates of the destination-n demand shifter $\frac{E_{nk,t}}{P_{nk,t}^{1-\sigma_k}\cdot\sum_{i\neq n}E_{ink,t}}$. The iceberg bilateral components $\tau_{ink,t}^{1-\sigma_k}$ are estimated by using the bilateral geographic distance and the common border dummy and corresponding distance and common border coefficients. The estimated bilateral component is proxied by $Distance_{in}^{\zeta_{kt}} \cdot \exp(\xi_{kt} \cdot Contig_{in})$.

B.4 The Post-Double-Selection Method

B.4.1 Estimating Equation

The estimating equation is

$$d\ln y_{i,t} \approx \kappa + \sum_{g\in G} \delta_g^x \cdot \left[\lambda_{ig,t}^x d\ln FMA_{ig,t}\right] + \zeta d\ln \mathbf{a}_{it} + \eta \mathbf{x}_{i,t} + \varepsilon_{i,t},$$

where the $d\ln FMA_{ig,t} = \sum_{k\in G} \lambda_{ik,t}^x d\ln FMA_{ik,t}$ are the log-differenced market access terms aggregated up to the cluster level. In describing the procedure, to streamline exposition we omit the fact that time fixed effects and the log initial per capita income are "protected regressors," that are always included and not subject to the control set selection procedure.

The vector $\mathbf{x}_{i,t}$ collects the industry-level initial equilibrium variables such as initial import and export shares ($\lambda_{ik,t}^m$ and $\lambda_{ik,t}^x$), and weighted initial firm and consumer market access ($\lambda_{ik,t}^x \cdot \ln FMA_{ik,t}$ and $\lambda_{ik,t}^m \cdot \ln CMA_{ik,t}$). The vector $d\ln \mathbf{a}_{it}$ collects the observed contemporaneous foreign supply shocks, i.e. $\lambda_{ig,t}^m d\ln CMA_{ig,t}$.

Since our estimating equation has a large number of controls relative to the sample size, the OLS estimation is infeasible, and dimension reduction is necessary. We estimate the above growth equation by implementing the "post-double-selection" method of Belloni et al. (2014b, 2017). We describe our implementation of the estimator below.

B.4.2 Post-Double-Selection Method

The post-double-selection procedure works in two steps. In the double-selection step, LASSO is applied to select control variables that are useful for predicting the dependent and independent variables respectively. In the post-selection step, coefficients are estimated via an OLS regression of dependent variables on the independent variables and the union of selected controls.

First, let's rewrite the estimation equation as follows:

$$d\ln y_{i,t} = \mathbf{d}_{i,t}\boldsymbol{\delta} + \mathbf{x}_{i,t}\boldsymbol{\beta}_y + \mu_{i,t},$$

where $\mathbf{d}_{i,t}$ denotes the vector of treatment variables $\lambda_{ig,t}^x d\ln FMA_{ig,t}$, and $\mathbf{x}_{i,t}$ is the vector of control variables, that with some abuse of notation now also includes $d\ln \mathbf{a}_{it}$.

Applying LASSO directly to our estimation equation above might lead to the omitted-variable bias if the LASSO procedure drops a control variable that is highly correlated with the treatment but the coefficient associated with the control is nonzero. To learn about the relationship between the treatment variables and the controls, let's introduce a reduced-form equation

$$d_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_d + v_{i,t}$$

for each element $d_{i,t}$ of the vector $\mathbf{d}_{i,t}$.

Substituting the reduced-form $d_{i,t}$ into the growth estimation equation we get

$$\begin{aligned} d \ln y_{i,t} &= \mathbf{x}_{i,t}(\boldsymbol{\beta}_d \boldsymbol{\delta} + \boldsymbol{\beta}_y) + (v_{i,t}\boldsymbol{\delta} + \mu_{i,t}) \\ d_{i,t} &= \mathbf{x}_{i,t}\boldsymbol{\beta}_d + v_{i,t} \quad \forall d_{i,t}. \end{aligned}$$

Both equations are used for variable selection. The first equation is used to select a set of variables that are useful for predicting the dependent variable $d \ln y_{i,t}$ and the second equation is used to select a set of controls that are useful for predicting each of the treatment variables $d_{i,t}$. The reduced form system could be further rewritten as

$$\mathbf{z}_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \varepsilon_{i,t}$$

where $\mathbf{z}_{i,t}$ is the vector of dependent variable $d \ln y_{i,t}$ and all treatment variables $d_{i,t}$. A feasible double-selection procedure via LASSO is then defined as follows

$$\min_{\boldsymbol{\beta}} E(\mathbf{z}_{i,t} - \mathbf{x}_{i,t}\boldsymbol{\beta})^2 + \frac{\lambda}{n}||L\boldsymbol{\beta}||_1$$

where $L = diag(l_1, l_2, \ldots, l_p)$ is a diagonal matrix of penalty loadings and λ is the penalty level. The LASSO estimator is used for variable selection by simply selecting the controls with nonzero estimated coefficients.

The double-selection procedure first selects a set of controls that are useful for predicting the independent variable $d \ln y_{i,t}$ and treatment variables $\mathbf{d}_{i,t}$. Then in the post-LASSO step, we estimate δ_g^x by ordinary least squares regression of $d \ln y_{i,t}$ on $\mathbf{d}_{i,t}$ and the union of the variables selected for predicting $d \ln y_{i,t}$ and $\mathbf{d}_{i,t}$.

B.4.3 K-fold Cross Validation

The penalty level λ controls the degree of penalization. Practical choices for λ to prevent overfitting are provided in Belloni et al. (2012, 2014a,b). We follow the online appendix of Belloni et al. (2014a) and choose λ by K-fold cross validation.

The K-fold cross-validation works as follows:

1. Randomly split the data $(y_{i,t}, \mathbf{x}_{i,t}, \mathbf{d}_{i,t})$ into K subsets of equal size, $S_1, S_2, \ldots, S_K$
2. Set the potential tuning parameter set to be $[\lambda^{RT} - 100 : grid : \lambda^{RT} + 100]$, where $\lambda^{RT} = 2.2\sqrt{n}\Phi(1 - \gamma/2p)$ is the rule of thumb tuning parameter suggested in Belloni et al. (2012, 2014b),

$\gamma = 0.1/log(p)$, n is the number of observations, p the number of variables, and $grid = 10$.

3. Given λ, for $k = 1, 2, \ldots, K$:

   (a) (*Training on* $(y_{i,t}, \mathbf{x}_{i,t}, \mathbf{d}_{i,t})$, $i \notin S_k$) Leave the kth subset out, and implement the post-double-selection method with tuning parameter λ on the $K-1$ subsets. Denote the estimated coefficients as $\hat{\boldsymbol{\delta}}^{-k}(\lambda)$ and $\hat{\boldsymbol{\beta}}_y^{-k}(\lambda)$.

   (b) (*Validating on* $(y_{i,t}, \mathbf{x}_{i,t}, \mathbf{d}_{i,t})$, $i \in S_k$) Given $\hat{\boldsymbol{\delta}}^{-k}(\lambda)$ and $\hat{\boldsymbol{\beta}}_y^{-k}(\lambda)$ compute the error in predicting the kth subset,

$$e_k(\lambda) = \sum_{i \in S_k} (d \ln y_{i,t} - \mathbf{d}_{i,t}\hat{\boldsymbol{\delta}}^{-k}(\lambda) - \mathbf{x}_{i,t}\hat{\boldsymbol{\beta}}_y^{-k}(\lambda))^2.$$

4. This gives the cross-validation error

$$CV(\lambda) = \frac{1}{K}\sum_{1}^{K} e_k(\lambda).$$

5. For each value of the tuning parameter $\lambda \in [\lambda^{RT} - 100, \lambda^{RT} + 100]$, repeat steps 3-4 and choose the tuning parameter that minimizes the $CV(\lambda)$.

B.5 Foreign Supply: The Role of $CMA_{ik,t}$

This appendix discusses the results of estimating the growth effects of foreign supply shocks, as captured by the external Consumer Market Access terms $CMA_{ik,t}$. Straightforward steps lead to an extension of equation (2.7) to include foreign supply shocks.[28]

$$d \ln y \approx \sum_k \delta_k^x \cdot \left[\lambda_k^x d \ln FMA_k\right] + \sum_k \delta_k^m \cdot \left[\lambda_k^m d \ln CMA_k\right]. \tag{B.2}$$

One can estimate the elasticities of real income with respect to foreign supply shocks by following similar steps as we do in estimating the impact of foreign demand. From (2.5), the $CMA_{nk,t}$ are expressed as follows:

$$CMA_{nk,t} = \sum_{i \neq n} c_{ik,t}^{1-\sigma_k} \cdot \tau_{ink,t}^{1-\sigma_k},$$

where n is importer. The (log) $c_{ik,t}^{1-\sigma_k}$ is recovered based on exporter fixed effects. After estimating the gravity specification (4.2), the foreign supply shock can be constructed as:

$$CMA_{nk,t} = \sum_{i \neq n} \kappa_{ik,t}^x \cdot Distance_{in}^{\zeta_{kt}} \cdot \exp\left(\xi_{kt} \cdot Contig_{in}\right), \tag{B.3}$$

[28] External consumer market access enters into the welfare expression (2.6) implicitly through the sectoral price indices $P_k \equiv (z_{H,k} c_{H,k}^{1-\sigma_k} + CMA_k)^{\frac{1}{1-\sigma_k}}$.

that are then aggregated into clusters exactly like foreign demand shocks.

Figure A9 reports the results of estimating the impact of foreign supply shocks on income. The left panel presents the OLS results, the right panel the double-LASSO results. Overall, the foreign supply shocks have both much larger magnitudes and standard errors. The latter feature makes it challenging to draw sharp conclusions about the impact of foreign supply shocks on income. The one significant coefficient (on the Consumption goods cluster) does not survive reasonable robustness checks. In practice, the variation in the FMA terms is an order of magnitude larger than the variation in CMA terms. This is sensible from an economic standpoint: examination of the functional forms for FMA and CMA in equations (4.3) and (B.3) reveals that foreign demand shocks are determined by both changes in foreign prices/costs as well as changes in the overall foreign expenditure. On the other hand, foreign supply shocks are driven purely by changes in foreign costs. As a result, the FMA terms have much greater variation in the data. Statistically, it is thus not surprising that a regressor with a smaller standard deviation has a higher point estimate. The large standard errors, however, imply a relative lack of confidence in those estimates.

Figure A10 reports the main results of the paper for foreign demand shocks when controlling for the vector of $d \ln CMA$'s. Note that throughout, all double-LASSO estimation admits foreign supply shocks as potential controls. In this robustness check, we make them "protected" controls, meaning that they are included as controls regardless of whether they are selected by the procedure. The main findings of the paper are robust to this exercise.

B.6 Developed vs. Developing Countries

Our main specification pools all countries and time periods together and clusters on the industry dimension alone. It is also interesting to consider clustering along the country dimension, i.e. whether the impact of foreign shocks exhibits heterogeneity across different groups of countries.[29] One of the more intriguing possibilities is that rich and poor countries systematically differ in the income impact of foreign shocks to different sectors. To investigate this hypothesis, we split the sample into two groups based on the World Bank's 2016 country classification by income. Developing countries are those assigned by the World Bank to "low income" and "lower middle income" categories, and the developed countries the remaining group. According to this classification, 70 countries belong to the developed group, and 57 to the developing group (Appendix Table A4). We then estimate elasticities of real income with respect to foreign shocks for the two country groups separately.

Figure A8 reports the results of the baseline specifications for the developed and developing groups. For both groups, the coefficients on demand shocks in complex intermediates are positive and precisely estimated, although the magnitude is larger for the developed country group. On the other hand, the capital goods coefficient behaves very differently in the two samples: it is slightly smaller than the baseline coefficient in the developed country sample, but much larger in the developing

[29] This heterogeneity could come from a combination of differences in underlying parameter values and in the point of approximation.

country sample. The standard error on the capital goods coefficient is actually smaller for the developed country sample that the full sample case, while it is larger for the developing country sample. These results suggest that the relatively large standard errors and sensitivity to classification errors observed for the capital coefficient in the full sample may be in part due to the heterogeneity across the country subsamples.

C. Additional Appendix Tables and Figures

Table A4: Country List

3-letter code	Country	Developed	3-letter code	Country	Developed
ALB	Albania	x	LBR	Liberia	
DZA	Algeria	x	LTU	Lithuania	x
AGO	Angola	x	MKD	Macedonia, FYR	x
ARG	Argentina	x	MDG	Madagascar	
ARM	Armenia		MWI	Malawi	
AUS	Australia	x	MYS	Malaysia	x
AUT	Austria	x	MLI	Mali	
AZE	Azerbaijan	x	MRT	Mauritania	
BGD	Bangladesh		MEX	Mexico	x
BLR	Belarus	x	MDA	Moldova	
BEL	Belgium	x	MNG	Mongolia	
BEN	Benin		MAR	Morocco	
BOL	Bolivia		MOZ	Mozambique	
BWA	Botswana	x	NAM	Namibia	x
BRA	Brazil	x	NPL	Nepal	
BGR	Bulgaria	x	NLD	Netherlands	x
BFA	Burkina Faso		NZL	New Zealand	x
BDI	Burundi		NIC	Nicaragua	
KHM	Cambodia		NER	Niger	
CMR	Cameroon		NGA	Nigeria	
CAN	Canada	x	NOR	Norway	x
CAF	Central African Republic		OMN	Oman	x
TCD	Chad		PAK	Pakistan	
CHL	Chile	x	PRY	Paraguay	x
CHN	China	x	PER	Peru	x
COL	Colombia	x	PHL	Philippines	
COG	Congo, Rep.		POL	Poland	x
CRI	Costa Rica	x	PRT	Portugal	x
CIV	Cote d'Ivoire		QAT	Qatar	x
HRV	Croatia	x	ROM	Romania	x
CZE	Czech Republic	x	RUS	Russian Federation	x
DNK	Denmark	x	RWA	Rwanda	
DOM	Dominican Republic	x	SAU	Saudi Arabia	x
ECU	Ecuador	x	SEN	Senegal	
EGY	Egypt, Arab Rep.		SLE	Sierra Leone	
SLV	El Salvador		SVK	Slovak Republic	x
ETH	Ethiopia		SVN	Slovenia	x
FIN	Finland	x	ZAF	South Africa	x
FRA	France	x	ESP	Spain	x
GEO	Georgia	x	LKA	Sri Lanka	
DEU	Germany	x	SDN	Sudan	
GHA	Ghana		SWE	Sweden	x
GRC	Greece	x	CHE	Switzerland	x
GTM	Guatemala		SYR	Syrian Arab Republic	
GIN	Guinea		TWN	Taiwan Province of China	x
HTI	Haiti		TJK	Tajikistan	
HND	Honduras		TZA	Tanzania	
HUN	Hungary	x	THA	Thailand	x
IND	India		TGO	Togo	
IDN	Indonesia		TUN	Tunisia	
IRQ	Iraq	x	TUR	Turkey	x
IRL	Ireland	x	TKM	Turkmenistan	x
ISR	Israel	x	UGA	Uganda	
ITA	Italy	x	UKR	Ukraine	
JPN	Japan	x	ARE	United Arab Emirates	x
JOR	Jordan	x	GBR	United Kingdom	x
KAZ	Kazakhstan	x	USA	United States	x
KEN	Kenya		URY	Uruguay	x
KOR	Korea, Rep.	x	UZB	Uzbekistan	
KWT	Kuwait	x	VEN	Venezuela, RB	x
KGZ	Kyrgyz Republic		VNM	Vietnam	
LAO	Lao PDR		YEM	Yemen, Rep.	
LBN	Lebanon	x	ZMB	Zambia	
LSO	Lesotho				

Notes: The "x"s indicate that the country is in the developed subsample.

Table A5: Summary Statistics of Clusters: Grouping the Manufacturing Industries to 5 Clusters

	cluster						
	1	2	3	4	5	Mean	Std. Dev.
Inv. Share	0.00	0.05	0.57	0.03	0.16	0.13	0.22
Int. Using	0.76	0.62	0.67	0.66	0.57	0.66	0.16
Int. Sales	0.85	0.71	0.26	0.31	0.52	0.57	0.31
Conc. Ratio	0.48	0.23	0.35	0.59	0.41	0.40	0.21
Sk. Share	0.33	0.23	0.30	0.32	0.54	0.32	0.13
Cap. Int.	0.69	0.55	0.54	0.69	0.55	0.61	0.10
Con. Int.	0.25	0.52	0.71	0.49	0.74	0.51	0.22
Num of ind.	54	70	36	44	29		
Trade share	0.31	0.20	0.15	0.07	0.20		
Label	Raw Materials Processing	Complex Intermediates	Capital Goods	Consumer Goods	Skill Intensive		
Abbreviation	RAW	INT	CAP	CONS	SI		

Notes: This table reports the summary statistics of the sectoral characteristics among the sectors selected into each cluster, when the number of clusters is 5. The last two columns report the mean and standard deviations of those characteristics among all manufacturing sectors. The row "Num. of ind" reports the number of sectors in each cluster, and "Trade share" reports the fraction of world trade accounted for by sectors in that cluster. The bottom panel lists the intuitive labels of the clusters, as well as 3-letter abbreviations. Both are heuristic and assigned by the authors.

Table A6: Control Variables Selected in the Double-Selection LASSO Procedure: Baseline Estimation

Admissible Controls	Controls Selected: Baseline	Controls Selected: Developed Countries	Controls Selected: Developing Countries
$\lambda^x_{ik,t}$	$\lambda^x_{i104,t}$ $\lambda^x_{i176,t}$	$\lambda^x_{i178,t}$	$\lambda^x_{i65,t}$
$\lambda^m_{ik,t}$			$\lambda^m_{i263,t}$
$\lambda^x_{ig,t}$			
$\lambda^m_{ig,t}$			
$\lambda^x_{ik,t} \cdot \ln FMA_{ik,t}$	$\lambda^x_{i114,t} \cdot \ln FMA_{i114,t}$ $\lambda^x_{i143,t} \cdot \ln FMA_{i143,t}$	$\lambda^x_{i92,t} \cdot \ln FMA_{i92,t}$	$\lambda^x_{i243,t} \cdot \ln FMA_{i243,t}$
$\lambda^m_{ik,t} \cdot \ln CMA_{ik,t}$	$\lambda^m_{i166,t} \cdot \ln CMA_{i166,t}$ $\lambda^m_{i176,t} \cdot \ln CMA_{i176,t}$	$\lambda^m_{i205,t} \cdot \ln CMA_{i205,t}$	
$\sum_{k \in g} \lambda^x_{ik,t} \cdot \ln FMA_{ik,t}$			
$\sum_{k \in g} \lambda^m_{ik,t} \cdot \ln CMA_{ik,t}$			
$\sum_{k \in g} \lambda^m_{ik,t} \cdot d \ln CMA_{ik,t}$			
$\ln population_{i,t}$			
$\ln k_{i,t}$			
$\ln y_{i,t}$	included	included	included
$(\ln y_{i,t})^2$			
Time effects	included	included	included
Number of Controls Selected	6	3	3
Estimates Figures	Figure 1	Figure A8	Figure A8

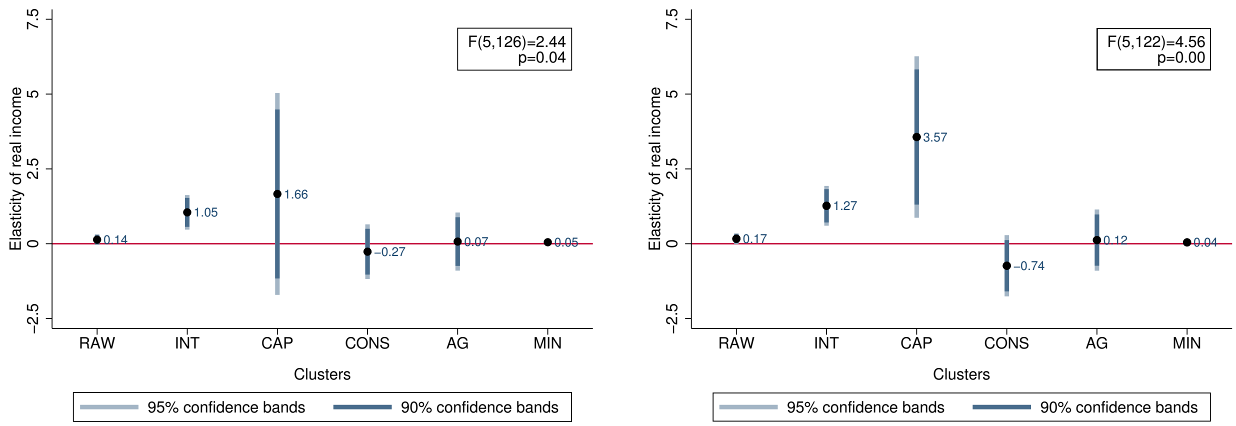
Notes: All specifications control for initial GDP per capita. Industries in our sample are relabeled by number from 1 to 268 for coding purposes, i.e. $k = 1, 2, \ldots, 268$. The numbers in the subscripts refer to the corresponding industries.

Table A7: Control Variables Selected in the Double-Selection LASSO Procedure: Robustness Checks

Admissible Controls	Controls Selected: Dropping Large Trading Partners	Controls Selected: Dropping Neighboring Countries
$\lambda^x_{ik,t}$	$\lambda^x_{i94,t}$ $\lambda^x_{i104,t}$ $\lambda^x_{i111,t}$ $\lambda^x_{i114,t}$ $\lambda^x_{i158,t}$ $\lambda^x_{i176,t}$	$\lambda^x_{i111,t}$ $\lambda^x_{i176,t}$ $\lambda^x_{i182,t}$
$\lambda^m_{ik,t}$		
$\lambda^x_{ig,t}$		
$\lambda^m_{ig,t}$		
$\lambda^x_{ik,t} \cdot \ln FMA_{ik,t}$	$\lambda^x_{i114,t} \cdot \ln FMA_{i114,t}$ $\lambda^x_{i152,t} \cdot \ln FMA_{i152,t}$ $\lambda^x_{i175,t} \cdot \ln FMA_{i175,t}$ $\lambda^x_{i186,t} \cdot \ln FMA_{i186,t}$ $\lambda^x_{i190,t} \cdot \ln FMA_{i190,t}$	$\lambda^x_{i143,t} \cdot \ln FMA_{i143,t}$ $\lambda^x_{i152,t} \cdot \ln FMA_{i152,t}$ $\lambda^x_{i186,t} \cdot \ln FMA_{i186,t}$ $\lambda^x_{i202,t} \cdot \ln FMA_{ik202,t}$ $\lambda^x_{i203,t} \cdot \ln FMA_{i203,t}$
$\lambda^m_{ik,t} \cdot \ln CMA_{ik,t}$	$\lambda^m_{i166,t} \cdot \ln CMA_{i166,t}$ $\lambda^m_{i229,t} \cdot \ln CMA_{i229,t}$	$\lambda^m_{i166,t} \cdot \ln CMA_{i166,t}$
$\sum_{k\in g} \lambda^x_{ik,t} \cdot \ln FMA_{ik,t}$		
$\sum_{k\in g} \lambda^m_{ik,t} \cdot \ln CMA_{ik,t}$		
$\sum_{k\in g} \lambda^m_{ik,t} \cdot d\ln CMA_{ik,t}$		
$\ln population_{i,t}$		
$\ln k_{i,t}$		
$\ln y_{i,t}$	included	included
$(\ln y_{i,t})^2$		
Time effects	included	included
Number of Controls Selected	13	9
Estimates Figures	Figure A6	Figure A6

Notes: All specifications control for initial GDP per capita. Industries in our sample are relabeled by number from 1 to 268 for coding purposes, i.e. $k = 1, 2, \ldots, 268$. The numbers in the subscripts refer to the corresponding industries. The set of neighboring countries is the union of contiguous countries and the top 5 closest countries by geographic distance.

Figure A6: Cluster-Specific Coefficients and Confidence Intervals, Robustness to Small Country Assumption

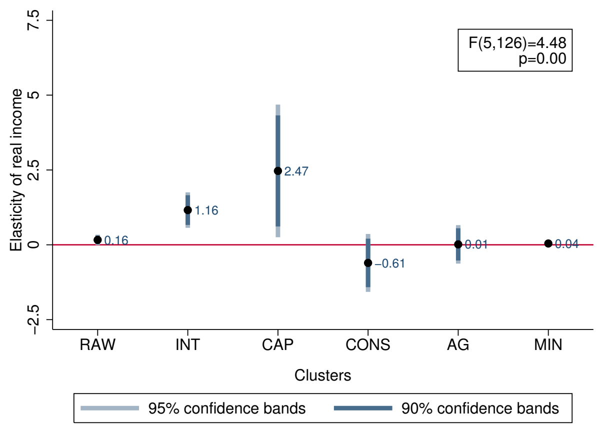


(a) Dropping Large Trading Partners

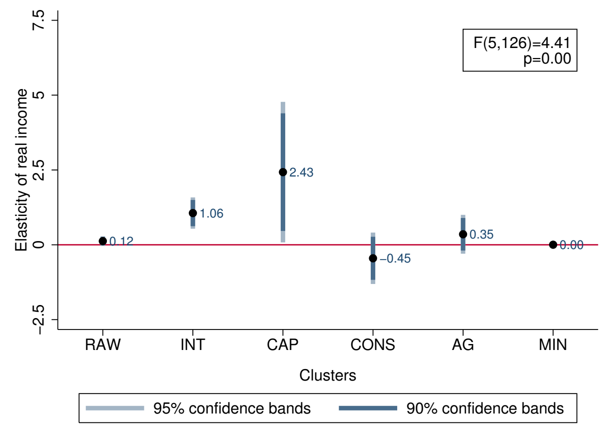
(b) Dropping Large Exporters

Notes: This figure reports estimates of the δ^x_g coefficients in equation (3.6) via post double-LASSO. All specifications control for (i) time effects and (ii) log initial GDP per capita. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display the results of an F-test for equality of the coefficients in each plot. In panel (a), the construction of the FMA terms omits foreign markets for which country i is a large trading partner. 13 control variables are selected in the double-selection step. Panel (b) drops from the estimation sample countries that represent the largest shares of world exports in any cluster. 3 control variables are selected in the double-selection step.

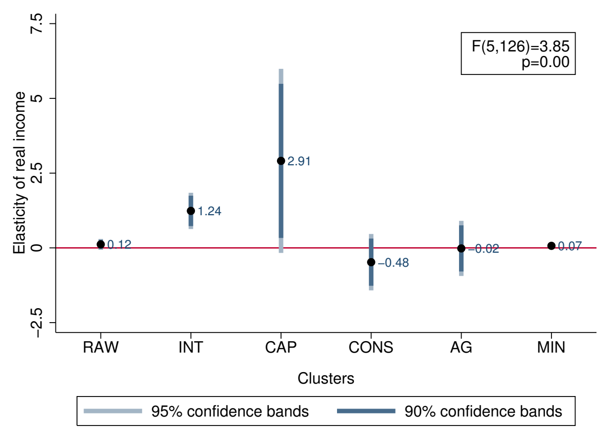
Figure A7: Cluster-Specific Coefficients and Confidence Intervals, Robustness to Spatial Correlation



(a) Dropping Neighboring Countries



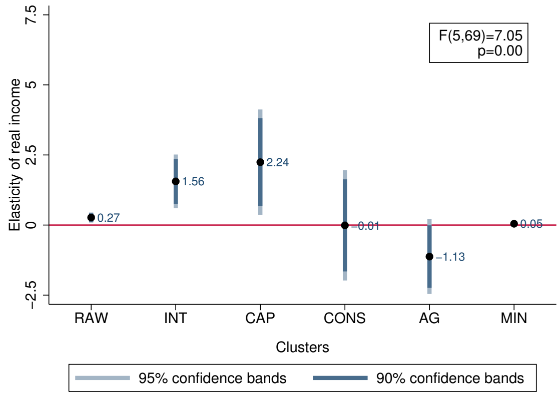
(b) Controlling for Neighboring Countries' TFP



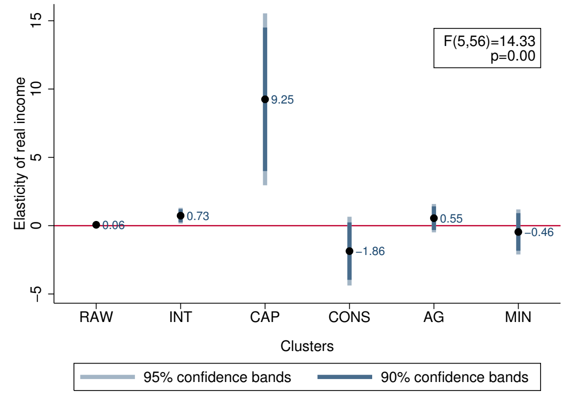
(c) Controlling for Unweighted $d \ln FMA_{igt}$

Notes: This figure reports estimates of the δ_g^x coefficients in equation (3.6) via post double-LASSO. All specifications control for (i) time effects and (ii) log initial GDP per capita. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display the results of an F-test for equality of the coefficients in each plot. In Panel (a) construction of the FMA terms omits neighboring countries. 9 control variables are selected in the double-selection step. Panel (b) controls for neighboring countries' average TFP growth in the post-LASSO OLS. The set of neighboring countries is the union of contiguous countries and the top 5 closest countries by geographic distance. Panel (c) controls for the non-export-share weighted $d \ln FMA_{igt}$ in the post-LASSO OLS.

Figure A8: Developed vs. Developing Countries



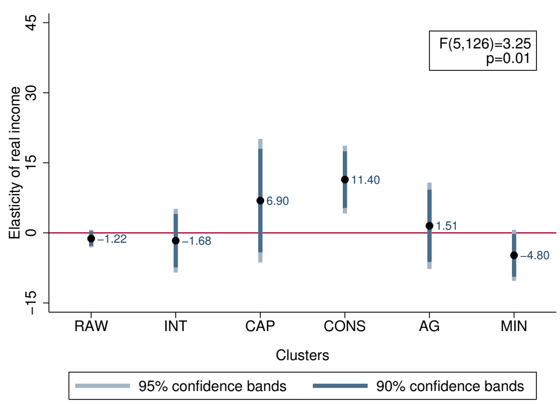
(a) Developed Countries



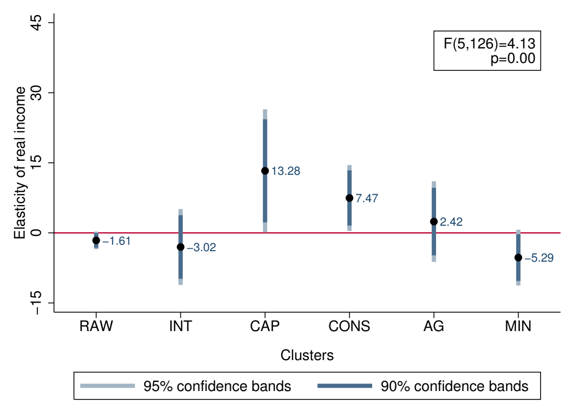
(b) Developing Countries

Notes: This figure reports estimates of the δ_g^x coefficients in equation (3.6) via the Post Double-LASSO. All specifications control for (i) time effects and (ii) log initial GDP per capita. The left panel displays the results for the sample of developed countries. 3 control variables are selected in the double-selection step. The right panel displays the results for developing countries. 3 control variables are selected in the double-selection step. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display the results of an F-test for equality of the coefficients in each plot.

Figure A9: Foreign Supply: Cluster-Specific Coefficients and Confidence Intervals for CMA



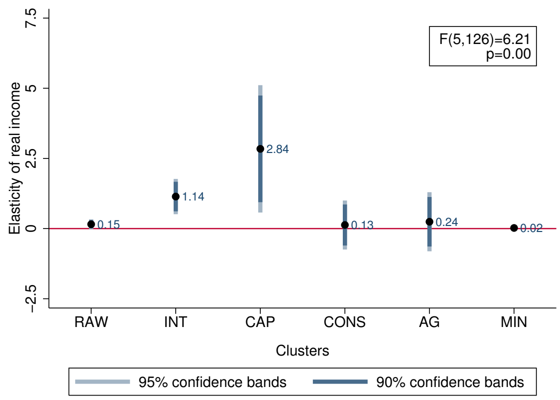
(a) OLS Estimates



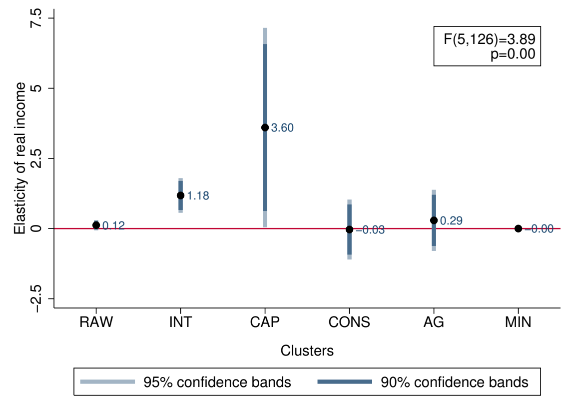
(b) LASSO Estimates

Notes: This figure reports the coefficients in estimating Equation (3.6), for the foreign supply shocks (CMA). All specifications control for (i) time effects and (ii) log initial GDP per capita. The left panel displays the baseline OLS estimates. The right panel displays the post double-LASSO estimates. 16 control variables are selected in the double-selection step. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display the results of an F-test for equality of the coefficients in each plot.

Figure A10: Foreign Demand: Cluster-Specific Coefficients and Confidence Intervals, Controlling for Foreign Supply



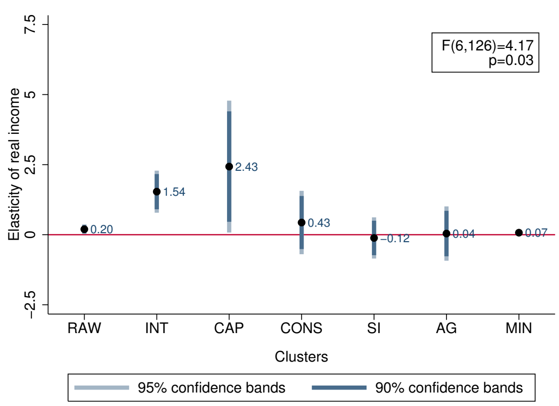
(a) OLS Estimates



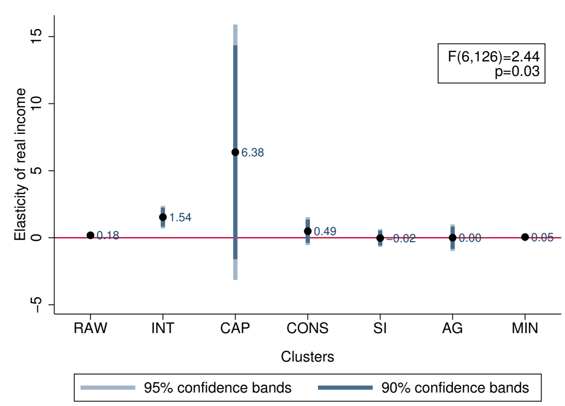
(b) LASSO Estimates

Notes: This figure reports the coefficients in estimating Equation (3.6), for the foreign demand shocks (FMA). All specifications control for (i) time effects and (ii) log initial GDP per capita, and (iii) foreign supply shocks (CMA). The left panel displays the baseline OLS estimates. The right panel displays the post double-LASSO estimates. 6 control variables are selected in the double-selection step. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display the results of an F-test for equality of the coefficients in each plot.

Figure A11: Cluster-Specific Coefficients and Confidence Intervals When Grouping the Manufacturing Industries to 5 Clusters



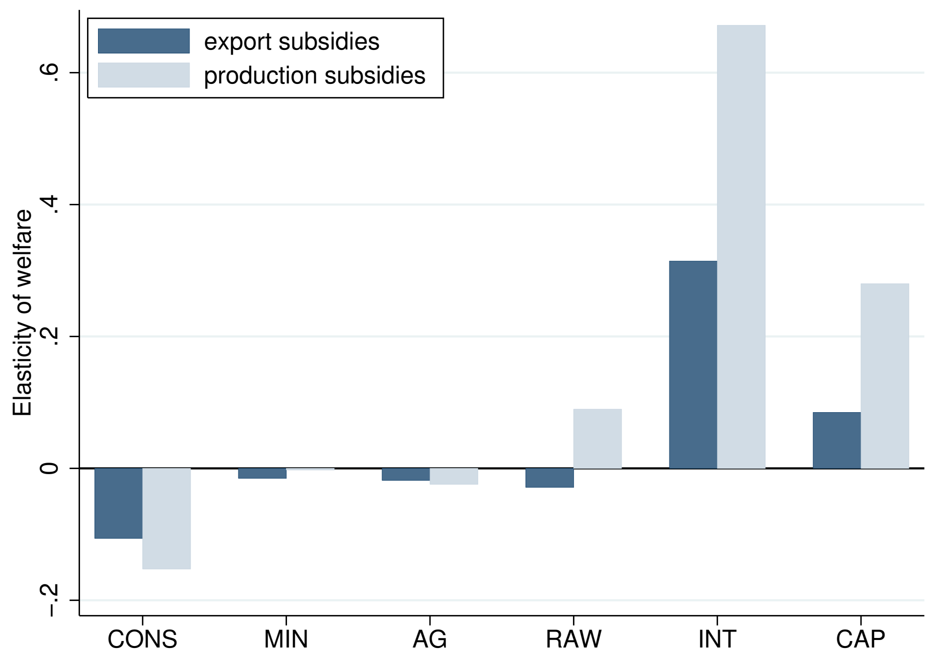
(a) OLS Estimates



(b) LASSO Estimates

Notes: This figure reports estimates of the δ_g^x coefficients in equation (3.6) via post double-LASSO, when grouping the manufacturing industries to 5 clusters. All specifications control for (i) time effects and (ii) log initial GDP per capita. The left panel displays the baseline OLS estimates. The right panel displays the post double-LASSO estimates. 9 control variables are selected in the double-selection step. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display the results of an F-test for equality of the coefficients in each plot.

Figure A12: First Order Gains from Export and Production Subsidies, Heterogeneous σ_k



Notes: Gains from policies computed using the formulas (2.9) and (A.17). The elasticities σ_k are cluster-specific, and sourced from Caliendo and Parro (2015), concorded to our 6 clusters.

PUBLICATIONS

Specialization, Market Access and Real Income

Working Paper No. WP/2024/051