NBER WORKING PAPER SERIES

SPECIALIZATION, MARKET ACCESS AND REAL INCOME

Dominick Bartelme
Ting Lan
Andrei A. Levchenko

Working Paper 28274
http://www.nber.org/papers/w28274

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020, Revised March 2021

We 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 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. The views expressed in this paper are those of the authors and should not be attributed to the International Monetary Fund, its Executive Board, its management, or the National Bureau of Economic Research.

Specialization, Market Access and Real Income
Dominick Bartelme, Ting Lan, and Andrei A. Levchenko
NBER Working Paper No. 28274
December 2020, Revised March 2021
JEL No. F43,F62

ABSTRACT

This paper estimates the impact of external demand shocks on real income. Our empirical strategy is based on a first order approximation to a wide class of small open economy models that feature sector-level gravity in trade flows. The framework allows us to measure foreign shocks and characterize their impact on income 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. We find clear evidence of heterogeneity in the income responses to different foreign shocks. Foreign demand shocks in complex intermediate and capital goods have large positive impacts on real income, whereas impacts in other sectors are negligible. The estimates imply that the pattern of sectoral specialization plays a quantitatively large role in how foreign shocks affect real income, while geographic position plays a smaller role. Finally, a calibrated multi-sector production and trade model can rationalize both the average and the heterogeneity in real income elasticities to foreign shocks under reasonable values of structural parameters.

Dominick Bartelme
Department of Economics
University of Michigan
238 Lorch Hall
611 Tappan Avenue
Ann Arbor, MI 48109
dbartelm@umich.edu

Ting Lan
700 19th St NW
Washington, DC 20431
tlan@imf.org

Andrei A. Levchenko
Department of Economics
University of Michigan
611 Tappan Street
Ann Arbor, MI 48109
and CEPR
and also NBER
alev@umich.edu

1. Introduction

The notion that the wealth of nations is shaped in part by international trade opportunities dates back to the origins of international economics. The size and sectoral composition of foreign demand varies substantially across countries due to trade costs and the uneven geography of development. How important is the variation in external demand for understanding the cross-country income distribution? A rich theoretical literature studies how foreign demand shocks affect real income and welfare, often emphasizing the role of a country's comparative advantage. An enduring theme is that the real income impacts of foreign shocks are heterogeneous across industries, and their signs and magnitudes depend crucially on the strength of the various mechanisms at play.[1] The goal of this paper is to empirically estimate the effects of external demand shocks in different industries on real income.

Empirical work on this question faces a number of challenges. There are many sectors and theories, but relatively few real income observations in the data. Econometric issues of endogeneity and omitted variable bias loom large. Faced with these challenges, the existing literature has coalesced around three basic approaches. One abstracts from sectoral heterogeneity altogether and focuses on the relationship between real income and the overall size of the external market, as determined by geography (e.g. Frankel and Romer, 1999; Redding and Venables, 2004). Another examines whether certain features of comparative advantage are associated with growth (e.g. Prebisch, 1959; Humphreys et al., eds, 2007). This approach abstracts from cross-country variation in external demand, and lacks a common theoretical foundation. The third calibrates fully specified general equilibrium models and conducts counterfactuals (e.g. Whalley, 1985; Eaton and Kortum, 2002). These methods deliver precise and interpretable answers, but depend heavily on the assumed theoretical framework and structural parameter values.

This paper develops a unified approach to quantifying the impact of foreign shocks in different sectors that strikes a balance between the clarity and rigor of the structural tradition and more model-robust statistical methods that "let the data speak." We begin by analyzing a class of small open economy models with many sectors that satisfy four key assumptions: i) sector-level gravity in bilateral trade flows, ii) a homothetic upper-tier utility aggregator, iii) competitive goods and factor markets, and iv) a unique and smooth equilibrium mapping from the primitives to the endogenous outcomes. The production side of the economy is quite general, allowing for any number of factors, intermediate goods linkages, and external effects within and across sectors. This class contains small open economy versions of most of the quantitative trade models in the literature as special cases, including isomorphisms with settings featuring monopolistic competition.

[1] Handbook chapters by Costinot and Rodríguez-Clare (2014) and Ventura (2005) review and quantify the impact of changes in openness on income levels and growth rates, respectively, under various assumptions on the structure of the economy. Rodríguez and Rodrik (2001) and Harrison and Rodríguez-Clare (2010) provide critical reviews of the empirical work on openness and income. Lederman and Maloney (2012) summarize the theoretical and empirical literature on trade patterns and income.

The framework delivers natural measures of sector-level foreign demand shocks, which we label external firm market access. These variables contain all relevant information about foreign demand for a country's exports, and are easily estimated from the trade data using standard techniques. We employ a first order approximation to express a log change in a country's real income in terms of export share-weighted averages of external firm market access, along with domestic demand and supply shocks. The elasticities of real income with respect to external firm market access reflect how different foreign demand shocks generate different general equilibrium income impacts, thus providing a direct answer to the question posed at the outset of the paper.[2]

We implement our approach on UN COMTRADE trade data and decadal real income changes from the Penn World Table 9.0 over 1965-2015, with a sample of 127 countries and 268 sectors spanning manufacturing, agriculture, and mining. Estimation of the model-derived equation must confront two primary challenges.

The first is that there are hundreds of traded sectors and thus potentially hundreds of elasticities of income to foreign shocks that can be estimated. This is clearly not feasible given the relatively small sample of available 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. Theory predicts that a sector's income impact depends on its features, such as position in the production network, factor intensities, and so on. Clustering industries with similar characteristics thus has the dual benefits of minimizing within-cluster heterogeneity in the income impacts of foreign shocks, and identifying the key sectoral properties that explain the heterogeneous impacts. We use the k-means clustering algorithm (MacQueen et al., 1967) along with 7 sectoral characteristics measured from US data to group 233 manufacturing industries into 4 clusters. It turns out that this procedure results in clusters with features that are easy to verbalize: 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 for a total of 6 clusters and therefore 6 cluster-level foreign demand shocks. We then estimate the much smaller number of cluster-level income elasticities.

The second challenge is the common one in cross-country growth regressions: omitted variables and endogeneity. We first provide formal conditions under which the average within-cluster income elasticities are identified by an OLS regression that fully conditions on the initial equilibrium 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

[2]As will become clear below, a country will also be affected by foreign supply shocks, which are reflected in its import prices. It turns out that there is much less variation across countries and time in the identifiable foreign supply shocks, and their estimated income impacts are quite noisy and rarely significant. For these reasons, the paper focuses on 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.

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.

Our main finding is that the average impact of foreign shocks on real income differs significantly across clusters. Foreign demand shocks to capital goods and complex intermediate sectors have the largest income impacts, with the capital goods elasticity being somewhat less precisely estimated. Foreign demand shocks in all other sectors have small and insignificant income impacts.

We subject our specification to robustness checks along a number of dimensions including the number of clusters, the tuning parameter used for selecting controls, measurement error in the cluster characteristics, and dropping important trading partners. The most robust result is that demand shocks in complex intermediate goods have high income elasticities and non-intermediate, non-capital goods sectors have small elasticities. The capital goods cluster always has the highest income elasticity point estimate, but with relatively large standard errors. Interestingly, when we split the sample into developed and developing countries, we obtain more precise estimates as well as a much higher capital goods elasticity for developing countries across all specifications. This suggests that the large standard error in the full sample may be partly due to differences in the income impact of foreign demand shocks for capital goods in different groups of countries. While the finding that the capital goods elasticity is higher for poor countries is intriguing, its practical importance is limited by the low shares of capital goods in developing countries' export baskets.

We examine the quantitative implications of our estimates through the lens of both data and theory. First, given our estimated cluster-level elasticities, the real income impacts of foreign shocks in individual countries are determined by the size and pattern of external demand ("geography") interacted with the initial sectoral specialization. To isolate the role of the specialization pattern, we hold geography constant and compute the total elasticity of income with respect to a uniform foreign demand shock for each country in our sample. There is substantial cross-country heterogeneity in the impacts, ranging from 0 in many of the poorest countries to 0.5 in upper-middle income countries such as Hungary, Slovakia, Malaysia, and Taiwan. We illustrate the role of geography by holding sectoral specialization constant and subjecting each country to the foreign shocks experienced by different countries in the same time period. We find that geography plays a modest but noticeable role in determining the growth experiences of different countries. For example, East Asian countries benefited by about half a percentage point of growth over the 2005-2015 decade (relative to the median country) from the rapid growth of surrounding countries, while Western European countries lost roughly 1 percentage point of growth over the same decade due to slow overall growth in the region.

Second, we ask whether our empirical estimates can be rationalized by theory. To that end, we set up a quantitative trade model of a small open economy with intermediate input linkages, endogenous capital accumulation and industry-level scale effects that are external to the firm (Bartelme et al., 2019;

Kucheryavyy et al., 2020). We calibrate the model using standard data on intermediate, final, and trade shares. We show that a parsimonious parameterization with only two structural elasticities – substitution and scale – can successfully match both the average level of estimated coefficients and their variation across clusters. Importantly, the model is quantitatively successful under a fairly broad range of these structural parameters, and for values that are reasonable in light of existing estimates. The model matches the heterogeneity in estimated coefficients purely through internal propagation of foreign demand shocks within the home economy, rather than different structural parameters across clusters. We show that input linkages in both intermediate and capital goods as well as the presence of substantial scale economies are quantitatively important for matching our econometric estimates.

We stress that this quantification is a "proof-of-concept" exercise, rather than a strong stand on the precise economic mechanisms behind the empirical estimates. There might be many theoretical models, and potentially infinitely many parameter combinations within each model, that could match the income elasticities in the data. The objective of this exercise is to highlight a set of economic mechanisms that can be quantitatively successful at matching the econometric estimates.

Related literature. Our paper contributes to the literature on trade and income, 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 literature on export patterns and 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; Hausmann et al., 2014), or skill intensity (Atkin, 2016; Blanchard and Olney, 2017).[3] We make two contributions to the empirics of trade patterns and income. 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.

In a sense, all of international trade theory is about the relationship between openness and income. Many mechanisms have been proposed for how the pattern of sectoral specialization can affect the income level, 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; Melitz, 2005) externalities, and to political economy (Tornell and Lane, 1999; Levchenko, 2013; Berman et al., 2017; Dippel et al.,

[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).

2020), to name a few. The wealth of potential theoretical mechanisms motivates the more data-driven approach in our paper.

Our use of theory 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 et al., 2012; Caliendo and Parro, 2015; Hsieh and Ossa, 2016; Levchenko and Zhang, 2016), as well as recent work on trade counterfactuals that apply across families of models (e.g. Arkolakis et al., 2012; Adão et al., 2017; Bartelme, 2018; Baqaee and Farhi, 2019; Allen et al., 2020; Kleinman et al., 2020). Our approach is rooted in theory but is centered on econometric estimation of the general equilibrium effects of trade shocks. Section 2.1 discusses in detail the relationship between our approach and the quantitative trade literature.

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

Economic environment. 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 will face downward sloping demand for its products in international markets (as in Armington, 1969).

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

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

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, 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,

$$Q_{n,k} = \left(z_{n,k}^{\frac{1}{\sigma_k}} \cdot q_{nn,k}^{\frac{\sigma_k - 1}{\sigma_k}} + \sum_{i \neq n} q_{in,k}^{\frac{\sigma_k - 1}{\sigma_k}} \right)^{\frac{\sigma_k}{\sigma_k - 1}}, \tag{2.1}$$

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 (2.1) implies that foreign demand for Home's exports in sector k takes the form

$$p_{Hn,k} \cdot q_{Hn,k} = (c_{H,k} \cdot \tau_{Hn,k})^{1-\sigma_k} \cdot \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}}, \tag{2.2}$$

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 (2.1) in country n.

All factor income in Home accrues to a representative consumer, who has homothetic preferences over sectoral quantity bundles $Q_{H,k}$.[4]

Foreign shocks. We now define the key object underlying our analysis. By summing export revenues across foreign export destinations, we get total foreign revenues as a function of Home costs and *External Firm Market Access* (FMA),[5]

$$\sum_{n \neq H} p_{Hn,k} q_{Hn,k} = c_{H,k}^{1-\sigma_k} \cdot \underbrace{\sum_{n \neq H} \tau_{Hn,k}^{1-\sigma_k} \cdot \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}}}_{FMA_{H,k}}. \tag{2.3}$$

External firm market access has three key features. First, $FMA_{H,k}$ is an exogenous demand shifter for sector k 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 holding the price fixed. Second, any change in foreign demand affects the Home equilibrium only through its effects on $FMA_{H,k}$. Importantly, $FMA_{H,k}$ 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 described in Section 4.

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

[4]Homotheticity allows us to utilize price indices from national accounts in the empirical section.

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

imports across source countries:

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

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 $c_{n,k}\cdot\tau_{nH,k}$, modulated by the demand elasticity. 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.[6] 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 impact on income.

Competitive equilibrium. We define a static competitive equilibrium as 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. Under our assumptions, 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 (all derivations are in Appendix A). 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 unique and smooth mapping from the exogenous variables to equilibrium outcomes, at least locally. Without uniqueness the data would contain little or no information on how different foreign shocks systematically affect real income.[7] There are no general results available on the equilibrium properties of this class of models, and we do not pursue them here.[8] We assume a unique and smooth equilibrium mapping for the rest of the paper.

[6] 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 $FMA_{H,k}$.

[7] Our framework does allow small differences in either domestic fundamentals or foreign market access to have large impacts on long-run real income, a feature that many models with multiple equilibria are designed to capture.

[8] See Kucheryavyy et al. (2020) and Allen et al. (2020) for some results that apply to special cases of our framework.

2.1 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

$$y = \frac{\sum_{k \in K} \mu_k c_k^{1-\sigma_k} \cdot \left(z_k \frac{E_k}{P_k^{1-\sigma_k}} + FMA_k \right)}{\mathbb{P}}. \tag{2.5}$$

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

Equation (2.5) highlights the two ways that changes in FMA_k 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 approximately

$$d \ln y \approx \sum_k \delta_k^{ex} \cdot \left[\lambda_k^{ex} d \ln FMA_k \right], \tag{2.6}$$

where λ_k^{ex} is the initial share of total sales accounted for by exports in sector k.[9]

Discussion. The elasticities δ_k^{ex} encapsulate the general equilibrium response of real income to small exogenous changes in foreign demand in different industries. The main goal of this paper is to estimate them. As evident from (2.3), foreign demand shocks in this environment can come from a variety of sources, such as foreign taste or productivity shocks and changes in aggregate foreign expenditure, iceberg trade costs, or foreign trade policy. 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^{ex}. These elasticities are thus directly relevant to understanding how the evolution and cross-country variation in foreign demand have shaped the level and distribution of income across countries. They are also closely related to the welfare impact of changes in iceberg trade costs and foreign tariffs, and thus map to counterfactual experiments common in the quantitative trade

[9]The foreign sales shares λ_k^{ex} serve to weight the foreign demand shocks by the importance of foreign sales by sector. A valid alternative approximation absorbs these foreign sales shares into the coefficients δ_k^{ex}. The expression in the text is preferable for both interpretation and estimation: see the discussion below and in Section 3.

literature.[10]

From here, there are several ways to proceed. The dominant approach in the quantitative 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 $U(\cdot)$. Having done that, the model will feature well-defined propagation mechanisms, and can be disciplined with data. Quantification can take the form of estimating structural parameters using the partial equilibrium relationships implied by the model.[11] 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 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 δ_k^{ex} econometrically. As such, for us δ_k^{ex} 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

[10]The first order welfare impact of a change in iceberg export trade barriers or foreign tariffs is given by

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

[11]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.

estimates can be used as moments to be targeted by models, either for estimation or as out-of-sample validation. Section 6 implements one example of such an exercise, by evaluating the ability of a series of quantitative trade models to match our estimates under different parameter values.

While there are some simple theoretical environments in which the δ_k^{ex} do not vary by sector, most quantitative trade models imply that they do.[12] In an efficient economy the δ_k^{ex} are completely determined by the direct impact of foreign demand shocks on the terms of trade, which varies across sectors inversely with the trade elasticity. However, the *laissez-faire* competitive equilibrium of an Armington economy is not generally welfare-maximizing from the individual country perspective due to its unexploited international market power and domestic distortions (such as industry-level external economies of scale). The size and industry variation in the δ_k^{ex} are determined by the interaction of these factors with other sector characteristics, such as openness to trade, the position in the input-output network, and the final use of the industry (consumption vs. investment). 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^{ex}. The formula for δ_k^{ex} in the more full-fledged model makes it clear that the δ_k^{ex} generically differ across sectors, and depend on the full structure of the economy in ways that are increasingly complex and sensitive to assumptions. This very complexity provides one of the primary motivations for our more agnostic, data-driven approach to quantification.

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 (2.6) and the same interpretation of the market access elasticities δ_k^{ex}. In addition, while the model described above is static, equation (2.6) is also valid for small shocks in the steady state of a dynamic economy with some reproducible factors of production. One example of a dynamic model with such a steady state representation is described in Section 6.

3. Identification and Estimation

3.1 Identification

We now consider identification of the market access elasticities δ_k^{ex} in equation (2.6). We will view data as a collection of small open economies (indexed by i) observed over a number of steady states

[12] Note that the null hypothesis that the δ_k^{ex} are the same across sectors is testable. Section 5 reports statistical tests of this null, which is rejected by our estimates.

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 $\{Z_k\}$:

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

We proceed in several steps. First, we assume that the equilibrium mapping $G(\{Z_k\}) \rightarrow (\{x_k\})$ between all relevant exogenous variables and observable equilibrium outcomes is locally smoothly invertible; this implies that the observables "reveal" all relevant aspects of the initial state of the economy.[13] Using this assumption, we can rewrite (after redefining F) the log change in real income as

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

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)

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

where $\{\tilde{a}_k\}$ is a set of observable shocks (e.g. $\{\lambda_k^{im} d \ln CMA_k\}$). 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

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

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 $\hat{\delta}_k^{ex}$ as the causal effect of foreign demand shocks on real

[13] This is a typical property of quantitative trade models, that justifies the widespread use of "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.

income, we need the additional assumption that

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

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^{ex} 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, factor supply growth, and changes in domestic tastes 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.

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 (2.3) includes iceberg trade 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 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.[14]

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.

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 ($\lambda^{ex} d \ln FMA$), adding the unweighted $d \ln FMA$ as 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.

[14]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.

3.2 Estimation

Estimation of equation (3.4) by OLS is consistent for the δ_k^{ex} 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 $\{\delta_k^{ex}\}$, 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^{ex} 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

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

where

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

The cluster-level elasticities δ_g^{ex} 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^{ex} 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^{ex}, 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.[15] We deal with this problem by using the Post-Double-Selection estimator developed by Belloni et al.

[15] Note that simply aggregating the control variables into clusters in the same way as we do the foreign demand shocks would not, in general, lead to consistent estimates of the group-level elasticities.

(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.[16] 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. We 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, 268 sectors and 5 decades from 1965 to 2015, with a total of 548 10-year GDP growth rate observations.

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

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 (Solow, 1956) and human (Becker, 1975) capital, position in the input network (Jones, 2011), and contracting institutions (Acemoglu et al., 2005). 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 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.[17]

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.

Table 1 summarizes the characteristics of the 4 clusters. Since each cluster has some salient features that distinguish it from others, we name the clusters based on these key features. It is important to stress that the clustering procedure does not produce these cluster labels, nor does our identification strategy hinge upon them. We use the cluster names, shown in the last row of Table 1, purely for expositional purposes. Note also that there is no information contained in cluster numbers (1, 2, ...).

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

[17]This step is prudent because k-means clustering is not invariant to the scale used to measure the characteristics. If a particular characteristic takes on a broader range of values than the others, it will be given higher weight when assigning industries to clusters.

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.

sales share. Thus we label it "consumer goods." Table 2 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.[18] For a given sector k at time t, the gravity equation (2.2) 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

[18]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.

Table 2: 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.

sector-time-specific coefficients, leading to our empirical specification

$$\lambda_{ink,t} = \kappa^{ex}_{ik,t} \cdot \kappa^{im}_{nk,t} \cdot Distance^{\zeta_{kt}}_{in} \cdot \exp\left(\xi_{kt} \cdot Contig_{in}\right) \cdot \varepsilon_{ink,t}, \tag{4.2}$$

where $\kappa^{ex}_{ik,t}$ is the exporter fixed effect, $\kappa^{im}_{nk,t}$ 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:

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

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, $\kappa^{ex}_{ik,t}$ and $\kappa^{im}_{nk,t}$ 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. Appendix B.3 provides a detailed discussion.

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.

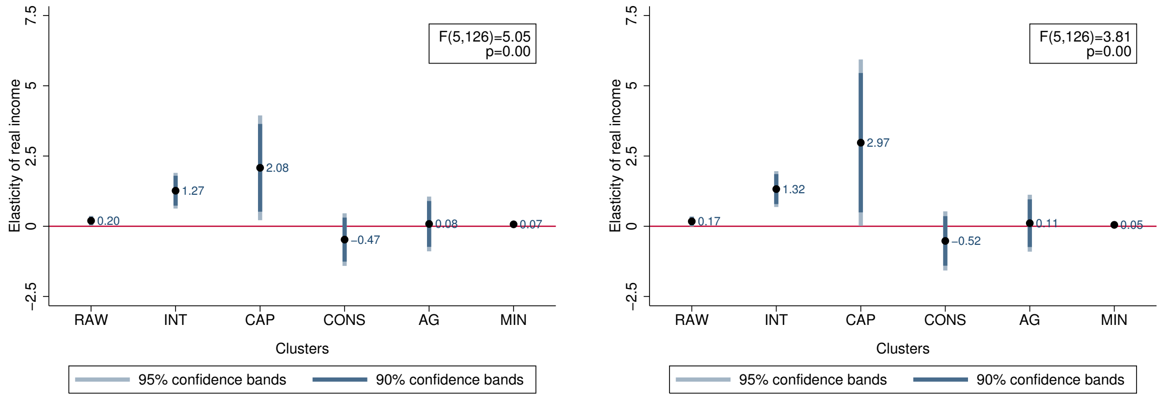
5. OLS regression of $d \ln y_{i,t}$ on $d \ln FMA_{ig,t}$ and selected controls to obtain estimates of δ_g^{ex}.

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

Figure 1: Cluster-Specific Coefficients and Confidence Intervals



(a) OLS Estimates

(b) LASSO Estimates

Notes: This figure reports estimates of the δ_g^{ex} 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.

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 A4 lists the potential and the selected controls in the Post-Double-Selection estimation.[19] Substantively the results are quite similar to the OLS specification, although some confidence intervals widen.

5.3 Robustness

Assignment of sectors to clusters. One concern with our approach is that clusters may be fragile due to some sectors being on the margins between clusters. If those sectors are particularly influential, then the results could be sensitive to the assignment of specific sectors to clusters. To assess the role

[19]We follow Belloni et al. (2014a) and choose the tuning parameter for the double-LASSO procedure through 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. We also check robustness to using a smaller tuning parameter for different specifications in Appendix Figures A5 and A9.

of marginal sectors in our results, we perform two exercises. First, we add a 5th manufacturing cluster. The results of re-clustering on 5 clusters are presented in Appendix Table A5. The basic characteristics of the original 4 clusters and the labels we attach to them remain the same. When given the opportunity to isolate a 5th cluster, the k-means procedure creates a cluster of skill-intensive industries.[20] The income regression results with 5 clusters are presented in Appendix Figure A6. The 5th cluster itself does not have a positive impact on income, indeed the coefficients are relatively precisely estimated zeros. The main findings regarding the income impacts of the other clusters are preserved.

In the second cluster robustness exercise, we assess the importance of sectors at the margins of the cluster classification. We add noise (standard deviation of 10% of the actual variability) to each characteristic of each sector, re-cluster sectors, and perform the full double-LASSO estimation using the new clusters. We repeat this procedure 1000 times. The goal is to see how the cluster-specific income-impact coefficients are affected by switching a small number of marginal sectors from one cluster to another.

Panel (b) of Appendix Figure A5 reports the results. The dots indicate our baseline coefficient estimates, whereas the dashed bars indicate the 95% range of outcomes across simulations (not confidence intervals). The figure reveals that many of the coefficient estimates are quite stable: the range of estimates across simulations for raw materials processing, agriculture, and mining clusters is very small. On the other hand, reclassification tends to boost the coefficients on consumption goods and to a lesser extent complex intermediates, at the expense of instability (in both directions) in coefficient on capital goods. These results indicate that our most robust findings are that foreign demand shocks in raw materials processing, consumption goods, agriculture, and mining have small income impacts and those in complex intermediates have significant and large income impacts, while the results for the the capital goods sector are less robust to this type of classification error.

Small country assumption. We next 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.

[20]The mean skilled labor share of this cluster, 0.54, is 21 percentage points higher than the skilled labor share of the second-most skill-intensive cluster.

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 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 A7. 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 A7 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 A8 reports 3 robustness checks. First, in panel (a) we omit contiguous countries from the calculation of the $d \ln FMA$ terms and re-estimate the model. Second, in panel (b) we control directly for the average TFP growth of 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. (Recall that the regressor of interest is the export-share weighted foreign demand shock, $\lambda_{ig,t} d \ln FMA_{ig,t}$.) 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.

Finally, an interesting question is whether the impact of foreign shocks has changed in important ways over time. Note that overall globalization trends will be picked up in our analysis, as the export/output shares $\lambda_{ig,t}$ are calculated for each decade, and will thus reflect the trends in trade openness. Note also that all of our specifications include time effects, that absorb the average effect of global phenomena (oil shocks, arrival of WTO, China's WTO accession, the Great Recession, etc) on trade and income. It could still be the case that the impact of foreign demand shocks changes over time conditional on export shares. Unfortunately, the relatively small size of our regression sample leaves us with insufficient power to explore this question statistically. 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.

5.4 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.[21] 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. We then estimate elasticities of real income with respect to foreign shocks for the two country groups separately.

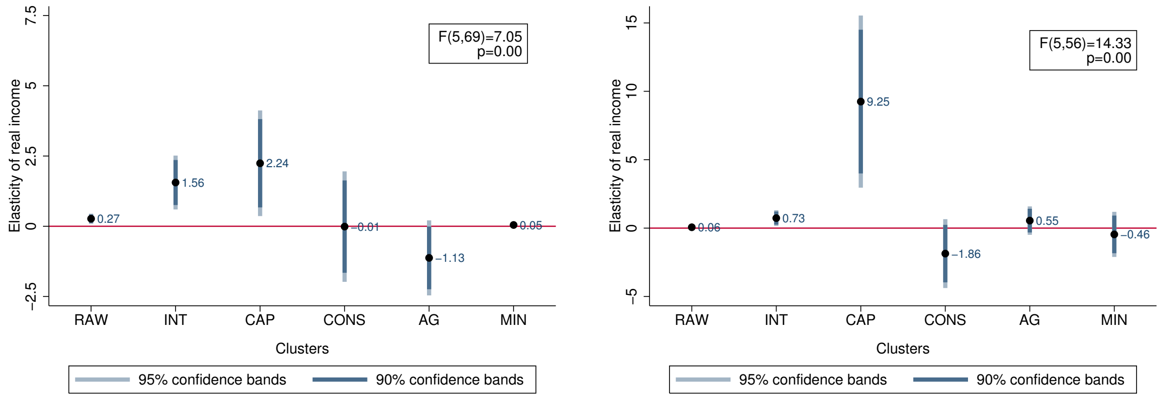
Figure 2 reports the results of the baseline specifications for the developed and developing groups. For both groups, the coefficients on demand shocks in complex intermediates are positive and precisely estimated, although the magnitude is larger for the developed country group. On the other hand, the 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 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. We repeat each of the robustness checks described above for the rich and poor country sample split, with the results reported in Appendix Figures A9-A12. The main results are robust to these different specifications. Interestingly, the measurement error simulation for the split sample indicates much more stability across simulations that the baseline case. Taken together, 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.

6. Quantitative Implications

We assess the economic significance of the estimated coefficients in two ways. The first is data-driven. We combine our estimated elasticities with information on countries' trade patterns and geographic location to quantify the heterogeneous impact of foreign demand shocks on real income across countries. The second is model-based. We set up a quantitative small open economy model of production and trade featuring an input-output matrix, endogenous capital accumulation, and sector-level scale economies, and explore whether it can reproduce the magnitude and pattern of the empirically estimated elasticities.

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

Figure 2: Developed vs. Developing Countries: Cluster-Specific Coefficients and Confidence Intervals



(a) Developed Countries

(b) Developing Countries

Notes: This figure reports estimates of the δ_g^{ex} 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.

6.1 Data: Observed Cross-Country Heterogeneity

Sectoral specialization. In this exercise, we compute the elasticity of each country's income to a worldwide uniform log-change in FMA, that is the same in every foreign sector and every foreign country. Above, we found that foreign shocks in certain sectors have a higher income impact than in others. As a result, even a foreign shock that is completely uniform across sectors would be predicted to change real income differently across countries, depending on their initial trade shares. A simple transformation of our estimating equation leads to the following expression for this elasticity:

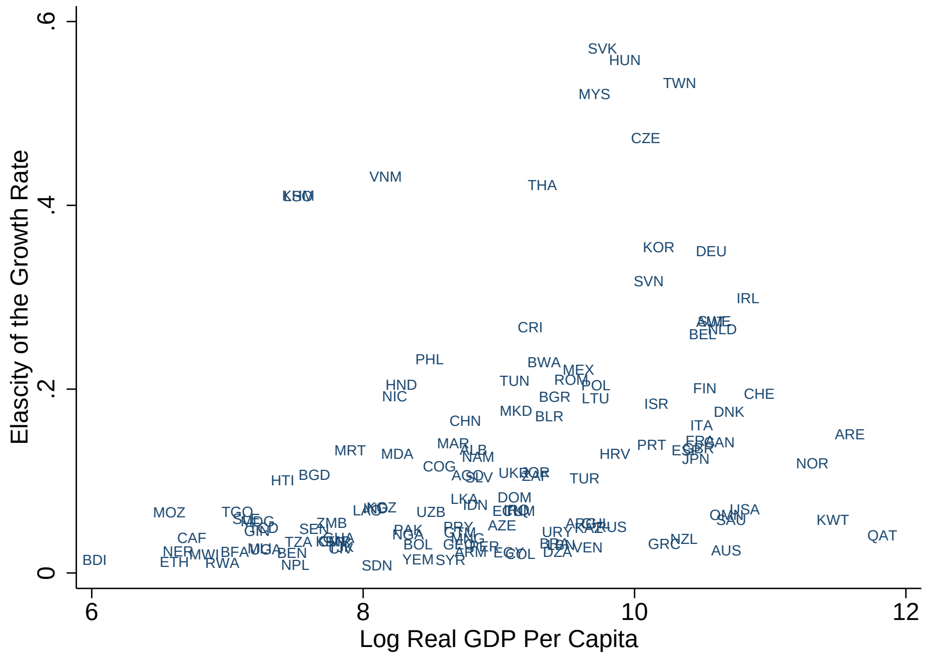
$$\frac{d \ln y_{i,t}}{d \ln FMA} = \sum_{g \in G} \widehat{\delta}_g^{ex} \lambda_{ig,t}^{ex}. \tag{6.1}$$

By imposing uniform foreign shocks across all countries and sectors, this counterfactual allows us to focus purely on the role of industrial specialization, as reflected in the $\lambda_{ig,t}^{ex}$'s. Countries that have high export shares in clusters with a high estimated income impact will have a larger positive real income response.

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 3 plots them against log PPP-adjusted income per capita.[22]

[22] As a robustness check, Appendix Figure A13 plots the same elasticities using the δ_g^{ex} estimates from Figure 2, which vary across countries according to income. Despite large differences in the estimated elasticities for capital goods, the resulting real income elasticities with respect to the uniform shock are quite similar for most countries. This is because while the capital goods foreign demand shocks have a large coefficients among developing countries, capital goods exports are quantitatively small for most poor countries (2.5% of exports on average).

Figure 3: Elasticity of Real Income with Respect to a Uniform Foreign Shock



Notes: This figure presents the scatterplot of the elasticity of real income with respect to a uniform foreign demand shock (6.1) against real GDP per capita. It is calculated using the baseline estimated δ_g^{ex} coefficients in equation (3.6) and the sectoral trade shares in 2015.

There is indeed a great deal of heterogeneity in the country impact of foreign shocks. The real income elasticity with respect to 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. This exercise suggests that a country's trade patterns matter quantitatively in how foreign shocks translate into domestic income levels.

Geography. Having illustrated how the heterogeneity in countries' sectoral specialization affects their real income response to foreign demand shocks, our next counterfactual is designed to illustrate the role of geography. The same vector of worldwide demand shocks for a particular sector (which in our model manifest themselves as changes in each country's sector-specific importer effect) translates into different changes in the external firm market access faced by each exporter due to its geographic position. As an example, suppose that in a particular period the importer effects reveal that China has a much larger demand shock for capital goods than does Germany. This pair of importer-specific shocks will affect Belgium and Vietnam quite differently, as Vietnam is closer to China than to Germany, and the opposite is true for Belgium. We would like to understand the size of this heterogeneity induced by countries' geographic positions. We thus construct counterfactual real

Table 3: Predicted Decennial Growth Difference, 2005-2015

	Growth difference, actual vs:		
	Median	25th pctile	75th pctile
G7			
Canada	-3.25	-4.45	-1.24
France	-1.88	-2.52	-1.02
Germany	-3.78	-5.14	-1.60
Italy	-0.59	-1.19	0.12
Japan	5.59	5.12	6.30
UK	-0.85	-1.40	-0.10
US	-0.26	-0.49	0.00
BRICS			
Brazil	0.34	0.09	0.79
China	-8.53	-9.19	-7.85
India	-0.10	-0.35	0.29
Russia	0.30	-0.19	0.91
South Africa	0.19	-0.30	1.03

Notes: This table reports the differences in real income growth, in percent per decade, between the actual growth and the counterfactual growth that the country would experience if it were moved to the median (resp. 25th and 75th percentile) geographic position.

income changes that would occur if Belgium experienced Vietnam's market access shocks. This counterfactual answers the following question: how much would Belgium's real income change if in a particular time period if it were picked up and moved to the place on the globe occupied by Vietnam? We do this for every pair of countries and in each decade.

To begin getting a sense of the magnitudes involved, Table 3 reports the results for a set of prominent countries, namely the G7 and the BRICS. The first column reports the difference between the country's actual decennial growth over 2005-2015 and the growth that would obtain if the country were moved to the position of the median country, where "median" means the median difference among all the possible counterfactual geographic positions. So, a value of 1 in the first column implies that the country grew 1 percentage point faster over the course of the 2005-2015 decade in its actual geographic position, relative to being moved to the median position in the world. The second and third columns report the counterfactual growth differences due to being moved to the 25th and the 75th percentile geographic position for that country.

These countries' geographic positions had a modest but noticeable impact on income. Among the G7, Japan's income is nearly 6% higher at the end of the decade in its actual geographic position relative to the hypothetical median location. By contrast, European G7 countries' income is 1-4 percentage points lower in their actual geographic position when compared to the median. The picture for the BRICS is less clear, with medians closer to zero. The exception is China, which would have been substantially better off locating in the median position.

Table 4 reports the summary statistics from the same exercise using all countries, by region and decade. Most medians are between −1% and 1% over a decade. Interestingly, within the same region the signs often flip from decade to decade. For instance, Western Europe/North American countries' income in their actual geographic locations is 0.8% lower in 2005-2015, but 0.9% higher in 1995-2005, relative to the median location. The opposite is true for East Asia and Pacific. Note that these comparisons capture the impact of *changes* in foreign demand on economic growth rates. So the negative growth differentials are perfectly consistent with West European or East Asian countries having high market access *levels*. What these results reveal is that in the last decade, the wealthy West European countries were located next to countries with relatively slow-growing demand in key sectors, and thus foreign demand has expanded more slowly for them than they would have if they had been located in faster-growing regions of the world.

In other groups of countries, the growth impact of geographic location is somewhat smaller in absolute terms, and switches sign over time. The absolute impact of geography on growth tends to rise over time, as countries become more open overall. In the last decade, the Middle East, South Asia, and Sub-Saharan Africa have enjoyed a modest benefit from their geographic position, whereas location has had a modest cost for Latin America and Eastern Europe/Central Asia.

Finally, we evaluate which geographic locations are most advantageous from each country's perspective. Thus, instead of asking how countries would fare relative to being in the geographic position of the median country in the world, we ask what would have happened if it were moved to a particular region. Appendix Table A7 presents the results for the period 2005-2015. It reports the per annum change in growth for the median country in the row region relative to its growth if it were moved to the median geographic location in the column region. For the regions at the extremes, the geographic (dis)advantage is quite pervasive. In this decade, East Asia/Pacific countries tend to exhibit higher actual growth relative to being moved to almost any region. By contrast, Western European/North American countries would grow faster in most other regions. For other regions the picture is more nuanced, and the sign of the growth impact switches across counterfactual regions. By and large, countries would experience higher growth if they moved to East and South Asia, and slower growth if they moved to Western Europe.

6.2 Theory and Quantification: "Proof-of-Concept"

We now ask whether the average level and the variation across clusters of the estimated coefficients can be generated by a quantitative trade model that embodies mechanisms that have been explored in the previous literature. To that end, we return to the small open economy model of Section 2, specify mechanisms and functional forms, calibrate it, and compare the real income elasticities with respect to foreign demand shocks inside the model to those estimated above. We stress that this is a "proof-of-concept" exercise, rather than a strong stand on the precise economic mechanisms behind the empirical estimates. There are many models, and potentially infinitely many parameter

Table 4: Predicted Decennial Growth Difference Relative to Median Geographic Location, Medians by Region and Time Period

Region	1975	1985	1995	2005	2015
East Asia & Pacific	0.27	0.35	0.56	-1.01	0.46
	[-0.06 , 0.61]	*[0.01 , 1.20]*	*[0.27 , 2.45]*	*[-2.29 , 0.08]*	*[-0.10 , 2.47]*
	10	14	14	14	14
Eastern Europe & Central Asia	0.01	-0.39	0.32	0.98	-0.87
	[-0.02 , 0.03]	*[-2.42 , -0.21]*	*[0.00 , 0.64]*	*[0.28 , 1.94]*	*[-1.56 , -0.29]*
	2	6	6	24	24
Latin America & Caribbean	-0.19	0.10	-0.21	-0.14	-0.27
	[-0.45 , 0.05]	*[-0.26 , 0.49]*	*[-0.58 , 0.07]*	*[-0.40 , 0.12]*	*[-1.41 , 0.12]*
	18	18	18	18	18
Middle East & North Africa	0.03	-0.08	-0.10	-0.10	0.33
	[-0.17 , 0.24]	*[-1.74 , 0.96]*	*[-0.66 , 0.37]*	*[-0.91 , 0.55]*	*[-0.04 , 0.75]*
	7	14	14	15	15
South Asia	0.04	0.13	-0.19	-0.79	0.07
	[-0.81 , 0.23]	*[0.05 , 0.34]*	*[-0.25 , 0.10]*	*[-1.49 , -0.29]*	*[-0.10 , 0.10]*
	4	5	5	5	5
Sub-Saharan Africa	0.02	0.03	-0.25	-0.03	0.19
	[-0.09 , 0.26]	*[-0.06 , 0.21]*	*[-0.38 , -0.06]*	*[-0.15 , 0.54]*	*[-0.05 , 0.41]*
	28	30	30	30	33
West Europe/North America	1.05	-0.30	1.39	0.92	-0.82
	[0.16 , 2.26]	*[-1.57 , 0.11]*	*[0.53 , 2.38]*	*[-0.85 , 2.25]*	*[-2.46 , -0.45]*
	18	18	18	18	18

Notes: This table reports the region- and period-specific differences in economic growth, in percent per decade, between the actual growth and the counterfactual growth that the country would experience if it were moved to the median geographic position. The numbers in square brackets are the interquartile range across countries in that region and time period. The bottom rows report the number of countries in each cell.

combinations within each model, that could in principle match our empirical estimates. The objective of this exercise is to explore whether the coefficients that we estimated can be generated by a relatively standard quantitative trade model.

To complete the description of the model, we need to 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:

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

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:

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

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 2.1). 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 (2.1), 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 (Bartelme et al., 2019; Kucheryavyy et al., 2020). 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. Since our δ^{ex}'s are estimated for the 6 tradeable sector clusters, with some abuse of notation in this section k indexes clusters, and the model is calibrated to data at the cluster level. We introduce a non-tradeable service sector, and calibrate its size and role in production to the data. We use the WIOD database (Timmer et al., 2015) 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. To do this, we treat the elasticities of substitution and of scale as free parameters, and select them to best match the vector of δ^{ex}'s across clusters estimated in the data. Since there are 6 δ^{ex} 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^{ex}. 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^{ex} from the data and the same objects in the model. We stress that the δ^{ex} 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 shares, intermediate input shares, expenditure shares and final use, as illustrated by the analytical solution in Appendix A.4.

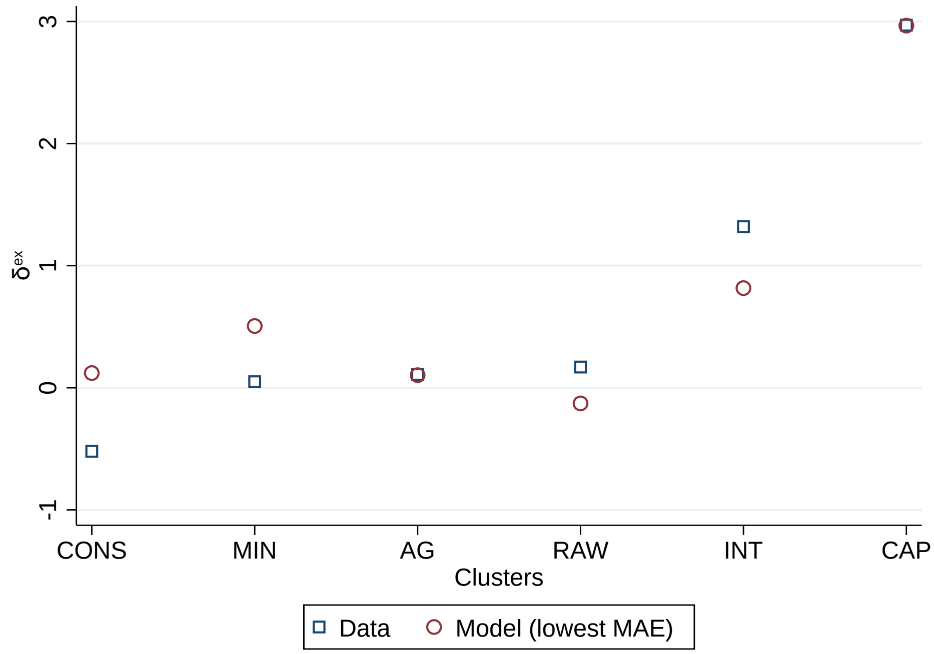
Figure 4 displays the result. It plots, for each cluster, the δ^{ex} estimated from the data and those implied by the model. The clusters are shown in increasing order of estimated δ^{ex}. The model is quite successful at replicating the estimated coefficients; the correlation between the δ^{ex} implied 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. To achieve this performance, the MAE-minimization procedure selects an elasticity of substitution $\sigma = 3.2$, and a scale elasticity $\gamma = 0.29$. 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. Bartelme et al. (2019) find a somewhat lower average value of about 0.13. Importantly, the model generates the variation in δ^{ex} 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.

Appendix Table A1 presents some diagnostics on the model performance. Appendix Figure A1 plots the MAE against γ.[23] While strictly speaking the minimum MAE criterion selects a relatively high γ of 0.29, the MAE is actually quite flat from about $\gamma = 0.17$ to 0.30. This suggests that the variation in the δ^{ex} coefficients can actually be accounted for fairly well by a wide range of reasonable parameter values. The dashed line displays the average δ^{ex} 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^{ex} right. Interestingly, the model matches the δ^{ex} for the capital goods cluster – by far the highest δ^{ex} in the data – almost exactly for all γ between 0.13 and 0.3. Thus, the sensitivity of the average δ^{ex} to γ in the model is driven by other clusters.

Mechanisms. 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. Appendix A.4 presents the formulas and the more detailed discussion. Here, "first-" and "higher-order" are used in the input-output sense of intermediates being used directly vs. indirectly, not to be confused with the first-order Taylor approximation to the solution that is used throughout. Intuitively, the direct effect only applies

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

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



Notes: This figure plots the δ^{ex} coefficients as estimated in the data, and generated by the model, selecting σ and γ to minimize the MAE between the data and model δ^{ex}'s.

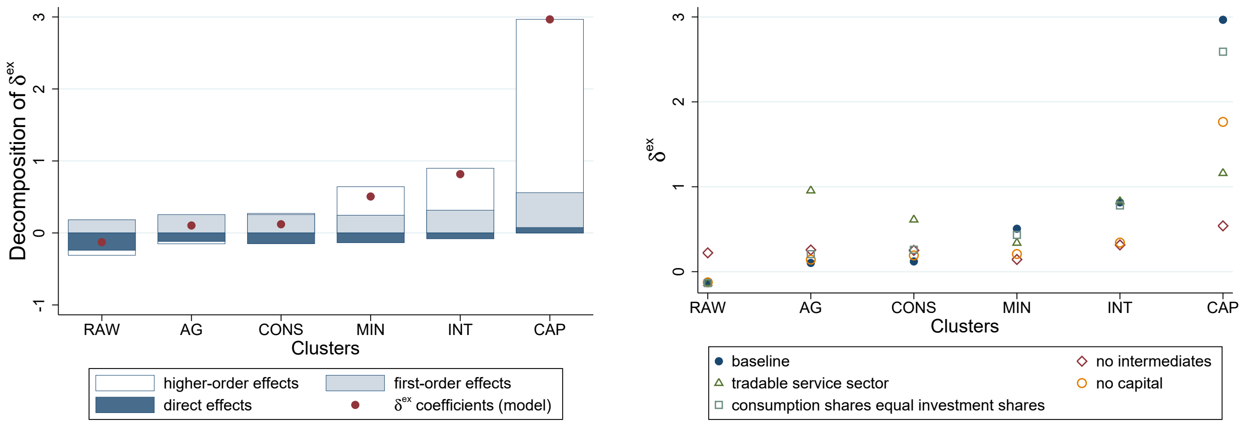
to the sector experiencing the foreign demand shock and reflects how an increase in foreign sales translates into higher aggregate sales in partial equilibrium. The first-order effect reflects the fact that the sector experiencing a foreign demand shock changes its purchases of intermediates, and the change in its value added affects final demand inside the home economy. Importantly, it also captures the changes in unit costs, both through wages and scale effects (see equation A.16). 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.[24]

To highlight which determinants of higher-order propagation are key, we examine a set of al-

[24]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.

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^{ex} into direct, first-order, and higher-order effects. The right panel displays model δ^{ex} under alternative production structures.

ternative 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. This is essentially a static trade model with labor as the only primary factor. The third 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^{ex}) 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). This 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).[25] Only internal propagation mechanisms inside the economy differ between the alternative models and the baseline.

[25] As pointed out by Baqaee and Farhi (2019), 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), and avoids the need to reshuffle the entire trade share vector.

The right panel of Figure 5 displays the model-implied δ^{ex} coefficients in the baseline and the alternative models. The last 4 columns of Appendix Table A1 report the averages of δ^{ex} in these alternative models, as well as the MAE and correlations between these models and the data. Removing the input-output linkages has the largest impact on the model-implied δ^{ex}. 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^{ex} and a coefficient of 1.8 in the capital sector, much closer to the data. Nonetheless, its average δ^{ex} 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^{ex} falls by more than 50%, and the model does not generate coefficients that closely match the observed variation across sectors.

Finally, Appendix A.4 explores how successful this model can be without scale effects. It can match the average δ^{ex} under $\sigma = 2.2$, which is low relative to conventional wisdom, but cannot match the observed dispersion across sectors. 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^{ex} about one-third of the data value, low dispersion across sectors, and the complex intermediate and capital goods coefficients that are too small. In the absence of scale effects, the calibrated model lacks sufficient internal propagation mechanisms necessary to generate the observed amount of dispersion in δ^{ex} across sectors for any values of σ.

The model with a positive but common scale elasticity generates large dispersion in the δ_k^{ex} 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 (Bartelme et al., 2019). These effects are further amplified when the sector is relatively upstream from the non-traded sector, which (all else equal) is also too small due to the positive terms-of-trade effects associated with its expansion.[26]

Our main conclusions from this 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 cluster-level 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. However, substantial scale effects appear

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

important for the current crop of quantitative trade models to match the long run general equilibrium response of economies to foreign demand shocks.

Second, 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.

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 have a significantly higher real income impact than shocks in other sectors. Our estimates imply that the interaction between initial specialization and the pattern of foreign shocks is important for understanding the variety of growth experiences across countries. 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 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 speak to these debates insofar as they affirm the importance of the external environment for economic development and validate a focus on the sectoral dimensions of policy. At the same time, it is important to stress that our results do not in and of themselves imply that countries should use government policy to encourage specialization in certain sectors, or that productivity growth is more valuable in some sectors relative to others. These propositions pertain to the impact of domestic policies and productivity shocks, and our paper does not provide any direct evidence on these effects. Our findings do imply that, all else equal, countries should pursue increased market access more vigorously in some sectors relative to others. A fuller understanding of optimal sectoral policy requires considering domestic policies as well, along with the ever-mysterious drivers of productivity growth.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "Institutions as a Fundamental Cause of Long-Run Growth," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1, Elsevier, 2005, chapter 6, pp. 385–472.

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

_ , 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.

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, "Trade costs and economic geography: evidence from the US," *mimeo, University of Michigan*, 2018.

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

Becker, Gary, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, NBER, 1975.

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.

Beshkar, Mostafa and Ahmad Lashkaripour, "Interdependence of trade policies in general equilibrium," 2019. mimeo, Indiana University.

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.

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

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.

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

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.

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, Cesar Hidalgo, Sebastian Bustos, Michele Coscia, Sarah Chung, Juan Jimenez, Alexander Simoes, and Muhammed A. Yildirim, *The atlas of economic complexity: Mapping paths to prosperity*, MIT Press, 2014.

_ , Jason Hwang, and Dani Rodrik, “What You Export Matters,” *Journal of Economic Growth*, 2007, *12*, 1–25.

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.

Jones, Charles I, “Intermediate Goods and Weak Links in the Theory of Economic Development,” *American Economic Journal: Macroeconomics*, April 2011, *3* (2), 1–28.

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.

Kleinman, Benny, Ernest Liu, and Stephen J Redding, “International Friends and Enemies,” July 2020. NBER Working Paper 27587.

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.

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, "When and How Should Infant Industries Be Protected?," *Journal of International Economics*, 2005, *66*, 177–196.

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.

Solow, Robert M., "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, February 1956, *70* (1), 65–94.

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.

Ventura, Jaume, "A Global View of Economic Growth," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1, Elsevier, 2005, chapter 22, pp. 1419 – 1497.

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:

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

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

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

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

$$Y_k = c_k^{1-\sigma_k} \left(z_k \frac{e_k \cdot E + \sum_{l \in K} \alpha_{l,k} Y_l}{P_k^{1-\sigma_k}} + FMA_k \right)\ \ \forall k \in K. \tag{A.5}$$

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 $\alpha_{l,k}$ is the fraction of industry l's gross revenue (Y_l) 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 the relevant price. Note that 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 (E) or industry gross output.

The first set of conditions (A.1) are the industry factor demand equations, which can be summed to generate aggregate factor demand. The second set of conditions (A.2) equates factor demand with fixed 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 Approximation

A general expression for our first order welfare approximation is

$$d\ln y = \sum_{k\in K} \lambda_k^{ex} d\ln FMA_k + \sum_{k\in K}\left(\frac{e_k}{\sigma_k - 1} - \lambda_k^d\right)\theta_k^f d\ln CMA_k + \sum_{k\in K} d\ln \mu_k + \sum_{k\in K}\lambda_k^d d\ln E_k + \sum_{k\in K}\left((1-\sigma_k)(\lambda_k - \lambda_k^d\theta_k^d) - e_k\theta_k^d\right) d\ln c_k,$$

where λ_k^{ex} (resp. λ_k^d) is the share of total sales attributable to industry k's export (resp. domestic) sales, $\lambda_k = \lambda_k^d + \lambda_k^{ex}$, e_k is the consumer expenditure share on industry k, and θ_k^d (resp. θ_k^f) is the share of expenditure on industry k that is sourced domestically (resp. foreign).

Since $d\ln \mu_k$, $d\ln c_k$ and $d\ln E_k$ are all ultimately functions of the exogenous variables $d\ln FMA_k$, we can substitute in for these variables to arrive at equation (2.6) in the main text.

We can also interpret the elasticities δ_k^{ex} in terms of parameters more directly relevant to trade policy as well as to the types of counterfactual experiments more often conducted in the literature. Imagine that Home's trading partners lower their import tariffs on Home's exports in sector k. From Home's perspective, this policy change is equivalent to a reduction in the iceberg costs of exporting to the rest of the world. The general equilibrium impact of such a policy on Home's welfare will be given by

$$\frac{\partial \ln y}{\partial \ln \tau_k^{ex}} = (\sigma_k - 1)\cdot \delta_k^{ex}, \tag{A.6}$$

where the trade elasticity $\sigma_k - 1$ scales the market access elasticity by the effect of the tariff change on Home's market access. The object on the LHS of equation (A.6) is of crucial importance to policymakers; given the trade elasticity, the δ_k^{ex} capture all the remaining welfare effects of a trading partner's import policy on a small economy.[27]

A.3 The Determinants of δ_k^{ex}

We now discuss the factors that determine the size of the δ_k^{ex} and their variation across industries.

[27]The LHS of equation (A.6) is more directly relevant to trade policy. The primary reasons why we do not define it as the object of interest are data availability and statistical power. Changes in tariffs are simply not widely observable enough (across countries and time) nor large enough in magnitude to directly estimate $\frac{\partial \ln y}{\partial \ln \tau_k^{ex}}$. In contrast, our estimation approach leverages variation in market access due to aggregate and sectoral demand and technology shocks in addition to variation in tariffs, and utilizes sectoral trade data that are widely available from the 1960s onwards. See Sections 3 and 4 for more details. In addition, understanding the consequences of the large variations in foreign demand over time and space is of direct interest.

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 $q_{n,k}^{ex}$, and the production function in each sector by F_k. Define $D_{n,k} \equiv \tau_{n,k}^{1-\sigma_k} E_{n,k}/P_{n,k}^{1-\sigma_k}$.[28]

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

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

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

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

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

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

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

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 $q_k^{ex} = \sum_{n\in N} q_{n,k}^{ex}$, $q_k^{c,f} = (\sum_{n\in N} (q_{n,k}^{c,f})^{\frac{\sigma_k-1}{\sigma_k}})^{\frac{\sigma_k}{\sigma_k-1}}$ and

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

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

$q_k^{m,f} = (\sum_{n\in N}(q_{n,k}^{m,f})^{\frac{\sigma_k-1}{\sigma_k}})^{\frac{\sigma_k}{\sigma_k-1}}$ such that the problem above is equivalent to

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

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

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

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

We now derive the formulas for δ_k^{ex} for an efficient economy. A simple application of the Envelope Theorem gives

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

where ϑ is the multiplier on the trade balance constraint. Our assumption of homotheticity allows us to normalize this constant to equal 1.

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 $\sum_{k\in K} \lambda_k^{ex} \ln p_k$; 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.[29]

Economy without 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 is a monopolist and 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.[30] 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 (labor) economy, the percentage increase in labor demand is

[29] As $\sigma_k \to \infty$, we approach the case of a small open economy that can sell any amount at a fixed foreign price. Since the size of the market is already effectively infinite, the welfare effect of an increase in the size of the market tends to zero.

[30] See Bartelme et al. (2019) and Beshkar and Lashkaripour (2019) for optimal trade policy in this environment.

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^{ex}, the market access elasticities δ_k^{ex} 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

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

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

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

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

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

To solve for the changes in real income, we need to consider the effect on the overall price index $\mathbb{P} = \prod_{k\in K}P_k^{e_k}$. 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, we get

$$\begin{aligned}
d\ln y &\approx d\ln w - \sum_{k\in K}(e_k-\lambda_k^{im})d\ln w \\
&= \lambda^{im}d\ln w \\
&= \lambda^{im}\cdot\sum_{k\in K}\frac{\lambda_k^{ex}d\ln FMA_k}{1-\sum_{k'\in K}\left(\lambda_{k'}^d+(1-\sigma_{k'})\left(\lambda_{k'}^d\theta_{k'}^f+\lambda_{k'}^{ex}\right)\right)} \\
&= \delta^{ex}\cdot\sum_{k\in K}\lambda_k^{ex}d\ln FMA_k,
\end{aligned}$$

with

$$\delta^{ex} = \frac{\lambda^{im}}{1 - \sum_{k' \in K} \left(\lambda^{d}_{k'} + (1 - \sigma_{k'}) \left(\lambda^{d}_{k'} \theta^{f}_{k'} + \lambda^{ex}_{k'} \right) \right)},$$

where $\lambda^{im} = \sum_{k \in K} \lambda^{im}_{k}$. Thus, in this environment δ^{ex}_{k} do not vary across sectors. This expression simplifies further when we set the domestic sales share in each industry, $\theta^{d}_{k'}$, equal to zero:

$$\delta^{ex} = \frac{1}{1 - \sum_{k' \in K} (1 - \sigma_{k'}) \lambda^{ex}_{k'}}.$$

For a general homothetic upper tier, the formula would have to be modified to account for changes in industry expenditure shares, although the δ^{ex}_{k} would still be common across industries.

In the presence of multiple factors with different intensities across sectors or general production networks, foreign demand shocks in different sectors imply different changes in factor demand and hence heterogeneous indirect effects. Section 6 and Appendix A.4 spell out one such economy, and equations (A.12), (A.13) and (A.15) state an analytical solution for δ^{ex}_{k}. Clearly, the δ^{ex}_{k} generically differ across sectors in this economy, even in the absence of any domestic distortions ($\gamma = 0$), and/or differences in trade elasticities across sectors ($\sigma_k = \sigma \ \forall k$). While the analytical solution is not sufficiently transparent to see how individual parameters affect the δ^{ex}_{k} of individual sectors, it is clear that in general the δ^{ex}_{k}'s depend on the full structure of the economy.

External Economies

The second 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 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

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

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

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

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

where

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

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

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

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.[31]

In more elaborate quantitative environments that feature multiple factors of production, input-output networks and other mechanisms, the interaction between international and domestic distortions that determine the values of the δ_k^{ex} becomes increasingly complex and sensitive to assumptions. We provide an example of this type of exercise in Section 6, with details in the appendix section below.

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

[31] 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.

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:

$$Y_k = \frac{z_k c_{kt}^{1-\sigma_k}}{P_{kt}^{1-\sigma_k}} \sum_l \left(f_k \mu + \widetilde{\alpha}_{k,l}\right) Y_l + c_k^{1-\sigma_k} FMA_k \ \ \forall k \tag{A.7}$$

$$P_k^{1-\sigma_k} = z_k c_k^{1-\sigma_k} + CMA_k \ \ \forall k \tag{A.8}$$

$$c_k = T_k L_k^{-\gamma_k} w^{\beta\mu} \prod_l P_l^{\widetilde{\alpha}_{l,k}+\mu(1-\beta)\nu_l} \ \ \forall k \tag{A.9}$$

$$wL_k = \mu\beta Y_k \ \ \forall k \tag{A.10}$$

$$\sum_k L_k = L. \tag{A.11}$$

Mapping to regression coefficients. Note that while in the statement of equilibrium conditions (A.7), FMA_k enters by itself, in actual empirical estimation the regressor is weighted by the export share: $\lambda_k^{ex} FMA_k$, see (2.6). Thus, we state the model solution directly in terms of export-share-weighted firm market access: $FMA_k^W \equiv \lambda_k^{ex} FMA_k$. 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:

$$\begin{aligned} d\ln \mathbf{Y} &= \left\{ \mathbf{I} - \left[\mathbf{\Pi}^d + (\mathbf{I} - \boldsymbol{\sigma}) \left((\mathbf{I} - \boldsymbol{\pi}^{ex}) \left(\mathbf{I} - \boldsymbol{\theta}^d \right) + \boldsymbol{\pi}^{ex} \right) \left(\mathbf{I} - \mathbf{A}\boldsymbol{\theta}^d \right)^{-1} \left(\gamma \left(\boldsymbol{\lambda} \otimes \mathbf{1} - \mathbf{I} \right) + \mu\beta \boldsymbol{\lambda} \otimes \mathbf{1} \right) \right] \right\}^{-1} \\ &\quad \times diag\left(\boldsymbol{\lambda}\right)^{-1} d \ln \mathbf{FMA^W} \end{aligned} \tag{A.12}$$

$$d \ln \mathbf{P} = \boldsymbol{\theta}^d \left(\mathbf{I} - \mathbf{A}\boldsymbol{\theta}^d \right)^{-1} \left(\gamma \left(\boldsymbol{\lambda} \otimes \mathbf{1} - \mathbf{I} \right) + \mu\beta \boldsymbol{\lambda} \otimes \mathbf{1} \right) d \ln \mathbf{Y}. \tag{A.13}$$

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

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

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

- Matrix $\mathbf{A}$ 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.12)-(A.13) 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

$$d\ln y = \boldsymbol{\lambda}\cdot d\ln\mathbf{Y} - \mathbf{f}\cdot d\ln\mathbf{P}. \tag{A.14}$$

Plugging (A.12)-(A.13) into (A.14), we obtain the following model elasticities with respect to foreign market access shocks:

$$\frac{d\ln y}{d\ln\mathbf{FMA^W}} \quad = \quad \left(\boldsymbol{\lambda} - \mathbf{f}\cdot\boldsymbol{\theta}^d\left(\mathbf{I}-\mathbf{A}\boldsymbol{\theta}^d\right)^{-1}\left(\gamma\left(\boldsymbol{\lambda}\otimes\mathbf{1}-\mathbf{I}\right)+\mu\beta\boldsymbol{\lambda}\otimes\mathbf{1}\right)\right)\frac{d\ln\mathbf{Y}}{d\ln\mathbf{FMA^W}}, \tag{A.15}$$

where $\frac{d\ln\mathbf{Y}}{d\ln\mathbf{FMA^W}}$ is given by (A.12).

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.15) is the theoretical counterpart of the econometrically estimated elasticities of GDP with respect to foreign demand shocks, δ^{ex}.

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 $\boldsymbol{\lambda}$, $\mathbf{f}$, $\boldsymbol{\theta}^d$, $\mathbf{A}$, $\boldsymbol{\pi}^{ex}$, 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. The matrices $\mathbf{f}$, $\boldsymbol{\theta}^d$, 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 averages across the 40 countries available in that database. The matrices $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}^{ex}$ are constructed to be consistent with the averages of import and export shares used in the econometric estimation. 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":

$$\boldsymbol{\Omega} \quad \equiv \quad \boldsymbol{\Pi}^d + (\mathbf{I}-\boldsymbol{\sigma})\left((\mathbf{I}-\pi^{ex})\left(\mathbf{I}-\theta^d\right)+\pi^{ex}\right)\left(\mathbf{I}-\mathbf{A}\theta^d\right)^{-1}\left(\gamma\left(\boldsymbol{\lambda}\otimes\mathbf{1}-\mathbf{I}\right)+\mu\beta\boldsymbol{\lambda}\otimes\mathbf{1}\right). \tag{A.16}$$

Then, the model solution can be stated as:

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

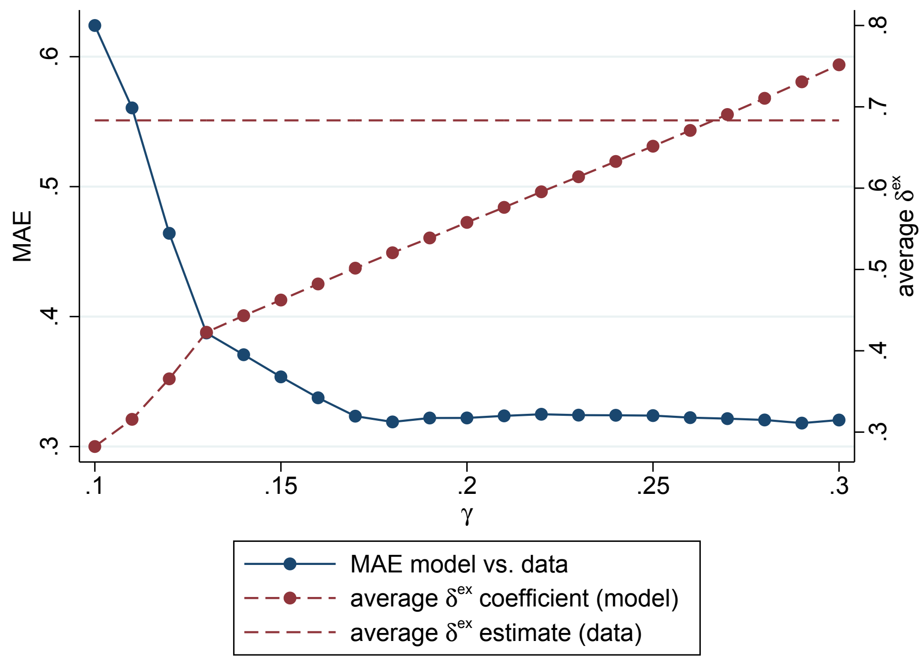
where the second line writes the Leontief inverse as an infinite expansion. The first term is the direct effect of a foreign demand in a sector. The matrix is diagonal, and thus the direct effect only applies to the sector experiencing the foreign demand shock. The first-order effect is given by the impact matrix $\mathbf{\Omega}$. Examining (A.16), the first-order effect is in turn comprised of two terms. The first, $\mathbf{\Pi}^d$, 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 $(\mathbf{I}-\boldsymbol{\sigma})\left((\mathbf{I}-\boldsymbol{\pi}^{ex})\left(\mathbf{I}-\boldsymbol{\theta}^d\right)+\boldsymbol{\pi}^{ex}\right)\frac{d\ln \mathbf{c}}{d\ln \mathbf{FMA^W}}$. The unit costs will change both because of the fact that factor reallocation affects production scale (captured by $\boldsymbol{\gamma}(\boldsymbol{\lambda}\otimes\mathbf{1}-\mathbf{I})$), 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 $(\mathbf{I}-\boldsymbol{\sigma})\boldsymbol{\pi}^{ex}$), as well as domestic sales (by $(\mathbf{I}-\boldsymbol{\sigma})(\mathbf{I}-\boldsymbol{\pi}^{ex})\left(\mathbf{I}-\boldsymbol{\theta}^d\right)$). 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.

No scale economies. We end by evaluating how well this model can match estimated growth effects of foreign shocks without appealing to external scale economies. Table A2 considers a range of models with constant returns to scale ($\gamma = 0$). We first report the model-implied δ^{ex} under a range of σ from 1 to 10. The average δ^{ex} falls in the Armington elasticity. It is not difficult to get average growth effects to be the same as estimated in the data, by simply lowering σ. However, this leads to 2 problems: first, lowering σ leads to δ^{ex} that are much too high in 4 out of 6 clusters, where the data $\widehat{\delta}^{ex}$ are near zero or negative. So just varying σ can get the average level right at the expense of missing the dispersion. This is made clear by the MAE's, which are 2-3 times higher in this table than under the optimal combination of σ and γ in Table A1. Second, the σ needed to match the average estimated $\widehat{\delta}^{ex}$, 2.2, is low relative to conventional wisdom.

Perhaps we can do better by appealing to variation in σ across sectors. The last two columns of Table A2 report the results of 2 exercises. First, we use the σ_k from Caliendo and Parro (2015). Second, we ask the model to select 6 sector-specific σ_k to minimize the MAE with respect to data. Interestingly, in both cases, the average δ^{ex}'s are much too low, and the MAE is barely lower than simply using a single σ. We conclude that the constant-returns to scale version of this particular model can match the average, but not the dispersion in δ^{ex} that we estimate in the data. We acknowledge that this should not be interpreted as conclusive evidence in favor of scale effects, as many more constant-returns models are possible than the one we consider here.

A feature of Table A2 that stands out is that the correlation between the model and data δ^{ex} is high throughout the table, at 0.91. What is interesting is that while all of these models miss on both the level and the dispersion of δ^{ex} across sectors, they all obtain the correct *ranking* of the δ^{ex}.

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^{ex} Coefficients: Data and Model

	$\widehat{\delta}^{ex}$ data	δ^{ex} 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^{ex}	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 $\widehat{\delta}^{ex}$ (first column), and the model δ^{ex} 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^{ex} for each case, and for the theoretical models reports the MAE and the correlation of δ^{ex} implied by the model and data.

Table A2: δ^{ex} Coefficients: Data and Model, No Scale Effects

	$\widehat{\delta}^{ex}$ data	δ^{ex} model ($\gamma = 0$)				
		$\sigma = 1.01$	$\sigma = 3$	$\sigma = 10$	σ_k from Caliendo-Parro	σ_k lowest MAE
RAW	0.17	1.27	0.36	0.10	0.09	0.17
INT	1.32	1.91	0.54	0.15	0.13	0.26
CAP	2.97	2.86	0.80	0.23	0.20	0.38
CONS	-0.52	1.64	0.46	0.13	0.11	0.22
AG	0.11	1.38	0.39	0.11	0.10	0.19
MIN	0.05	1.39	0.39	0.11	0.10	0.19
Average δ^{ex}	0.68	1.74	0.49	0.14	0.12	0.23
MAE model vs. data		1.10	0.79	0.78	0.79	0.77
Corr. model vs. data		0.91	0.91	0.91	0.91	0.91

Notes: This table reports the econometric estimates of $\widehat{\delta}^{ex}$ (first column), and the model δ^{ex} 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 uses sector-specific σ_k estimates from Caliendo and Parro (2015), and the last column selects sector-specific σ_k's that minimize the Mean Absolute Error (MAE) between model and data. The bottom panel reports the average δ^{ex} for each case, and for the theoretical models reports the MAE and the correlation of δ^{ex} implied by the model and data.

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

Rousseeuw (1987) introduces the silhouette plot as a means for selecting the number of clusters. 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:

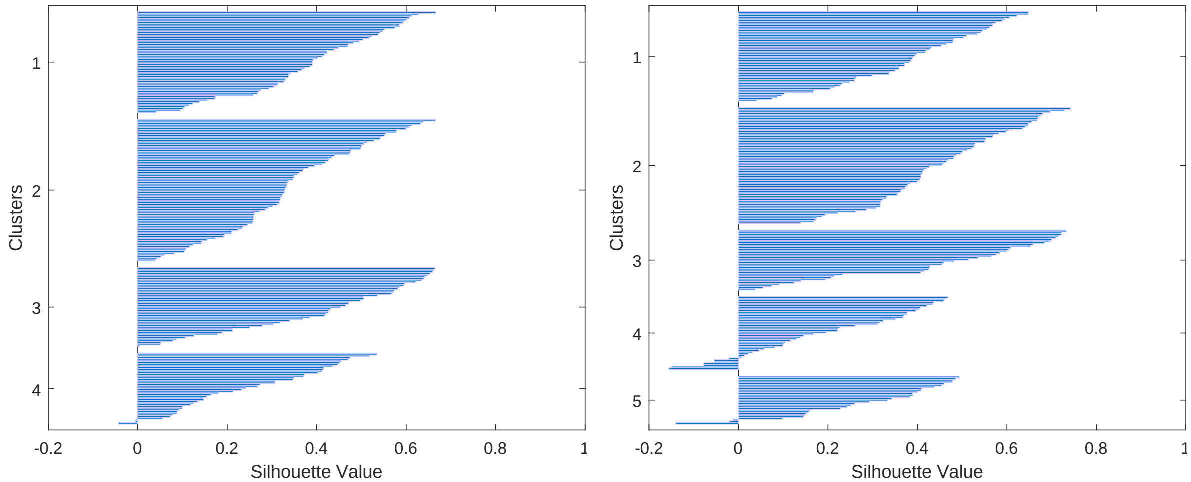
1. (*Measuring the cohesion*) Denote by a_k the average distance between point k and all other points in the same cluster.

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 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.

Figure A2: 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).

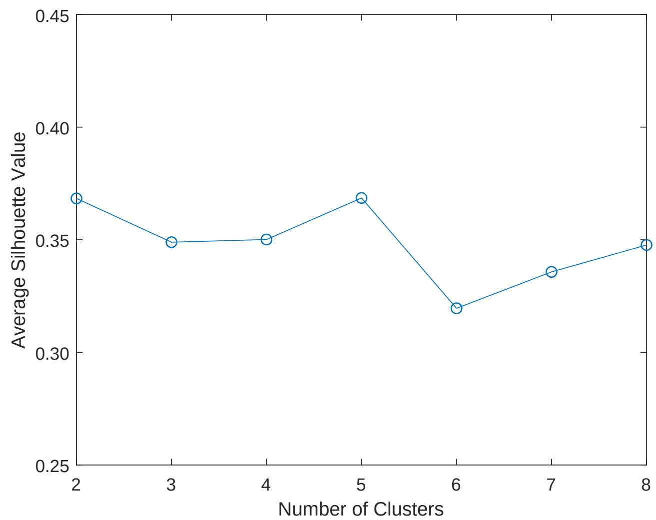
The average silhouette width provides an evaluation of clustering validity, and can be used as way to select an appropriate number of clusters. A high average silhouette width indicates a strong clustering. The average silhouette method computes the average silhouette of observations for different numbers of clusters G. The optimal number of clusters G is the one that maximizes the average silhouette over a range of possible values for G.

Figure A2 plots the silhouette width for industries in each cluster when there are 4 and 5 clusters, and Figure A3 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 slightly prefers 5 clusters to 4, the silhouette analysis suggests that with 4 clusters fewer industries are near the boundary.

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

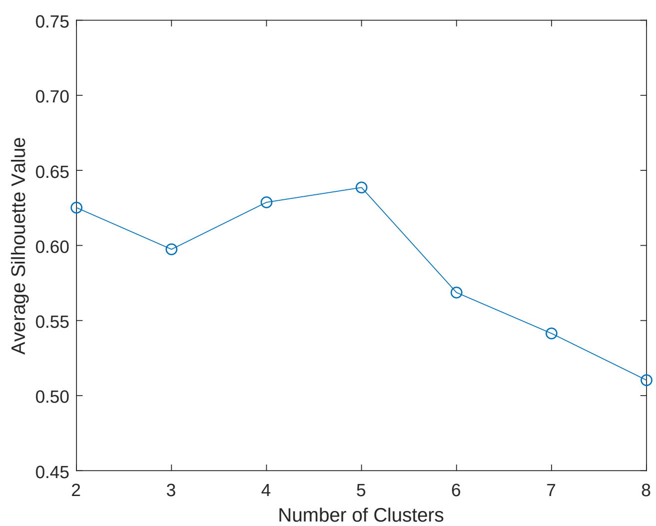
The average silhouette value under 4 clusters is about 0.35, which indicates that the cluster structure is somewhat weak. However, this could be due to the inclusion of irrelevant sectoral characteristics, which tend to drag down the average silhouette value. We investigate this hypothesis by implementing 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 important 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 A4), suggesting a strong cluster structure. Table A3 reports the summary statistics for sectoral characteristics of

Figure A3: Average Silhouette Value



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

Figure A4: 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.

each cluster. The 4 clusters based on these three characteristics closely replicate the baseline cluster structure reported in Table 1.

Table A3: 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.

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 (2.3) 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 (2.2) 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 $\{\kappa^{im}_{nk,t}(\omega)\ \kappa^{ex}_{ik,t}(\omega)\ \zeta_{kt}(\omega)\ \xi_{kt}(\omega)\}$ 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:

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

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, $\kappa^{im}_{nk,t}$ denote the renormalized importer fixed effects when country ω is omitted. These importer fixed effects are estimates of the destination-n demand shifter $\frac{E_{nk,t}}{P^{1-\sigma_k}_{nk,t} \cdot \sum_{i \neq n} E_{ink,t}}$. The iceberg bilateral components $\tau^{1-\sigma_k}_{ink,t}$ 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 $Distance^{\zeta_{kt}}_{in} \cdot \exp(\xi_{kt} \cdot Contig_{in})$.

B.4 The Post-Double-Selection Method

B.4.1 Estimating Equation

The estimating equation is

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

where the $d \ln FMA_{ig,t} = \sum_{k \in G} \lambda^{ex}_{ik,t} d \ln FMA_{ik,t}$ 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 ($\lambda^{im}_{ik,t}$ and $\lambda^{ex}_{ik,t}$), and weighted initial firm and consumer market access ($\lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t}$ and $\lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t}$). The vector $d \ln \mathbf{a}_{it}$ collects the observed contemporaneous foreign supply shocks, i.e. $\lambda^{im}_{ig,t} d \ln CMA_{ig,t}$.

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 $\lambda^{ex}_{ig,t} d \ln FMA_{ig,t}$, 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^{ex} 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 (2.6) to include foreign supply shocks.[32]

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

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 (2.4), 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:

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

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

Figure A14 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 A15 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

[32] External consumer market access enters into the welfare expression (2.5) 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 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.

C. Additional Appendix Tables and Figures

Table A4: Control Variables Selected in the Double-Selection LASSO Procedure: Baseline Estimation

Admissible Controls	Controls Selected: Baseline	Controls Selected: Developed Countries	Controls Selected: Developing Countries
$\lambda^{ex}_{ik,t}$	$\lambda^{ex}_{i104,t}$ $\lambda^{ex}_{i176,t}$	$\lambda^{ex}_{i178,t}$	$\lambda^{ex}_{i65,t}$
$\lambda^{im}_{ik,t}$			$\lambda^{im}_{i263,t}$
$\lambda^{ex}_{ig,t}$			
$\lambda^{im}_{ig,t}$			
$\lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t}$	$\lambda^{ex}_{i114,t} \cdot \ln FMA_{i114,t}$ $\lambda^{ex}_{i143,t} \cdot \ln FMA_{i143,t}$	$\lambda^{ex}_{i92,t} \cdot \ln FMA_{i92,t}$	$\lambda^{ex}_{i243,t} \cdot \ln FMA_{i243,t}$
$\lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t}$	$\lambda^{im}_{i166,t} \cdot \ln CMA_{i166,t}$ $\lambda^{im}_{i176,t} \cdot \ln CMA_{i176,t}$	$\lambda^{im}_{i205,t} \cdot \ln CMA_{i205,t}$	
$\sum_{k\in g} \lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t}$			
$\sum_{k\in g} \lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t}$			
$\sum_{k\in g} \lambda^{im}_{ik,t} \cdot d\ln CMA_{ik,t}$			
$\ln population_{i,t}$			
$\ln k_{i,t}$			
$\ln y_{i,t}$	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 2	Figure 2

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 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		

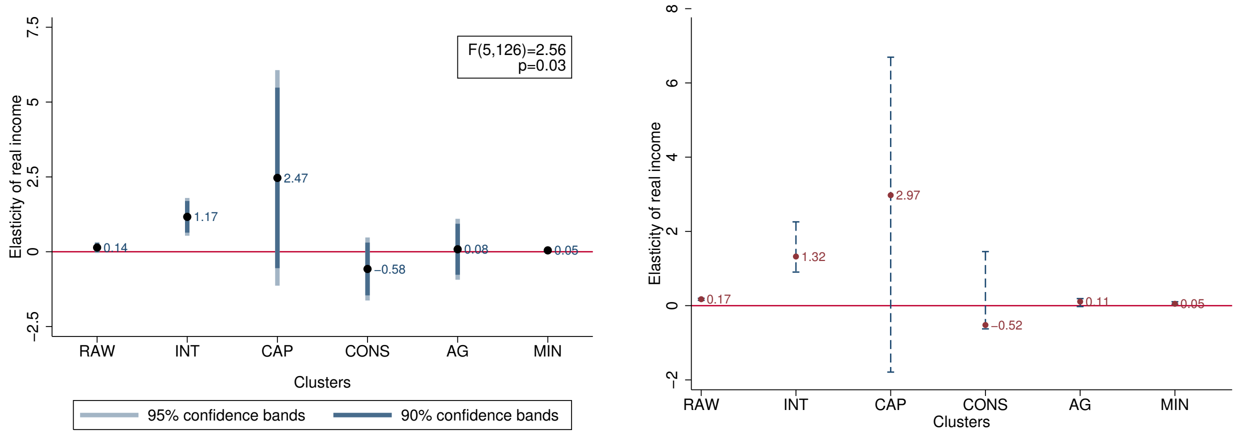
Notes: This table reports the summary statistics of the sectoral characteristics among the sectors selected into each cluster, when the number of clusters is 5. The last two columns report the mean and standard deviations of those characteristics among all manufacturing sectors. The row "Num. of ind" 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: Robustness Checks

Admissible Controls	Controls Selected: Dropping Large Trading Partners	Controls Selected: Dropping Contiguous Countries
$\lambda^{ex}_{ik,t}$	$\lambda^{ex}_{i94,t}$ $\lambda^{ex}_{i104,t}$ $\lambda^{ex}_{i111,t}$ $\lambda^{ex}_{i114,t}$ $\lambda^{ex}_{i158,t}$ $\lambda^{ex}_{i176,t}$	$\lambda^{ex}_{i111,t}$ $\lambda^{ex}_{i176,t}$ $\lambda^{ex}_{i182,t}$
$\lambda^{im}_{ik,t}$		
$\lambda^{ex}_{ig,t}$		
$\lambda^{im}_{ig,t}$		
$\lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t}$	$\lambda^{ex}_{i114,t} \cdot \ln FMA_{i114,t}$ $\lambda^{ex}_{i152,t} \cdot \ln FMA_{i152,t}$ $\lambda^{ex}_{i175,t} \cdot \ln FMA_{i175,t}$ $\lambda^{ex}_{i186,t} \cdot \ln FMA_{i186,t}$ $\lambda^{ex}_{i190,t} \cdot \ln FMA_{i190,t}$	$\lambda^{ex}_{i143,t} \cdot \ln FMA_{i143,t}$ $\lambda^{ex}_{i152,t} \cdot \ln FMA_{i152,t}$ $\lambda^{ex}_{i186,t} \cdot \ln FMA_{i186,t}$ $\lambda^{ex}_{i202,t} \cdot \ln FMA_{ik202,t}$ $\lambda^{ex}_{i203,t} \cdot \ln FMA_{i203,t}$
$\lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t}$	$\lambda^{im}_{i166,t} \cdot \ln CMA_{i166,t}$ $\lambda^{im}_{i229,t} \cdot \ln CMA_{i229,t}$	$\lambda^{im}_{i166,t} \cdot \ln CMA_{i166,t}$
$\sum_{k \in g} \lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t}$		
$\sum_{k \in g} \lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t}$		
$\sum_{k \in g} \lambda^{im}_{ik,t} \cdot d \ln CMA_{ik,t}$		
$\ln population_{i,t}$		
$\ln k_{i,t}$		
$\ln y_{i,t}$	included	included
$(\ln y_{i,t})^2$		
Time effects	included	included
Number of Controls Selected	13	9
Estimates Figures	Figure A5	Figure A5

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.

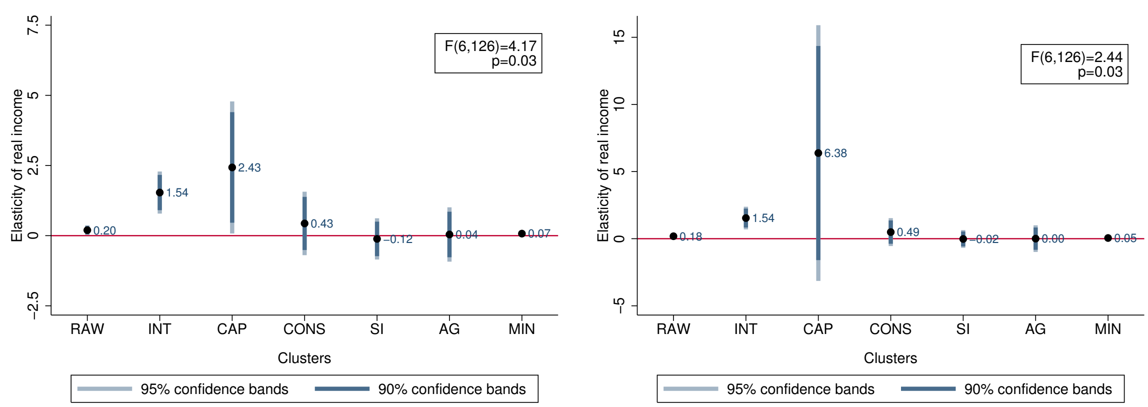
Figure A5: Cluster-Specific Coefficients and Confidence Intervals, Decreased Tuning Parameter and Cluster Measurement Error



(a) Decreased Tuning Parameter (b) Cluster Measurement Error Simulation

Notes: This figure reports estimates of the δ_g^{ex} 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. Panel (a) reports the results with a decreased tuning parameter. 15 control variables are selected in the double-selection step. Panel (b) reports the range in coefficient estimates in the measurement error simulations described in Section 5.3. The vertical bars report the 95% range of coefficient point estimates.

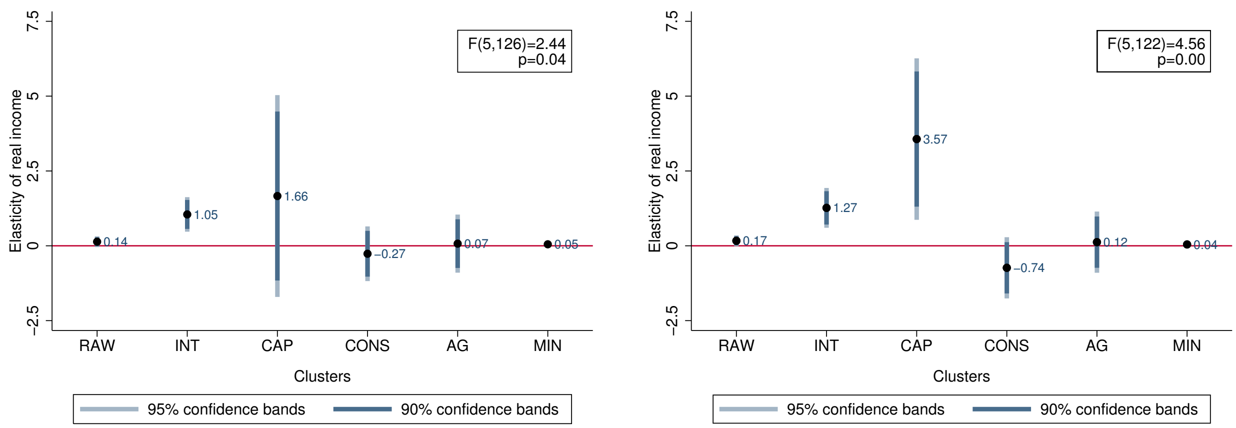
Figure A6: Cluster-Specific Coefficients and Confidence Intervals When Grouping the Manufacturing Industries to 5 Clusters



(a) OLS Estimates (b) LASSO Estimates

Notes: This figure reports estimates of the δ_g^{ex} 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 A7: Cluster-Specific Coefficients and Confidence Intervals, Robustness to Small Country Assumption

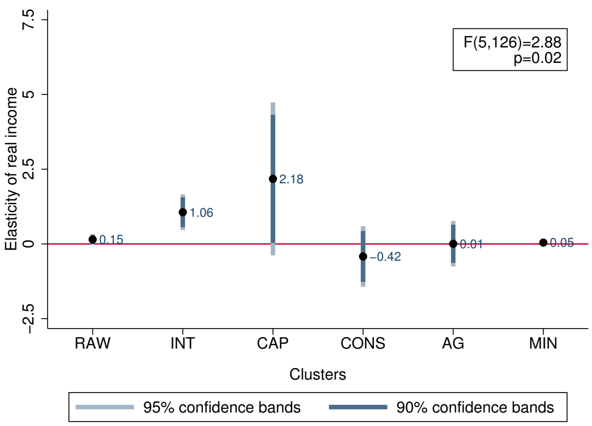


(a) Dropping Large Trading Partners

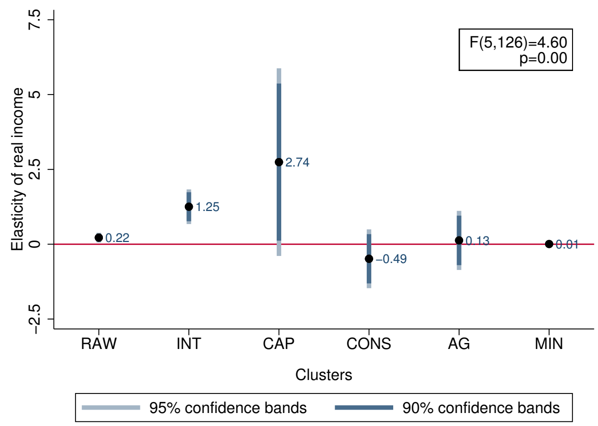
(b) Dropping Large Exporters

Notes: This figure reports estimates of the δ_g^{ex} 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.

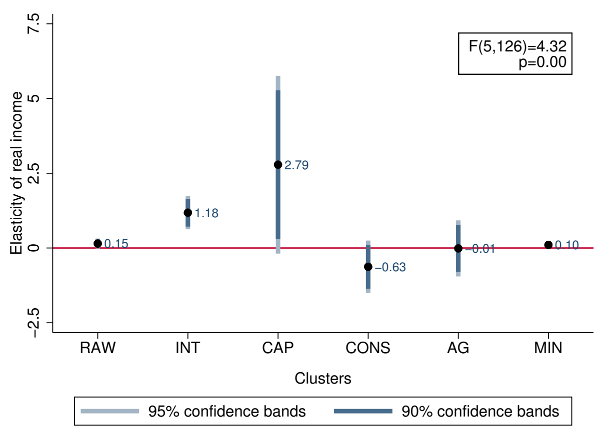
Figure A8: Cluster-Specific Coefficients and Confidence Intervals, Robustness to Spatial Correlation



(a) Dropping Contiguous Countries



(b) Controlling for Contiguous Countries' TFP

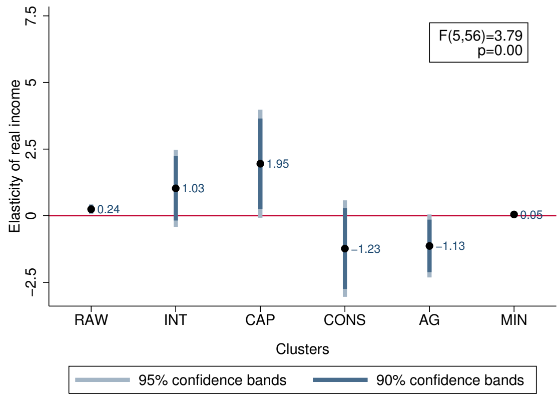


(c) Controlling for Unweighted $d \ln FMA_{igt}$

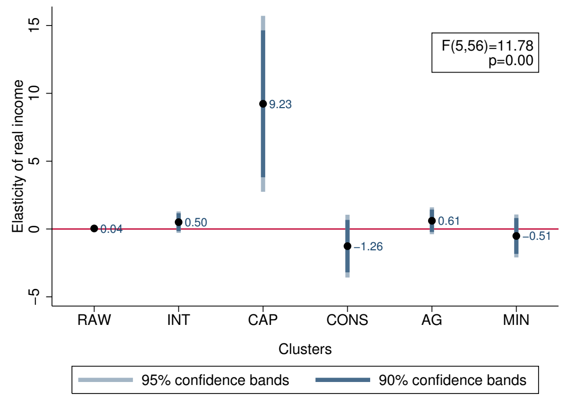
Notes: This figure reports estimates of the δ_g^{ex} coefficients in equation (3.6) via post double-LASSO. All specifications control for (i) time effects and (ii) log initial GDP per capita. The bars display the 90% and 95% confidence bands, that use standard errors clustered by country. The boxes display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

Figure A9: Developed vs. Developing Countries: Robustness to Decreased Tuning Parameter and 5 Manufacturing Clusters

I. Decreased Tuning Parameter

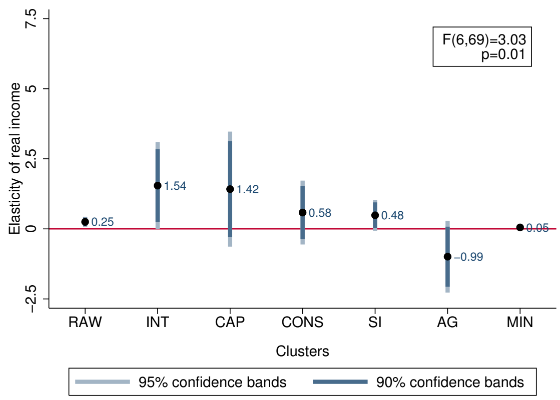


(a) Developed Countries

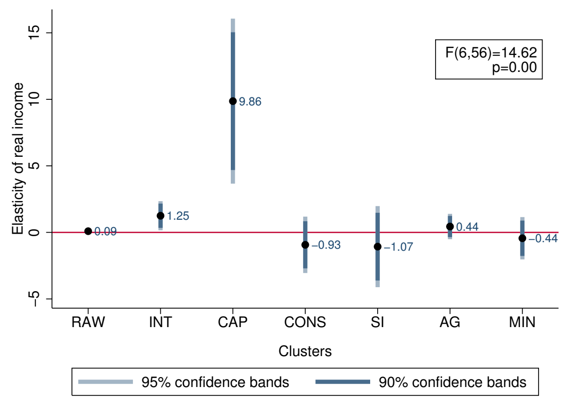


(b) Developing Countries

II. 5 Manufacturing Clusters



(c) Developed Countries

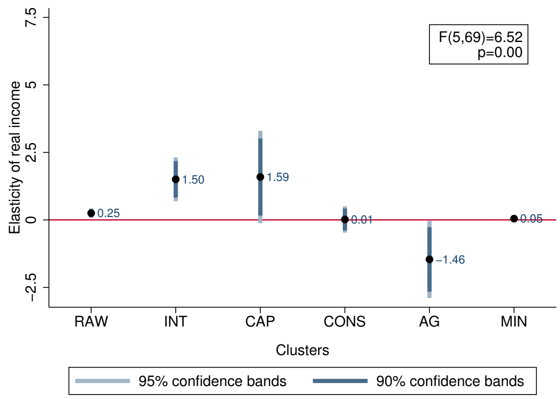


(d) Developing Countries

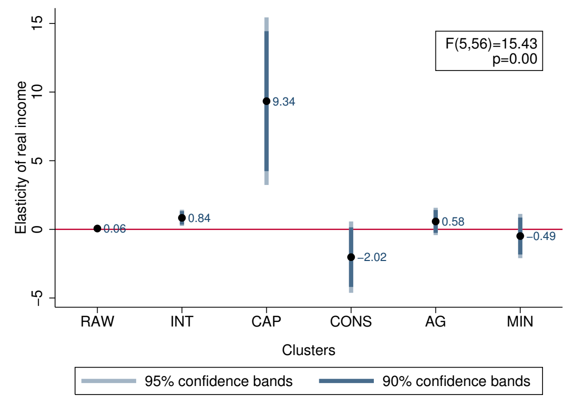
Notes: This figure reports estimates of the δ_g^{ex} coefficients in equation (3.6) via post double-LASSO, separately for developed (left side) and developing (right side) countries. All specifications control for (i) time effects and (ii) log initial GDP per capita. The bars display the 90% and 95% 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. The top panel reports the results with a decreased tuning parameter. For the sample of developed countries, 11 control variables are selected in the double-selection step. For developing countries, 6 control variables are selected in the double-selection step. The bottom panel reports the results when grouping the manufacturing industries to 5 clusters. For developed countries, 5 control variables are selected in the double-selection step. For developing countries, 0 control variables are selected in the double-selection step.

Figure A10: Developed vs. Developing Countries: Robustness to Small Country Assumption

I. Dropping Large Trading Partners

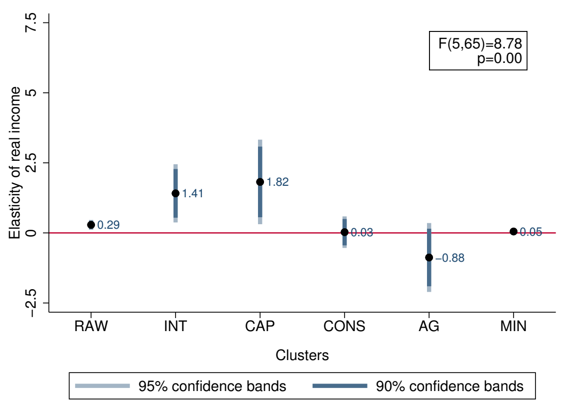


(a) Developed Countries

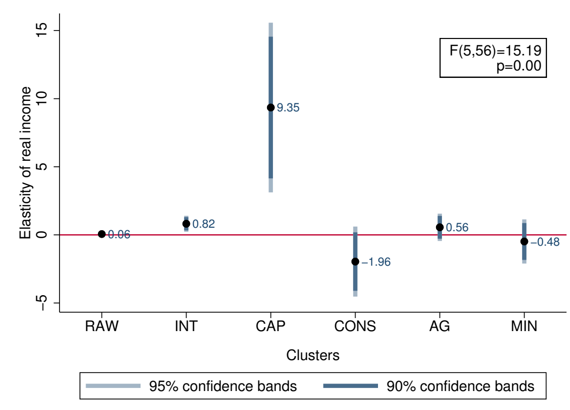


(b) Developing Countries

II. Dropping Large Exporters



(c) Developed Countries

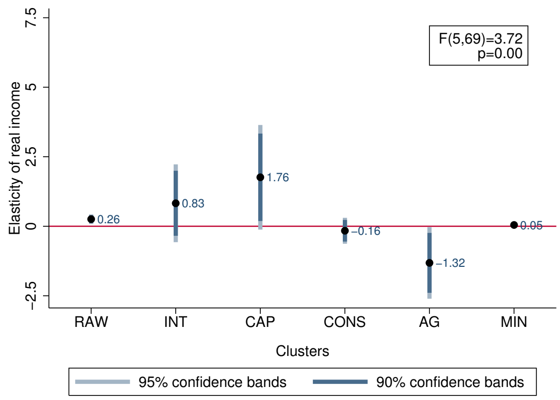


(d) Developing Countries

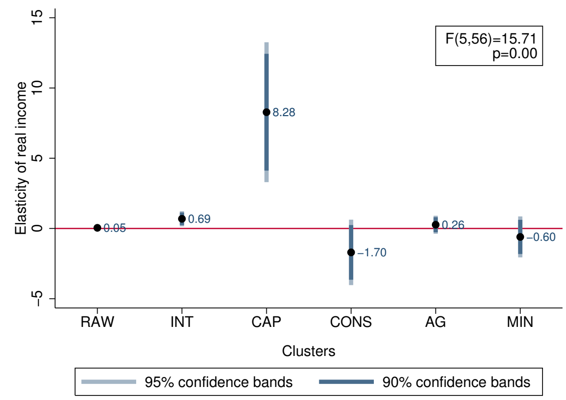
Notes: This figure reports estimates of the δ_g^{ex} coefficients in equation (3.6) via post double-LASSO, separately for developed (left side) and developing (right side) countries. All specifications control for (i) time effects and (ii) log initial GDP per capita. The bars display the 90% and 95% 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. The top panel reports the results while dropping large trading partners. For the sample of developed countries, 1 control variable is selected in the double-selection step. For developing countries, 0 control variables are selected in the double-selection step. The bottom panel drops from the estimation sample countries that represent the largest shares of world exports in any cluster. For both groups of countries, 0 control variables are selected in the double-selection step.

Figure A11: Developed vs. Developing Countries: Robustness to Spatial Correlation

I. Dropping Contiguous Countries

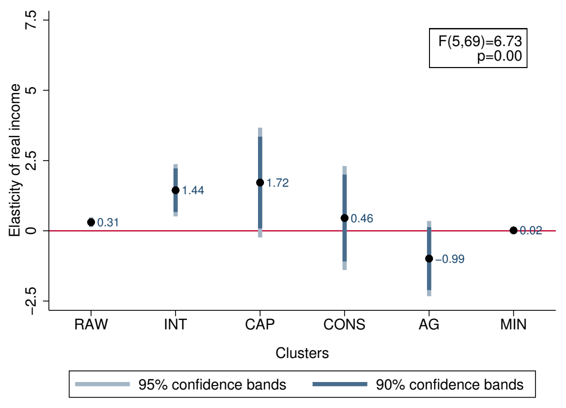


(a) Developed Countries

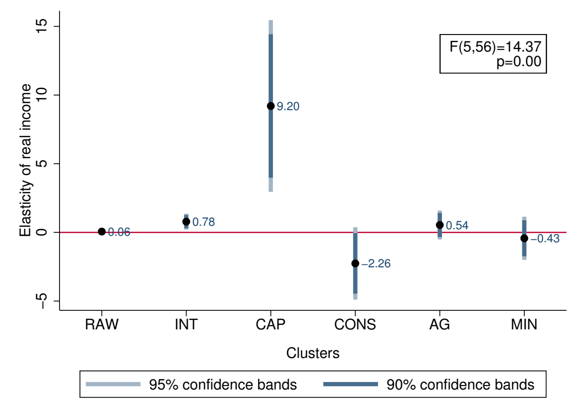


(b) Developing Countries

II. Controlling for Contiguous Countries' TFP

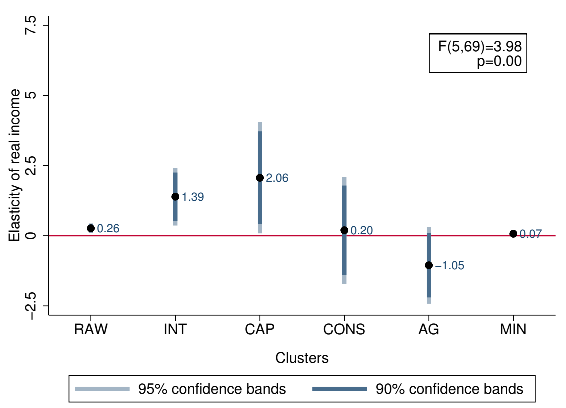


(c) Developed Countries

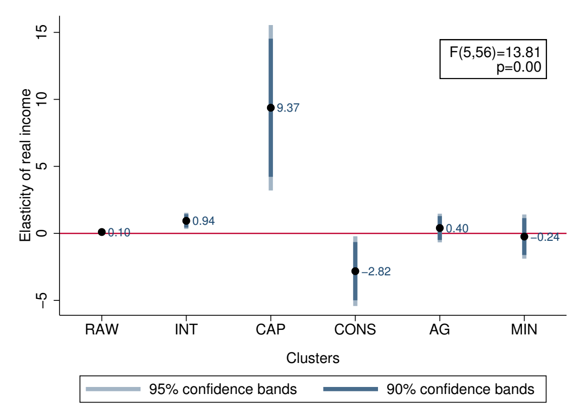


(d) Developing Countries

III. Controlling for Unweighted $d\ln FMA_{igt}$



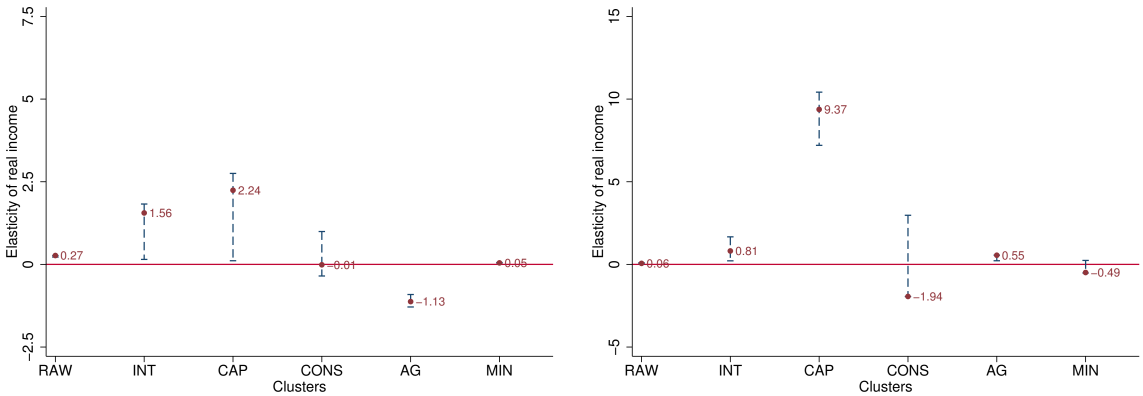
(e) Developed Countries



(f) Developing Countries

Notes: This figure reports estimates of the δ_g^{ex} coefficients in equation (3.6) via post double-LASSO, separately for developed (left side) and developing (right side) countries. 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. The top panel reports the results when dropping contiguous countries. For developed countries, 5 control variables are selected in the double-selection step. For developing countries, 0 control variables are selected in the double-selection step. The middle panel controls for contiguous countries' average TFP growth in the post-LASSO OLS. The bottom panel controls for the non-export-share weighted $d\ln FMA_{igt}$ in the post-LASSO OLS.

Figure A12: Developed vs. Developing Countries: Cluster Measurement Error Simulation

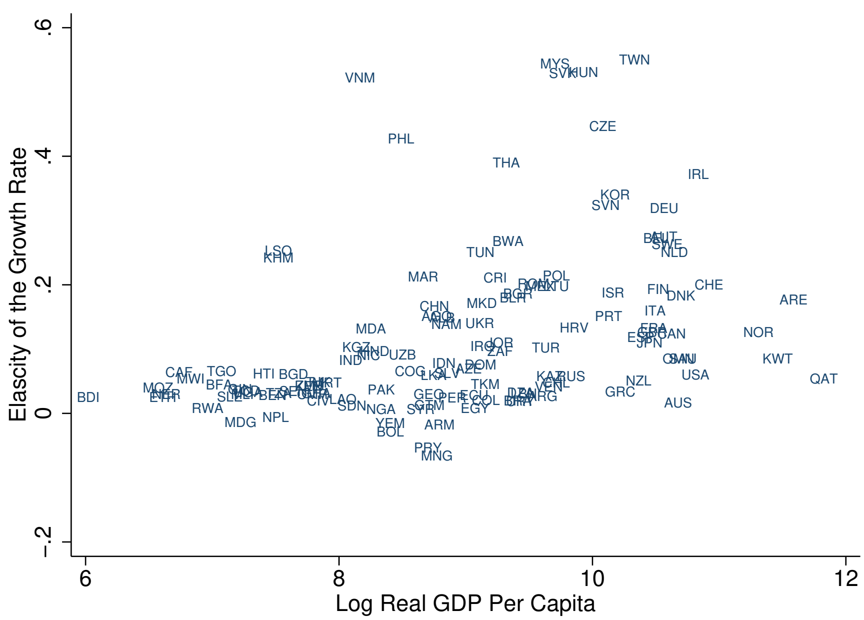


(a) Developed Countries

(b) Developing Countries

Notes: This figure reports estimates of the δ_g^{ex} coefficients in equation (3.6) via post double-LASSO in the measurement error simulations. 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. The right panel displays the results for developing countries. The vertical bars report the 95% range of coefficient estimates.

Figure A13: Elasticity of Real Income with Respect to a Uniform Foreign Shock, Developed vs. Developing Countries



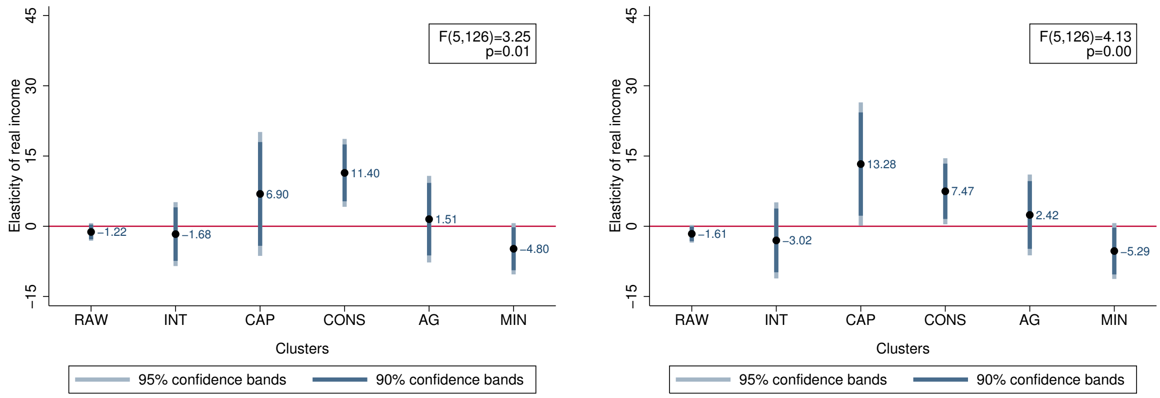
Notes: This figure presents the scatterplot of the elasticity of real income with respect to a uniform foreign demand shock (6.1) against real GDP per capita. It is calculated using the estimated coefficients in estimating equation (3.6) for developed and developing country subsamples separately, and the sectoral trade shares in 2015.

Table A7: Predicted Decennial Growth Difference Relative to Median Geographic Location, Medians by Region and Time Period

	Counterfactual Region						
Actual Region	East Asia & Pacific	Eastern Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	South Asia	Sub-Saharan Africa	West Europe/ North America
East Asia	0.43	0.91	0.54	-0.04	-0.28	0.17	1.87
& Pacific (N = 14)	*[-0.01 , 3.65]*	*[0.28 , 2.13]*	*[0.04 , 2.64]*	*[-2.50 , 1.51]*	*[-2.07 , 0.88]*	*[-0.41 , 2.86]*	*[0.40 , 3.78]*
Eastern Europe	-0.87	-0.25	-0.54	-1.12	-1.12	-1.38	-0.21
& Central Asia (N = 24)	*[-1.91 , -0.18]*	*[-0.88 , 0.19]*	*[-1.83 , -0.01]*	*[-1.78 , -0.56]*	*[-2.29 , -0.67]*	*[-2.20 , -0.72]*	*[-0.69 , 0.21]*
Latin America	-0.49	-0.04	-0.06	-0.53	-1.15	-0.51	0.16
& Caribbean (N = 18)	*[-1.55 , -0.04]*	*[-1.23 , 0.34]*	*[-0.44 , 0.52]*	*[-2.04 , -0.26]*	*[-1.80 , -0.49]*	*[-1.56 , -0.16]*	*[-1.00 , 0.44]*
Middle East	-0.02	0.78	0.92	0.31	-0.22	-0.17	1.07
& North Africa (N = 15)	*[-0.15 , 0.29]*	*[0.11 , 1.38]*	*[0.17 , 1.36]*	*[-0.17 , 0.77]*	*[-0.67 , -0.03]*	*[-0.52 , 0.22]*	*[0.28 , 1.96]*
South Asia (N = 5)	-0.09	0.11	0.29	-0.04	0.00	-0.01	0.37
	[-0.26 , 0.14]	*[0.06 , 0.19]*	*[0.20 , 0.54]*	*[-0.26 , -0.02]*	*[-0.49 , 0.20]*	*[-0.28 , 0.01]*	*[0.10 , 0.47]*
Sub-Saharan &	-0.11	0.43	0.36	-0.01	-0.15	-0.02	0.49
Africa (N = 33)	*[-0.25 , 0.52]*	*[0.16 , 1.03]*	*[0.10 , 0.72]*	*[-0.24 , 0.26]*	*[-0.62 , 0.29]*	*[-0.21 , 0.20]*	*[0.29 , 1.23]*
West Europe/	-1.20	0.01	-0.76	-1.28	-1.43	-1.29	0.13
North America (N = 18)	*[-2.75 , -0.23]*	*[-0.81 , 0.56]*	*[-2.37 , -0.55]*	*[-3.31 , -1.01]*	*[-3.42 , -0.68]*	*[-3.33 , -0.67]*	*[-1.02 , 0.63]*

Notes: This table reports the region pair-specific differences in economic growth, in percent per decade, between the actual growth and the counterfactual growth that the country would experience if it were moved to the counterfactual region. The numbers in square brackets are the interquartile range across countries in the region.

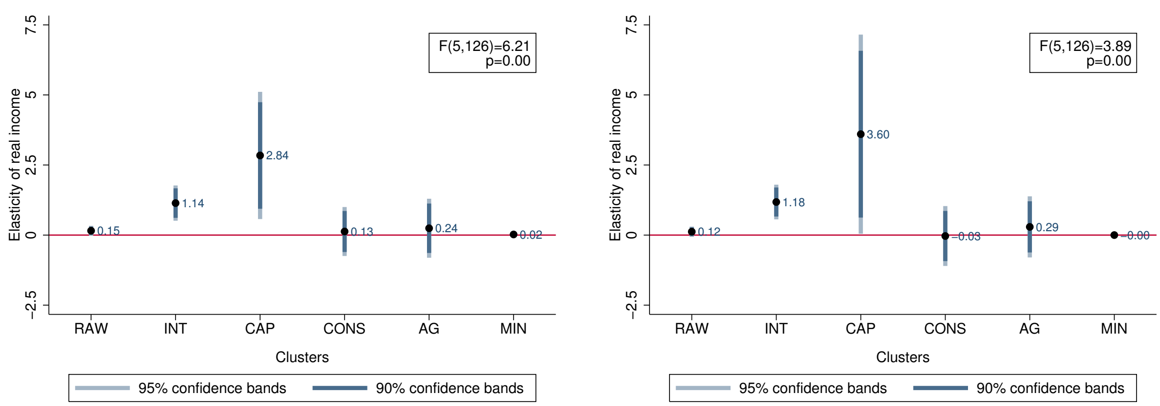
Figure A14: Foreign Supply: Cluster-Specific Coefficients and Confidence Intervals for CMA



(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 A15: 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.