Specialization, Market Access and Medium-Term Growth*

Dominick Bartelme
University of Michigan

Ting Lan
University of Michigan

Andrei Levchenko
University of Michigan
NBER and CEPR

January 2019

Abstract

This paper empirically estimates the impact of foreign demand and supply shocks to different sectors on medium-term economic growth. Our approach is based on a first order approximation to a wide class of small open economy models that feature sector-level gravity, which allows us to precisely measure foreign shocks and define their differential impact on growth in terms of reduced form elasticities. We use machine learning methods to cluster 4-digit manufacturing sectors into a smaller number of characteristics, and show that the cluster-level elasticities can be consistently estimated using high- dimensional statistical techniques. We find clear evidence of heterogeneity in the growth elasticities of different foreign shocks. Foreign demand shocks in complex intermediate and capital goods have large growth impacts, and both supply and demand shocks in capital goods have particularly large impacts on growth for poor countries. Counterfactual exercises show that both comparative advantage and geography have a quantitatively large growth impact.

JEL codes: F43, F62

Keywords: trade specialization, growth, gravity, k-means clustering

*We thank workshop participants at Berkeley and Monash for helpful suggestions. Email: dbartelm@umich.edu, tinglan@umich.edu, alev@umich.edu.

1 Introduction

The goal of this paper is to empirically estimate the effects of foreign demand and supply shocks on the medium-term growth rates of different countries. The question of how a country's external economic environment affects its development goes back to the classic work of Adam Smith and David Ricardo, who argued that changes in both the size of the external market and the sectoral composition of external demand and supply lead to changes in real income and therefore medium-term growth rates in the transition. Since these seminal contributions, a voluminous theoretical literature has elaborated on a number of mechanisms through which external demand and supply conditions, interacting with the structure of domestic comparative advantage, impact a country's medium-term growth rate. What has become clear is that the qualitative and quantitative impacts of foreign shocks depend crucially on the strength of the various mechanisms at play, and is therefore ultimately an empirical matter.

The empirical challenges in studying this question are formidable. There are many sectors and theories, but relatively few growth rates in the data. Econometric issues of endogeneity and omitted variable bias loom large. In the face of 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 size of the external market, as determined by geography (Frankel and Romer, 1999; Redding and Venables, 2004; Feyrer, 2019). Another examines whether certain feature of comparative advantage are associated with growth (e.g. Prebisch, 1959; Humphreys et al., eds, 2007; Hausmann et al., 2007). This approach abstracts from cross-country variation in external demand and supply, and tends to suffer from weak theoretical foundations and the econometric problems endemic to cross-country regressions. The third calibrates fully specified general equilibrium models and conducts counterfactuals (e.g. Whalley, 1985; Aguiar et al., 2016; Hsieh and Ossa, 2016). These methods deliver precise and interpretable answers, but depend heavily on the assumed model structure and a large number of parameters.

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 structural models 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) bilateral trade obeys sector-level gravity, 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 various frameworks featuring monopolistic competition.

This framework delivers natural measures of sector-level foreign demand and supply shocks, which we label external firm and consumer market access respectively. These variables contain all relevant information for Home's interaction with foreign markets, and are easily estimated from the trade data using standard techniques. We employ a first order approximation to express a country's real income growth in terms of export and import share-weighted averages of the foreign shocks, along with domestic demand and supply shocks. The elasticities on the foreign variables measure how different foreign shocks, interacted with the domestic sectoral composition, generate different general equilibrium growth impacts, thus providing a direct answer to the question posed by this paper. These elasticities also map directly to relevant parameters for trade policy.

Estimation must confront the "many sectors, few GDP growth rates" problem as well as the econometric issues of omitted variables and endogeneity. We first pool observations across countries and time, and employ a machine learning algorithm to cluster sectors based on their characteristics. The coefficients on the cluster-level variables are the average within-cluster elasticities. We then provide formal conditions under which these average effects are identified by an OLS regression that fully conditions on the initial equilibrium observables, which exploits the typical invertibility properties of gravity models. To deal with the high dimensionality of the control vector we employ the Post-Double-Selection method of Belloni et al. (2014b, 2017), which relies on the approximate sparsity of the control vector to select a lower-dimensional set of "important" controls while maintaining consistency and uniformly valid inference. We rely on the fact that most countries are small in foreign markets to eliminate any direct causal relationship between domestic and foreign shocks, and measure the foreign shocks in such a way as to minimize the practical relevance of this channel.

We implement our approach on UN COMTRADE trade data and decadal real income growth rates from the P.W.T. 9.0 over 1965-2015, with a sample of 127 countries and 268 sectors. We use the k-means clustering algorithm (MacQueen et al., 1967) along with 7 sectoral characteristics measured from U.S. data to cluster 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) producer non-durables (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 12 cluster-level foreign shocks.

We find significant heterogeneity in the average growth impact of different foreign shocks across clusters. Foreign demand shocks in complex intermediate and capital goods producing sectors have the largest growth impacts, with the capital goods elasticity being somewhat imprecisely estimated. Foreign demand shocks in all other sectors have small and positive growth impacts. Turning to the supply shocks, we find that the largest growth impacts come from the capital and consumer goods sectors, although the confidence intervals are rather large. This finding reflects in part the lack of variation across countries in the foreign supply shocks relative to the demand shocks.

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 growth elasticities and non-intermediate, non-capital goods sectors have small elasticities. The result that both types of foreign shocks in capital goods sectors have high growth elasticities is moderately robust. Interestingly, when we split the sample into developed and developing countries, we find that both capital goods elasticities are much higher (and relatively precisely estimated) for developing countries across all specifications. While intriguing, the practical importance of this finding on the demand side is limited by the low shares of these goods in the export baskets of developing countries.

We conclude by examining the quantitative implications of our estimates. Given our estimated elasticities, the growth impacts are determined by the size and pattern of foreign shocks ("geography") interacted with the trade shares ("comparative advantage"). Our first exercise holds geography constant and computes the total elasticity of growth with respect to uniform foreign demand and supply shocks for each country in our sample. There is substantial cross-country heterogeneity in the impacts, with rich countries benefiting more from foreign demand shocks on average due to their higher propensity to specialize in high growth-elasticity sectors. Our second exercise illustrates the role of geography by holding comparative advantage constant and subjecting each country to the foreign shocks experienced by different countries in the same time period. We find that geography plays an non-trivial role in determining the growth experiences of different countries. For example, East Asian countries benefited to the tune of roughly half a percentage point of growth per year (relative to the median country) over the sample period from the rapid growth of surrounding countries, while Western European countries lost roughly half a percentage point of growth due to slow overall growth in the region.

Our paper contributes to the literature on trade and growth. A number of influential papers estimate the impact of overall openness on growth (e.g. Frankel and Romer, 1999; Rodriguez and Rodrik, 2001; Feyrer, 2019; Redding and Venables, 2004). Our paper is closer to the literature on trade patterns and growth. Most of this literature studies either export or import patterns, but not both, and considers only one characteristic of trade patterns at a time. Some examples on export side include the natural resource curse literature (e.g. Humphreys et al., eds, 2007), the work on "high-income goods" (Hausmann et al., 2007), the location in the product space (Hidalgo et al., 2007), specialization in primary goods (Prebisch, 1959) or skill-intensity (Blanchard and Olney, 2017; Atkin, 2016). The literature also considered imports of capital goods (Eaton and Kortum, 2001; Caselli and Wilson, 2004), skill-intensive goods (Nunn and Trefler, 2010; Atkin, 2016; Blanchard and Olney, 2017), or intermediate inputs (e.g. (e.g. Amiti and Konings, 2007; Kasahara and Rodrigue, 2008). On the theory side, our framework is related to recent work using partially specified general equilibrium models to conduct trade counterfactuals (Adao et al., 2017; Allen et al., 2019; Bartelme, 2018).

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 Model

2.1 Economic Environment

We consider the steady state of a small open economy Home (H) in a world with N other countries (indexed by n) and K sectors indexed by k. 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 (the Armington assumption). For simplicity we assume that each sector in Home produces a homogeneous good.

Technology and Market Structure

There are J factors of production, indexed by j, that are in fixed supply $\bar{L}_{H,j}$ 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 that operate both within and across sectors. Given these assumptions, we can characterize the production technology in each sector by the unit cost function $c_k(\{w_{H,j}\}, \{P_{H,k}\}, \{L_{H,jk}\}, \{T_{H,k}\})$, where $\{w_{H,j}\}$ are factor prices, $\{P_{H,k}\}$ are intermediate goods prices, $\{L_{H,jk}\}$ are the factor allocations and $\{T_{H,k}\}$ are exogenous productivities. We allow this cost function to be very general, requiring only that it is continuously differentiable. Note that we allow for cross-sectoral productivity spillovers in that the allocation of factors to other sectors may affect the unit costs in sector j.

Demand

All factor income accrues to a representative consumer. Consumers have homothetic preferences over sectoral quantity bundles $Q^C_{H,k}$.[1] Within sectors, consumers combine Home and foreign varieties in a CES fashion,

$$Q^C_{H,k} = \left(z_{H,k}^{\frac{1}{\sigma_k}} \cdot (q^C_{H,k})^{\frac{\sigma_k - 1}{\sigma_k}} + \sum_{n \in N} (q^C_{nH,k})^{\frac{\sigma_k - 1}{\sigma_k}} \right)^{\frac{\sigma_k}{\sigma_k - 1}} \tag{1}$$

where $z_{H,k}$ is an exogenous demand shifter. This formulation allows consumers to have home bias in consumption, so that Home products can potentially have large market share in the Home market. We assume that producers use the same aggregator for intermediate goods. We denote the sectoral CES price indices as $P_{H,k}$ and the aggregate price index as $\mathbb{P}_H$.

These assumptions on the lower tier demand functions imply a sector-level gravity

[1] We assume homotheticity in order to equate welfare with real income via a well-defined aggregate price index, which in turn allows us to make contact with national accounts data in the empirical section.

equation for expenditure shares on goods from various sources. Foreign prices have two components: the source-specific costs and an iceberg bilateral component $\tau_{nH,k}$. With these assumptions, we can write the gravity equation as

$$p_{nH,k} \cdot q_{nH,k} = \frac{(c_{n,k} \cdot \tau_{nH,k})^{1-\sigma_k}}{P_{H,k}^{1-\sigma}} \cdot E_{H,k}, \;\; p_{HH,k} \cdot q_{HH,k} = z_{H,k} \frac{c_{H,k}^{1-\sigma_k}}{P_{H,k}^{1-\sigma}} \cdot E_{H,k} \tag{2}$$

where $P_{H,k}^{1-\sigma} = z_{H,k} c_{H,k}^{1-\sigma_k} + \sum_{n \in N} (c_{n,k} \tau_{nH,k})^{1-\sigma_k}$ and $E_{n,k}$ is Home sectoral expenditure on both consumption and intermediate goods. Foreign demand for Home's commodities also takes the gravity form, with foreign imports facing some iceberg bilateral trade barriers $\tau_{Hn,k}$,

$$p_{Hn,k} \cdot q_{Hn,k} = (c_{H,k} \cdot \tau_{Hn,k})^{1-\sigma_k} \cdot \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}}. \tag{3}$$

We now define two key quantities. 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),[2]

$$\sum_{n \in N} p_{Hn,k} q_{Hn,k} = c_{H,k}^{1-\sigma_k} \cdot \underbrace{\sum_{n \in N} \tau_{Hn,k}^{1-\sigma_k} \cdot \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}}}_{FMA_{H,k}} \tag{4}$$

Likewise, summing import expenditures across foreign sources, we get total imports as a function of Home expenditures, prices and external *Consumer Market Access* (CMA),

$$\sum_{n \in N} p_{nH,k} \cdot q_{nH,k} = \frac{E_{H,k}}{P_{H,k}^{1-\sigma_k}} \cdot \underbrace{\sum_{n \in N} (c_{n,k} \cdot \tau_{nH,k})^{1-\sigma_k}}_{CMA_{H,k}} \tag{5}$$

From Home's perspective, both external firm and consumer market access are exogenous. Moreover, they are sufficient statistics for Home's interaction with foreign markets. Any change in foreign variables affects the Home equilibrium only through their

[2] This concept differs from the usual definition of market access in that it excludes the contribution of domestic demand.

effects on FMA and CMA.

Competitive Equilibrium

We define a competitive equilibrium in the usual way, as a set of goods and factor prices and allocations such that firms and consumers maximize taking prices as given, factor and output markets clear and trade balances. Under the assumptions above, 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). If factor allocations are uniquely determined given factor prices, we can further reduce the system to a set of J simultaneous equations in factor prices, equating factor supply with factor demand. Regardless of uniqueness, the set of equilibria is completely determined by the functions $c_{H,k}$ and $U(Q_{H,k}^C)$, the elasticities σ_k and the exogenous variables.

Our first order approach to estimation and counterfactual welfare analysis requires a unique and smooth mapping from the exogeneous variables to equilibrium outcomes. Without uniqueness, the data would contain little or no information on how different foreign shocks systematically affect real income growth.[3] In general, without further restrictions on $c_{H,k}$ and $U(Q_{H,k}^C)$ there may be multiple equilibria, with the presence of external economies being the primary culprit. It is easy to construct examples of multiple equilibria, but difficult to characterize the conditions under which there is a unique equilibrium even in relatively simple settings (Kucheryavyy et al., 2016).[4] This is not to say that uniqueness rarely obtains, simply that it is difficult to provide sufficient conditions. Hence we do not pursue a characterization of the equilibrium properties of this class of models. Instead, we simply assume a unique and smooth equilibrium function in the relevant parameter space for the rest of the paper.

[3]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.

[4]Kucheryavyy et al. (2016) studies a setting with full international general equilibrium. Intuitively, uniqueness should obtain under a wider set of assumptions in an SOE because of the lack of international feedback.

2.2 First Order Welfare Approximation

We now drop the H subscript to economize on notation. Our assumption of homothetic preferences equates real expenditure with welfare, while our assumption of trade balance equates nominal GDP with nominal expenditure. Thus we can write Home's welfare as

$$\frac{Y}{\mathbb{P}} = \alpha \cdot \frac{\sum_{k \in K} c_k^{1-\sigma_k} \cdot \left(\frac{E_k}{P_k^{1-\sigma_k}} + FMA_{H,k} \right)}{\mathbb{P}} \tag{6}$$

where Y is nominal GDP and α is the share of value added in gross output. The term in the numerator of the RHS is total sales, domestic and foreign. External consumer market access enters into this expression implicitly through the sectoral price indices $P_k \equiv (z_{H,k} c_{H,k}^{1-\sigma_k} + CMA_k)^{\frac{1}{1-\sigma_k}}$.

External shocks will have two types of effects on Home's welfare in a competitive equilibrium. There will be direct effects through increased foreign sales (when FMA_k increases) and lower prices (when CMA_k increases). There will also be indirect effects as domestic producers and factor owners alter their prices and production plans and consumers alter their consumption patterns in response to these external shocks.

Our interest is in capturing the total effects of foreign shocks, both direct and indirect, in an empirical setting. To do so we make use of our assumption of a unique and smooth mapping from the domestic and foreign shocks to equilibrium quantities. Taking natural logs of Equation (6) and applying Taylor's theorem, the log change in real income with respect to a set of log changes in foreign and domestic shocks is approximately[5]

$$d \ln y \approx \sum_k \delta_k^{ex} \cdot \left[\lambda_k^{ex} d \ln FMA_k\right] + \sum_k \delta_k^{im} \cdot \left[\lambda_k^{im} d \ln CMA_k\right] + \sum_k \delta_k^{T} \cdot d \ln T_k, \tag{7}$$

where $y \equiv Y/\mathbb{P}$ denotes real income or welfare, λ_k^{ex} is the share of total sales accounted for by exports in sector k and λ_k^{im} is the share of total expenditures accounted for by

[5] For expositional purposes we assume that neither the factor supplies nor the domestic demand shifters change. It is straightforward but notationally cumbersome to add these terms.

imports in sector k.

The elasticities δ_k^{ex} and δ_k^{im} measure the total impact, direct and indirect, of foreign shocks in different industries on real income growth. To interpret these elasticities, consider the following natural experiment. Two small open economies, initially identical in every respect, experience a different pattern of foreign shocks. Specifically, suppose economy A experiences a 1% increase in foreign demand in industry 1 while economy B experiences a 1% increase in foreign demand in industry 2. Which economy will experience greater real income growth? Assuming both industries have the same initial export sales shares, the answer will be whichever economy gets the shock to the industry with the highest δ^{ex}. By focusing on external demand and supply shocks, rather than realized trade as in much of the literature, we can in principle separate the casual growth impact of external factors from that of domestic productivity or demand shifters.

These elasticities are not generally "structural" parameters, except in special cases. However, they do map to parameters that are relevant for trade policy. Imagine that an exporting sector faces an *ad valorem* non-revenue export barriers t_k^{ex} on all destinations. Using the definition of FMA, we have that δ_k^{ex} is the trade elasticity times the elasticity of real GDP with respect to τ_k^{ex}, or

$$\delta_k^{ex} = \frac{1}{1-\sigma_k} \cdot \frac{\partial \ln y}{\partial \ln \tau_k^{ex}}. \tag{8}$$

The same relationship applies to import trade barriers and δ_k^{im}.[6] Thus δ_k^{ex} and δ_k^{im} are sufficient statistics for the impact of non-revenue trade policy on real GDP, modulo the trade elasticity.

The determinants of δ_k^{ex} and δ_k^{im} are complex and difficult to characterize analytically in a general setting. Foreign shocks in different sectors generate different terms of trade effects, which in turn trigger different patterns of reallocation across sectors. These ini-

[6] If the trade barrier is an import tariff applied by a trading partner, then the formula for the export elasticity applies with $1/(1-\sigma_k)$ replaced by $1/\sigma_k$, while the formula for the import elasticity is unchanged in the case of a trading partner's export tax. Revenue-generating trade taxes applied by the Home country imply complex tariff revenue effects. In this case, the formula using $1/\sigma_k$ is approximately correct for real GDP (not real expenditure) at low tariff levels.

tial reallocations in turn generate factor price and productivity movements that imply further rounds of reallocation. These effects are especially complicated when sectors are linked through input-output relationships or productivity spillovers. Below we offer several simple examples to give some intuition for how the underlying structure of the economy determines the elasticities in different scenarios.

This very complexity provides one of the primary motivations for our approach. Rather than explicitly modeling and quantifying each aspect of the underlying structure of the economy, we aim to empirically recover the reduced form elasticities that are directly relevant to the relationship between trade and growth. Our estimates will thus be robust to model uncertainty within the wide class of trade models encompassed by our framework, which offers a clear advantage over methods that require a complete specification of the model. On the other hand, we provide enough structure to enable clear interpretation, provide precise conditions for identification, and conduct local counterfactuals. These elements are missing in the reduced form literature.

There are also some costs to achieving this robustness to model uncertainty. First, fully specifying a (correct) model permits more efficient estimation of the relevant parameters. Second, a fully structural model reveals the economic mechanisms that generate the results more clearly. Third, a structural model can be solved in its non-linear form, which enables more accurate counterfactuals with respect to large shocks. We thus view our strategy as complementary to fully structural approaches.

2.3 Examples

Efficient Economy

We consider the solution to the planner's problem in the general setting. We assume that the planner directly chooses quantities and factor allocations to maximize welfare. In Appendix A we show that an application of the Envelope Theorem gives

$$\delta_k^{ex} = \frac{1}{\sigma_k}, \;\; \delta_k^{im} = \frac{1}{\sigma_k - 1}, \;\; \forall k \in K. \tag{9}$$

Interestingly, a naive application of Hulten's Theorem to this economy fails in that the effects of foreign shocks are not simply proportional to the foreign sales and expenditure shares. An intuition for the export elasticity comes from the fact that the optimal export tax on industry k is $1/\sigma_k$. This implies that the country earns high margins on exports from industry k, relative to another industry with the same export sales but higher σ_k. Given equal initial sales, the planner prefers a proportional increase in sales in the high margin (low σ_k) industry. The intuition for the import elasticity is a bit different: the factor $\frac{1}{1-\sigma_k}$ simply translates the increase in market access into a decrease in prices.

Single Factor Economy with No Spillovers

We now specialize our general setting to the competitive equilibrium of a single factor economy with no intermediate goods and no external economies of scale. We also assume that the upper tier utility is Cobb-Douglas. Given these assumptions,

$$\delta_k^{ex} = \kappa, \;\; \delta_k^{im} = \left(\frac{1}{\sigma_k - 1} - \kappa\bar{\theta}_k\right), \;\; \forall k \in K, \tag{10}$$
$$\kappa = \frac{\sum_{k\in K}\lambda_k^{im}}{\sum_{j\in K} 1 - \left[\sum_{j\in K}\lambda_k^d + (1-\sigma_k)(\lambda_j^{ex} + (1-\bar{\theta}_k)\lambda_k^d)\right]}$$

where λ_k^d is the initial share of domestic sales in industry k in total sales, and $\bar{\theta}_k = \frac{\lambda_k^d}{\lambda_k^d + \lambda_k^{im}}$. Unlike the case of an efficient economy, here the export elasticity is constant across industries. This is because in a single factor economy without spillovers, labor allocations to exports in each industry are proportional to the export sales share λ_k^{ex}. This implies that the indirect effect (through the wage) of a shock to $\ln FMA_k$ is proportional to the export share; since the direct effect is also proportional to the export share the overall effect is proportional as well. The constant of proportionality κ reflects the overall importance of trade to the economy as well as the distribution of sales across foreign and domestic customers and their covariance with the trade elasticities. The import elasticity is modified (relative to the efficient case) to account for the negative impact of foreign competition on domestic producers.

Single Factor Economy with Industry Spillovers

We now augment the single factor economy above with endogeneous within-industry productivity spillovers as in Kucheryavyy et al. (2016) and Bartelme et al. (2018), so that $c_k = \frac{w}{T_k L_k^{\gamma_k}}$. To simplify the analysis we assume a Cobb-Douglas upper tier and zero domestic sales, as well as the condition $\gamma_k(\sigma_k - 1) < 1, \ \forall k$ to ensure a unique interior equilibrium. The elasticities are now given by

$$\delta_k^{ex} = \kappa \cdot \frac{1}{1-\gamma_k(\sigma_k - 1)}, \ \ \delta_k^{im} = \left(\frac{1}{\sigma_k - 1}\right), \ \ \forall k \in K, \tag{11}$$
$$\kappa = \frac{1}{\sum_{j\in K} 1 - \left[\sum_{j\in K} \frac{(1+\gamma_k)(1-\sigma_k)}{1-\gamma_k(\sigma_k-1)})\lambda_j^{ex}\right]}.$$

All else equal, foreign demand shocks in sectors with larger productivity spillovers generate higher income growth. Notice that, for a given γ_k, higher σ_k also implies a higher growth elasticity. This reflects the fact that scale economies are more valuable in sectors with more elastic international demand; in less elastic sectors, achieving higher productivity comes at the expense of significantly lower export prices.[7]

2.4 Isomorphisms and Extensions

We have derived our results using the competitive equilbrium of an Armington economy to maximize clarity and simplicity. However, the crucial assumptions are the gravity assumption on trade flows, homothetic upper tier preferences and the unique equilibrium mapping that validates our first order approach. Thus 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 iso-

[7]Our assumption of zero domestic sales implies that foreign supply shocks do not affect domestic prices or production decisions. With positive domestic sales the formulas become quite messy, but the general intuition is still that countries prefer demand shocks in high $\gamma_k(\sigma_k - 1)$ sectors and prefer supply shocks in low $\gamma_k(\sigma_k - 1)$ sectors.

morphic to our model in the sense that they have a first order approximation of the same form as Equation (7) and the same interpretation of the market access elasticities.

Our framework is static, and thus should be interpreted as capturing long run differences across steady states. Our assumption of fixed factor endowments formally rules out dynamic models of factor accumulation, but we can extend our approach to allow for this feature as well by letting the steady state factor supplies depend on the other exogenous variables of the model through long run factor supply equations.

3 Identification and Estimation

3.1 Identification

We now consider identification of the elasticities δ_k^{ex} and δ_k^{im} based on Equation (7). To match our empirical setting, we consider a world populated by many small open economies (indexed by i) over many time periods (indexed by t), with a fixed set of industries indexed by k. The log change in real income in country i between time t and $t+1$ is approximately

$$d\ln y_{i,t} \approx \sum_k \delta_{ik,t}^{ex} \cdot \left[\lambda_{ik,t}^{ex} d\ln FMA_{ik,t}\right] + \sum_k \delta_{ik,t}^{im} \cdot \left[\lambda_{ik,t}^{im} d\ln CMA_{ik,t}\right] + \sum_k \delta_{ik,t}^{T} \cdot d\ln T_{ik,t} \quad (12)$$

where $d\ln x_{ik,t} = \ln x_{ik,t+1} - \ln x_{ik,t}$ for $x = FMA, CMA, T$.

The variables $d\ln FMA_{ik,t}$, $d\ln CMA_{ik,t}$ and $d\ln T_{ik,t}$ on the right hand side of this equation are not directly observable. However, $FMA_{ik,t}$ and $CMA_{ik,t}$ can be consistently estimated using conventional gravity equation techniques (Head and Mayer, 2014). We defer a detailed discussion of our estimation strategy for these variables to Section 4, assuming that they are known with certainty for the remainder of this section. In contrast, the domestic productivity shocks $T_{ik,t}$ cannot be observed or estimated without knowledge of the full model structure. We treat the domestic shocks as unobservable, which leads to the empirical specification

$$d\ln y_{i,t} = \nu_t + \sum_k \delta^{ex}_{ik,t} \cdot \left[\lambda^{ex}_{ik,t} d\ln FMA_{ik,t}\right] + \sum_k \delta^{im}_{ik,t} \cdot \left[\lambda^{im}_{ik,t} d\ln CMA_{ik,t}\right] + \epsilon_{i,t}, \tag{13}$$

where ν_t is the mean time-t domestic shock term, $\epsilon_{i,t} = \sum_k \delta^T_{ik,t} \cdot d\ln T_{ik,t} - \nu_t$.[8]

Equation (13) has a larger number of parameters ($2K \times N \times T$) than observations ($N \times T$). In some simple examples the elasticities depend only on industry characteristics, but in general they also depend on the initial equilibrium (the point of approximation) and are thus country and time-specific as well. This issue is compounded by the fact that we observe a large number of distinct traded industries relative to the number of long run country-time growth rates in the sample, making even the estimation of industry-specific elasticities problematic in our finite sample.

We begin by clustering "similar" industries together, where similarity is defined as closeness in the space of industry characteristics. We measure a number of industry characteristics that are likely to affect the elasticities, then cluster the industries using the k-means algorithm commonly used in machine learning and statistics. We describe the clustering algorithm and the industry characteristics below in Section 4. For now, we simply assume that we have arrived at some clustering scheme $g \in G$. Using this notation, we can rewrite Equation (13) as

$$d\ln y_{i,t} = \nu_t + \sum_{g\in G} \delta^{ex}_g \cdot [d\ln FMA_{ig,t}] + \sum_{g\in G} \delta^{im}_g \cdot [d\ln CMA_{ig,t}] + \mu_{i,t} + \epsilon_{i,t}, \tag{14}$$

[8]Again, to simplify the notation we are ignoring the possibility of changes in domestic factor supplies and demand shocks $d\ln z_{ik,t}$. All our results regarding identification in the presence of unobserved productivity shocks apply to these variables as well.

where

$$\delta_g^{ex} = \frac{1}{K_g}\sum_{k\in g} E_{i,t}[\delta_{ik,t}^{ex}], \;\; d\ln FMA_{ig,t} = \sum_{k\in G}\lambda_{ik,t}^{ex} d\ln FMA_{ik,t}, \tag{15}$$

$$\mu_{i,t} = \sum_{g\in G}\sum_{k\in g}(\delta_{ik,t}^{ex} - \delta_g^{ex})\left[\lambda_{ik,t}^{ex} d\ln FMA_{ik,t}\right] + \sum_{g\in G}\sum_{k\in g}(\delta_{ik,t}^{im} - \delta_g^{im})\left[\lambda_{ik,t}^{im} d\ln CMA_{ik,t}\right], \tag{16}$$

K_g is the number of industries in g and similar definitions apply to δ_g^{im} and $d\ln CMA_{ig,t}$. The parameters of interest are the δ_gs, which are the within-cluster average of the average partial effects $E_{i,t}[\delta_{ik,t}]$. They can be interpreted as the best guess for the growth impact of a unit shock to log market access in industry $k \in q$ for a randomly chosen country and time period, conditional only on the identity of the cluster.

Identification requires the conditional independence of the foreign shocks and the two error components, $\mu_{i,t}$ and $\epsilon_{i,t}$. As it stands, Equation (14) does not satisfy this condition, since both the foreign shocks and the error components depend on the initial equilibrium. The foreign shocks are obviously functions of the initial equilibrium, via the trade share weights $\lambda_{ik,t}^{ex}$ and $\lambda_{ik,t}^{im}$. Less obviously, the error components $\mu_{i,t}$ and $\epsilon_{i,t}$ are also functions of the initial equilibrium. This dependence stems from several sources, primarily the dependence of the country-industry-time-specific elasticities $\delta_{ik,t}$ on the initial equilibrium and any serial correlation in the domestic shocks $d\ln T_{ik,t}$. Intuitively, the identification challenge is to ensure that "all else is equal" across countries receiving different "treatments," i.e. different patterns of foreign shocks. Note that the large number of potential channels for correlation between the errors and the independent variables makes it impossible to sign the bias that would arise from estimating Equation (14) using OLS.

This discussion suggests that we could identify the cluster-level average treatment effects if we condition on all relevant information on the initial equilibrium. We exploit the structure of the gravity to rigorously show how we can do so. Recall from Section 2 that we assume the existence of a smooth and one-to-one equilibrium map which determines every endogenous variable, including the $\delta_{ik,t}$, as a function of the set of exogenous variables $\{\{T_{ik,t}\}, \{z_{ik,t}\}, \{FMA_{ik,t}\}, \{CMA_{ik,t}\}, \{\bar{L}_{ij,t}\}\}$. In principle, the $FMA_{ik,t}$, $CMA_{ik,t}$ and $\bar{L}_{ij,t}$ are all observable while the domestic supply and demand shifters $T_{ik,t}$ and $z_{ik,t}$ are not. However, gravity models of trade typically have the property that, conditional on the rest of the exogenous variables and the parameters of the model, the trade flows $\lambda^{ex}_{ik,t} \cdot Y_{i,t}$ and $\lambda^{im}_{ik,t} \cdot E_{i,t}$ can be inverted to recover the $T_{ik,t}$ and $z_{ik,t}$ that generated them. We assume that the underlying model has this property as well, which allows us to characterize any variable in the initial equilibrium as functions of observables. Once we condition on the initial equilibrium via these observables, identification of the elasticities follows, provided that the residual innovations in domestic productivity and demand are uncorrelated with the foreign shocks. Our small open economy assumption makes this identification condition internally consistent with our model in the sense that there can be no direct causal relationship between the domestic and foreign shocks, and thus it involves only restrictions on the joint distribution of the exogenous variables.[9]

We now provide formal sufficient conditions for identification for two special cases of the general model in Equation (14), then discuss the general case. Our discussion assumes that the mapping from the initial equilibrium observables to the unobservables is sufficiently smooth to be well approximated by linear combinations of functions initial observables, such as dummies, polynomials, splines, and interactions. We denote the (potentially high dimensional) vector of approximating variables by $w_{i,t}$, and WLOG assume that each component has mean zero.

[9] In a large economy, domestic shocks will affect foreign variables. We measure our foreign shocks so as to minimize the effect of any violations of this assumption in the data, and conduct robustness checks with respect to this assumption in Section 5.

Constant Treatment Effects Within Clusters

In this case, the elasticities are constant within cluster, i.e. $\delta^{ex}_{ik,t} = \delta^{ex}_{g}$ and $\delta^{im}_{ik,t} = \delta^{im}_{g}$. Under this assumption, we can write Equation (14) as

$$d\ln y_{i,t} = \nu_t + \sum_{g\in G} \delta^{ex}_{g} \cdot [d\ln FMA_{ig,t}] + \sum_{g\in G} \delta^{im}_{g} \cdot [d\ln CMA_{ig,t}] + \gamma w_{i,t} + \tilde{\epsilon}_{i,t}, \tag{17}$$
$$\tilde{\epsilon}_{i,t} = \sum_{k\in K} \gamma_k w_{i,t} \cdot \xi^{T}_{ik,t}, \quad E[\tilde{\epsilon}_{i,t}] = 0, \quad E[\xi^{T}_{ik,t}|w_{i,t}] = 0 \ \ \forall k.$$

Here the $\xi^{T}_{ik,t}$ are the component of the $d\ln T_{ik,t}$ that is unforecastable by the initial equilibrium variables $w_{i,t}$. Then a sufficient condition for an OLS regression that controls for $w_{i,t}$ to identify the δ_gs is that the conditional expectation of the productivity innovations with respect to the foreign shocks and controls is zero,

$$E_{i,t}[\xi^{T}_{ik,t}|w_{i,t}, \{d\ln FMA_{ig,t}\}, \{d\ln CMA_{ig,t}\}] = 0, \ \ \forall k. \tag{18}$$

This condition implies that once we control for the initial equilibrium, the foreign shocks vary independently from the domestic shocks and thus provide exogenous variation that can be leveraged for identification.

Constant Treatment Effects Within Cluster-Country-Time

Our identification result above assumed away the problem of inference in the presence of heterogeneous treatment effects within clusters. We now allow the treatment effects to vary by country and time period, but not within sectors for a given cluster-country-time, i.e. $\delta^{ex}_{ik,t} = \delta^{ex}_{ig,t}$ and $\delta^{im}_{ik,t} = \delta^{im}_{ig,t}$. Unlike the typical application, the heterogeneity in our treatment effects is not random after conditioning on the initial equilibrium. However, we can fully control for the remaining dependence using interactions of the initial equilibrium variables with the treatments. Formally, let $s_{i,t}$ denote the vector of interactions between the initial equilibrium variables $w_{i,t}$ and the g-level foreign shocks

$d\ln FMA_{ig,t}$ and $d\ln CMA_{ig,t}$. Then we can write Equation (14) as

$$d\ln y_{i,t} = \nu_t + \sum_{g\in G} \delta_g^{ex} \cdot [d\ln FMA_{ig,t}] + \sum_{g\in G} \delta_g^{im} \cdot [d\ln CMA_{ig,t}] + \gamma w_{i,t} + \theta s_{i,t} + \tilde{\epsilon}_{i,t}, \quad (19)$$
$$\tilde{\epsilon}_{i,t} = \sum_{k\in K} \gamma_k w_{i,t} \cdot \xi_{ik,t}^T, \;\; E[\tilde{\epsilon}_{i,t}] = 0, \;\; E[\xi_{ik,t}^T | w_{i,t}] = 0 \;\; \forall k.$$

Once we control for both the initial equilibrium and the dependence of the individual treatment effects on the initial equilibrium, our condition for identification remains the same as in the constant elasticity case. Note that our de-meaning of $w_{i,t}$ ensures that there is not full collinearity between the cluster-level treatments and the control $s_{i,t}$.

General Treatment Effects

We now examine the case where the treatment effects also vary by industry within each country-time-cluster. Here we face a more difficult challenge to identification: the mean treatment effects by industry within a cluster vary in a way that we cannot control for without introducing collinearity with the treatments. Formally, and with a slight abuse of notation, let $s_{i,t}$ now denote the vector of interactions between the initial equilibrium variables $w_{i,t}$ and the k-level foreign shocks $\lambda_{ik,t}^{ex} d\ln FMA_{ik,t}$ and $\lambda_{ik,t}^{im} d\ln CMA_{ik,t}$. Then we can write Equation (14) as

$$d\ln y_{i,t} = \nu_t + \sum_{g\in G} \delta_g^{ex} \cdot [d\ln FMA_{ig,t}] + \sum_{g\in G} \delta_g^{im} \cdot [d\ln CMA_{ig,t}] + \gamma w_{i,t} + \theta s_{i,t} + \tilde{\epsilon}_{i,t}, \quad (20)$$
$$\tilde{\epsilon}_{i,t} = \sum_{k\in K} \gamma_k w_{i,t} \cdot \xi_{ik,t}^T + \sum_{g\in G}\sum_{k\in g} (\delta_k^{ex} - \delta_g^{ex}) \lambda_{ik,t}^{ex} d\ln FMA_{ik,t} + \sum_{g\in G}\sum_{k\in g} (\delta_k^{ex} - \delta_g^{ex}) \lambda_{ik,t}^{im} d\ln CMA_{ik,t},$$
$$E[\tilde{\epsilon}_{i,t}] = 0, \;\; E[\xi_{ik,t}^T | w_{i,t}] = 0 \;\; \forall k,$$

where δ_k^{ex} and δ_k^{im} are the mean treatment effects at the industry level. Since the q-level treatments are just sums of the k-level treatments, there is a structural correlation between the error term and the treatments that may lead to bias.

Intuitively, the source of the bias comes from potential for certain sectors to contribute disproportionately to the variation of the cluster level treatment, either because they comprise a larger share of trade or because they face more volatile foreign shocks. If that is the case, then the estimated cluster-level mean treatment effects will disproportionately reflect the contributions of those more highly weighted sectors. As an extreme example, suppose that in a given cluster with 100 industries, only one industry ever experiences a foreign shock. Clearly we cannot use any amount of data to recover the cluster-level mean treatment effect; what we will recover instead is the mean treatment effect for that industry.[10] In the more general case, the elasticities that we recover will be weighted averages of the industry-level mean treatment effects, where the weights reflect the likelihood of treatment conditional on the controls.

3.2 Estimation

We have shown that the group-level treatment effects are identified under reasonable conditions once we adequately control for the initial equilibrium observables. However, the vector of controls may be quite high-dimensional relative to the sample size. This is certainly the case in our application, where we have hundreds of medium-term growth rates but thousands of controls if we include initial import and export shares, interactions, etc. Thus conventional OLS estimation is infeasible.

To address this issue, we use the Post-Double-Selection estimator developed by Belloni et al. (2014b, 2017). This approach involves selecting a subset of "important" controls by regressing each dependent and independent variable on the full set of potential controls using an estimator that sets some or all of the coefficients to zero (e.g. LASSO). The selection is "double" in that the controls are selected based on their correlations with both the dependent and independent variables. The union of the sets of controls that are thus selected (i.e. have non-zero coefficients) in each regression then form the control set for an OLS regression of the dependent variable on the independent varia-

[10] To further build intuition, it may be helpful to consider the following special case in which there is no bias: trade shares are constant within clusters for any given country-time period, and the changes in foreign market access are $i.i.d.$ within cluster-country-time period.

bles.

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.[11] 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 likely to be satisfied in our setting.

4 Data, Clustering and Foreign Shock Estimation

4.1 Data

Our empirical implementation requires data on (i) real GDP across countries, (ii) sectoral bilateral trade flows and trade barriers, and (iii) sectoral characteristics. This section provides summary information on our data sources and measurement, with further details provided in the Data Appendix.

Income per capita data are from Penn World Table version 9.0. We calculate real GDP per capita by using the real GDP at constant national prices and population variables. We compute growth rates at 10-year intervals for a maximum of 5 ten-year growth rates per country (there are some missing values).

The bilateral trade flow data between 1965 and 2015 are from UN Comrade Database at the 4-digit SITC Rev 2 level. We convert the trade data from the SITC product classifications to 1997 NAICS classifications. Appendix B.1 describes the construction of the concordance in detail. All in all, the 784 4-digit SITC items are matched to 268 sectors. Among them, there are 233 manufacturing, 26 agricultural, and 9 mining sectors. We drop countries with population of less than 2 million from our sample. The final sample

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

includes 127 countries and 268 sectors, with a total of 548 10-year growth rates covering the 5 decades from 1965 to 2015. Finally, we obtain geographic variables (bilateral distance and contiguity measures) from CEPII.

The 233 manufacturing sectors are further grouped into clusters based on their sectoral characteristics. We use data from the United States to measure the sectoral characteristics, since sectoral data at comparable 4-digit level of sectoral disaggregation are not available for a large sample of countries. We collect data on 7 sectoral 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 is 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 exchange and not reference priced to capture the importance of relationship-specific investments in a sector.

4.2 K-means Clustering

As discussed above in Section 3, given our sample size and the large number of industries, we focus on estimating average treatment effects within groups or clusters of industries. While average treatment effects for any set of industry groups are identified, it is more useful and interesting to group industries according to characteristics that are both observable and related to the treatment effects. We implement this approach by measuring 7 characteristics (described in the previous sub-section) for each industry, then assigning industries to clusters based on their proximity in the space of characteristics. We apply this approach to the manufacturing industries in our sample.

We use the k-means clustering algorithm (MacQueen et al., 1967) to group sectors into clusters. 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 n different sectoral characteristics, $x^{(i)} \in \mathbb{R}^n$, $i = 1, \ldots, m$, assign the m sectors into J clusters. The J clusters are labeled as $j = 1, 2, \ldots, J$.

1. Initialize cluster centroids $\mu_1, \mu_2, \ldots, \mu_J$ for each cluster.

2. Assign each sector $x^{(i)}$ to closest cluster centroids. The cluster assignment is $c(i) \in \{1, 2, \ldots, J\}$,
$$c(i) = \underset{j \in \{1,\ldots,J\}}{\operatorname{argmin}} ||x^{(i)} - \mu_j||^2.$$

3. Replace cluster centroid μ_j by the coordinate-wise average of all points (sectors) in the jth cluster,
$$\hat{\mu}_j = \frac{\sum_{i=1}^{m} \mathbf{1}(c(i) = j) \cdot x^{(i)}}{\sum_{i=1}^{m} \mathbf{1}(c(i) = j)}.$$

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. As is standard practice, we normalize the values of each characteristic to have zero mean and unit variance.[12]

The algorithm above require a choice of the number of clusters. There is no unambigously optimal method, although there are a number of conceptually similar approaches based on maximizing various measures of cluster fit with respect to the number of clusters. We use the silhouette width (Rousseeuw, 1987) as our measures of cluster fit. Loosely speaking, the silhouette width measures how similar industries within a cluster are to each other relative to industries in the nearest cluster. A good clustering scheme will maximize the average silhouette width while minimizing the number of clusters near the boundaries. The silhouette analysis suggests that either 4 or 5 are good values for number of clusters: Appendix B shows the results of the silhouette analysis along with a full discussion. In the interest of parsimoney we choose to group the 233 manufacturing industries into 4 clusters for our main specification, and show that our results are insensitive to this choice in Appendix B.

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 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 "Producer non-durables." Cluster 3 stands out most clearly as capital goods producers, with an

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

average investment share of 0.52 compared to investment shares ranging from 0.00 to 0.06 in the other clusters. Cluster 4 has the lowest average intermediate sales share, and a negligible average investment sales share. Thus we label it "Consumer goods." Table A1 in Appendix B displays the 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 to Cluster 5, and all mining sectors to Cluster 6. In total, the 268 sectors are grouped into 6 clusters.

Table 1: Summary Statistics of Clusters

	cluster					
	1	2	3	4	Mean	Std. Dev.
Inv. Share	0.00	0.05	0.52	0.04	0.13	0.22
Int. Using	0.78	0.58	0.65	0.66	0.66	0.16
Int. Sales	0.84	0.7	0.27	0.28	0.57	0.31
Conc. Ratio	0.47	0.27	0.38	0.56	0.4	0.21
Sk. Share	0.32	0.28	0.35	0.36	0.32	0.13
Cap. Int.	0.68	0.55	0.54	0.7	0.61	0.1
Con. Int.	0.26	0.56	0.73	0.52	0.51	0.22
Num of ind.	60	84	47	42		
Trade share	0.33	0.26	0.23	0.11		
Label	Raw Materials Processing	Producer Non-durables	Capital Goods	Consumer Goods		
Abbreviation	RAW	INT	CAP	CONS		

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

To obtain $FMA_{ik,t}$ and $CMA_{ik,t}$ for country i sector k at time t, we estimate structural sector-specific gravity equations using the matrix of sectoral bilateral trade flows at decade intervals.[13] For a given sector k at time t, the gravity equation (2) can be rewritten as

$$\lambda_{in,kt} = c_{i,kt}^{1-\sigma_{kt}} \cdot P_{n,kt}^{\sigma-1} \cdot \tau_{in,kt}^{1-\sigma_{kt}}, \tag{21}$$

where $\lambda_{in,kt}$ 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 $\lambda_{in,kt}$ as the share of import expenditure. We model the bilateral resistance term $\tau_{in,kt}^{1-\sigma_{kt}}$ as a function of geographic distance and contiguity with sector-time-specific coefficients, leading to our empirical specification

$$\lambda_{in,kt} = \kappa_{i,kt}^{ex} \cdot \kappa_{n,kt}^{im} \cdot Distance_{in}^{\zeta_{kt}} \cdot \exp\{Contig_{in}\}^{\xi_{kt}} \cdot \epsilon_{i,kt}, \tag{22}$$

where κ_{ikt}^{ex} is the exporter fixed effect, κ_{nkt}^{im} is the importer fixed effect, ζ_{kt} and ξ_{kt} are the distance and common border coefficients. We use PPML to directly estimate the non-linear equation (22), following the methods of Eaton et al. (2012), separately for every sector and time period.

We use our estimates from equation (22) to construct the extermal market access terms as follows:

$$FMA_{i,kt} = \sum_{n \neq i} E_{n,kt} \cdot \kappa_{n,kt}^{im} \cdot distance_{in}^{\zeta_{kt}} \cdot exp\left(\xi_{kt} \cdot Contig_{in}\right) \tag{23}$$

$$CMA_{n,kt} = \sum_{i \neq n} \kappa_{i,kt}^{ex} \cdot distance_{in}^{\zeta_{kt}} \cdot exp\left(\xi_{kt} \cdot Contig_{in}\right), \tag{24}$$

where $E_{n,kt}$ is n's total foreign expenditure in k at time t.

In practice, we add two wrinkles to the method described above. First, we remove any direct effect of a country's exports and imports on the fixed effects of their trading partners by estimating equation (22) N times for each sector and time period, each time leaving out the trade flows from a particular country i. We then construct country i's foreign shocks using the estimates from the regression that omitted their data. Second, as

[13] To reduce measurement error, we use three year averages of the trade flows.

is well known, κ^{ex}_{ikt} and κ^{im}_{nkt} 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 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. We provide a fuller discussion in Appendix B.3.

5 Empirical Results

The top panel of Figure 1 presents the estimation results graphically, by displaying the coefficients on the foreign demand shocks in the left panel, and for the foreign supply shocks in the right panel, by cluster. Clusters 1-4 are manufacturing clusters obtained by the k-means algorithm, cluster 5 is agriculture, and cluster 6 mining and quarrying. The bars depict 95% confidence intervals, obtained using standard errors clustered at the country level. The specification includes the log of initial GDP per capita.

The first apparent feature of the results is the considerable heterogeneity in the coefficients. Indeed, the F-test for the equality of these coefficients rejects it at the 1% level of significance. When it comes to foreign demand shocks, two clusters stand out: export opportunities in cluster 2 ("Producer Non-durables", labeled "INT"), and cluster 3 ("Capital Goods", or "CAP") seem to have a larger and statistically significant pro-growth effect than the other clusters.

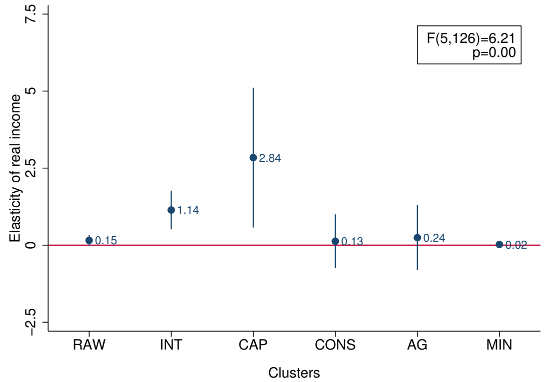
On the foreign supply shock side, there is also some heterogeneity in the coefficients (equality is rejected at the 1% level), but only the shock to the consumer goods supply exhibits a significantly positive impact on growth. Overall, the foreign supply shocks have both much larger magnitudes and standard errors. The latter feature makes it challenging to draw sharp conclusions about the impact of foreign supply shocks on growth. 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 (23) and (24) reveals that foreign demand shocks are detemined 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.

The bottom panel of Figure 1 displays the Post-Double-Selection estimation results (Belloni et al., 2014b). The procedure is described in detail in Appendix B.4. The specification includes a full set of potential controls, namely the industry-level initial equilibrium variables (initial import and export shares, weighted initial firm and consumer market access levels, the squares, and the interactions), interactions between the initial equilibrium variables and the industry-level foreign shocks, initial capital, and initial real GDP per capita. In total, 3219 potential control variables are included and 14 of them are selected in the double-selection procedure via LASSO. Appendix Table A4 lists the selected controls in the Post-Double-Selection estimation.[14] Substantively the results are quite similar, though the confidence intervals widen somewhat. Foreign demand shocks in the Producer Non-Durables cluster still have a significantly positive effect on growth. Foreign demand in the Capital Goods cluster continues to have the largest coefficient, but becomes insignificant. The (low) point estimates on the other clusters are unchanged. For foreign supply, the Consumer Goods continue to be significant at the 5% level, and now the supply shocks to the Capital Goods cluster have a significant impact on growth. Once again, however, the confidence intervals are quite wide.

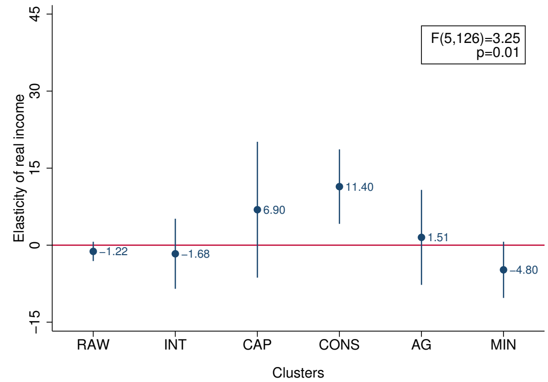
[14]We follow Belloni et al. (2014a) and choose the tuning parameter for the double-LASSO procedure through K-fold cross validation. Appendix B.4.3 describes the procedure in detail. The statistics literature often chooses the tuning parameter to be one standard deviation above the maximizing value in the interests of selecting a more parsimonious model. Our baseline specification uses the maximizing value, which results in more controls being selected. We also check robustness to using a smaller tuning parameter in Appendix Figure A8.

Figure 1: Cluster-Specific Coefficients and Confidence Intervals

A. OLS Estimates

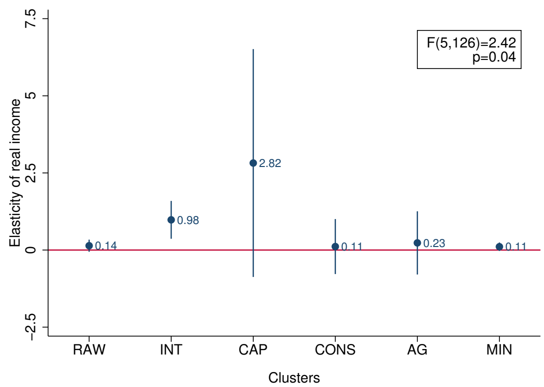


(a) Foreign Demand Shocks

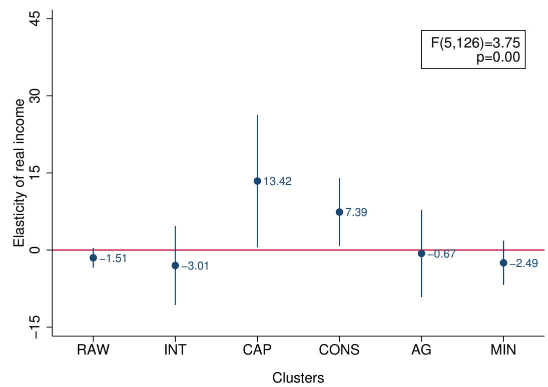


(b) Foreign Supply Shocks

B. LASSO Estimates



(c) Foreign Demand Shocks



(d) Foreign Supply Shocks

Notes: This figure reports the coefficients in estimating Equation ((14)), for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The top panel displays the baseline OLS estimates. The specifications control for initial GDP per capita. The bottom panel displays the post double-LASSO estimates. 14 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country.

5.1 Robustness

5.1.1 Assignment of Sectors to Clusters

One concern with our approach is that clusters may be fragile, in the sense that there are sectors on the margins between the clusters, and the results could be sensitive to assigning specific sectors to clusters. To assess the role 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 A3. The basic characteristics of the original 4 clusters and the labels we attach to them remain similar. When given the opportunity to isolate a 5th cluster, the k-means procedure creates a cluster of skill-intensive industries. 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.

The growth regression results with 5 clusters are presented in Appendix Figure A4. The 5th cluster itself does not have a positive impact on growth, indeed both the foreign demand and foreign supply coefficients are relatively precisely estimated zeros. The main finding that foreign demand shocks to Producer Non-Durables have the most robust association with growth is preserved.

In the second cluster robustness exercise, we assess how important are sectors at the margins of clusters. We add noise (standard deviation of 10% of the actual variability) to each characteristic of each sector, re-cluster sectors, and perform the final estimation of the impact of cluster-specific shocks on growth. We then repeat it 1000 times. The goal of this procedure is to see how the cluster-specific growth-impact coefficients are affected by switching a small number of marginal sectors from one cluster to another.

Appendix Figure A5 reports the results. The dots indicate our baseline coefficient estimates, whereas the bars indicate the 95% range of outcomes. (It is important to stress that these are not "confidence intervals", but rather the range of outcomes when sectors switch clusters.) The figure reveals that many of the cluster-specific coefficient estimates on the foreign demand side are actually quite stable. In particular, the Producer Durables cluster, whose positive impact on growth was the most robustly significant in the baseline and LASSO results, does not feature a large range of outcomes from changing cluster boundaries. In fact, in this simulation the baseline coefficient is on the low end of the range. On the other hand, the Capital Goods cluster, which has both the largest coefficient and the largest standard error in the actual data, also has a very large range of coefficients. This is another reason to be cautious about the inferring a positive impact of foreign demand shocks in that cluster.

5.1.2 Dropping Big and Contiguous Trading Partners

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 when we estimate the fixed effects from the gravity equations, we use the leave-one-out approach, whereby we drop country i from the gravity sample when estimating the fixed effects that go into building country i's FMA's and CMA's. Nonetheless, we check the robustness of the results by dropping the countries for whom i is a large trading partner from the computation of the market access terms.

Specifically, when constructing the country i's external Firm Market Access (FMA) in sector k, we drop importer n from the summation in equation (23) if more than 25% of its imports in sector k is from country i, i.e. $\frac{E_{inkt}}{\sum_{i \neq n} E_{inkt}} > 0.25$. When constructing the country n's external Consumer Market Access (CMA) in sector k, we drop exporter i from the summation in equation (24) if more than 25% of its exports in sector k is to country n, i.e. $\frac{E_{inkt}}{\sum_{n \neq i} E_{inkt}} > 0.25$.

The results are reported in Appendix Figure A6. It is clear that none of the basic results are affected by dropping the large trading partners from the computation of the market access terms.

Our identification relies on the assumption that country i's unobserved productivity shocks are uncorrelated with the foreign market access regressors. It may be, however, that productivity shocks are spacially correlated, so that nearby countries are subject to similar productivity shocks. To assess the possible impact of spatially correlated shocks, we omit contiguous countries from the calculation of the market access terms. The results are reported in Appendix Figure A7. The reveal very little change relative to the baseline.

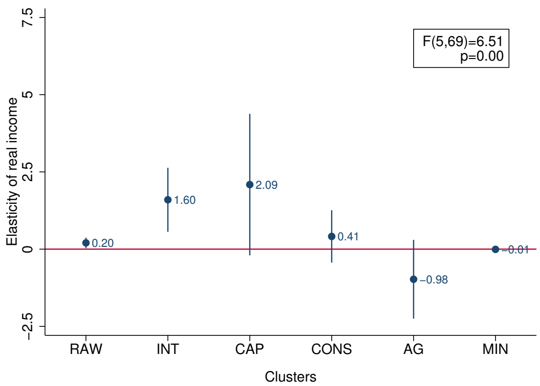
5.1.3 Developed vs. Developing Countries

Finally, we assess whether the growth impact of the foreign market access shocks differs between developed and developing countries. We split the sample into two groups, based on the World Bank's 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. We then estimate elasticities of real income growth with respect to foreign shocks for the two country groups separately. According to this classification, 70 countries belong to the developed group, and 57 to the developing group.

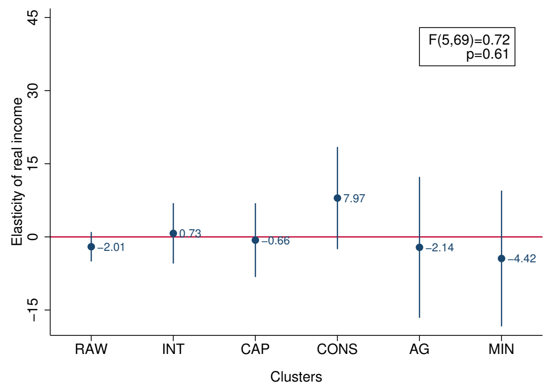
Figure 2 reports the results of the baseline specifications for the developed and developing groups. For both country groups, the Producer Non-Durables coefficients are positive and significant, although the magnitude is larger for the developed country groups. On the other hand, the Capital Goods coefficients behave very differently in the two samples. While the Capital Goods coefficient is small and insignificant in the developed country sample, it is actually quite high and highly significant in the developing country sample. Since the impact of capital goods exports is so different in the two subsamples, Appendix Figures A9-A12 repeat the robustness checks splitting the developed and developing countries, and confirms that the results are unchanged. Then, Appendix Figure A14 computes the country-specific growth elasticities (described in the next section) using country-group-specific coefficients, and confirms that the patterns of growth elasticities are similar whether or not we use group-specific coefficients. This is because while the capital goods foreign demand shocks have a large coefficients among developing countries, the capital goods exports are not a large category quantitatively.

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

A. Developed Countries

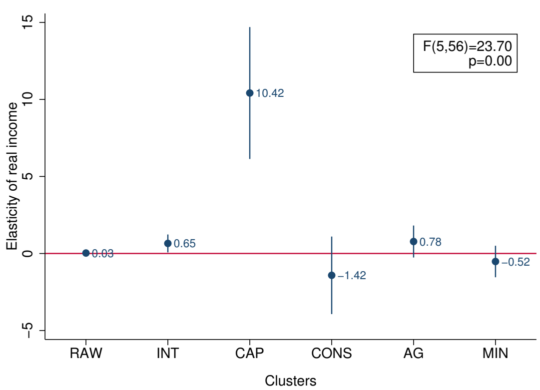


(a) Foreign Demand Shocks

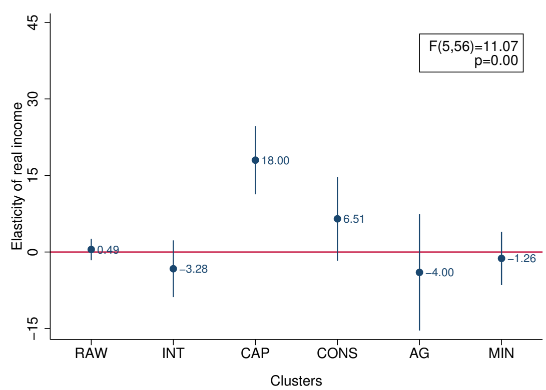


(b) Foreign Supply Shocks

B. Developing Countries



(c) Foreign Demand Shocks



(d) Foreign Supply Shocks

Notes: This figure reports the coefficients in estimating Equation ((14)), for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The top panel displays the results for the sample of developed countries. 9 control variables are selected in the double-selection step. The bottom panel displays the results for developing countries. 2 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

6 Quantitative Implications

To assess the economic significance of the estimated coefficients, we perform two counterfactuals. The first is designed to illustrate the role of comparative advantage. Above, we found that foreign shocks in certain sectors have a higher growth impact than in others. As a result, even a foreign shock that is completely uniform across sectors would be predicted to change economic growth differently across countries, depending on their comparative advantage. To get a sense of the extent of this heterogeneity, we compute the elasticity of each country's economic growth to a worldwide uniform log-change in FMA and CMA, that is the same in every foreign sector and every foreign country. A simple transformation of our estimating equation leads to the following expression for this elasticity:

$$\frac{d\ln y_{i,t}}{d\ln FMA} = \sum_{g\in G} \widehat{\delta}_g^{ex} \sum_{k\in G} \lambda_{ik,t}^{ex},$$

and

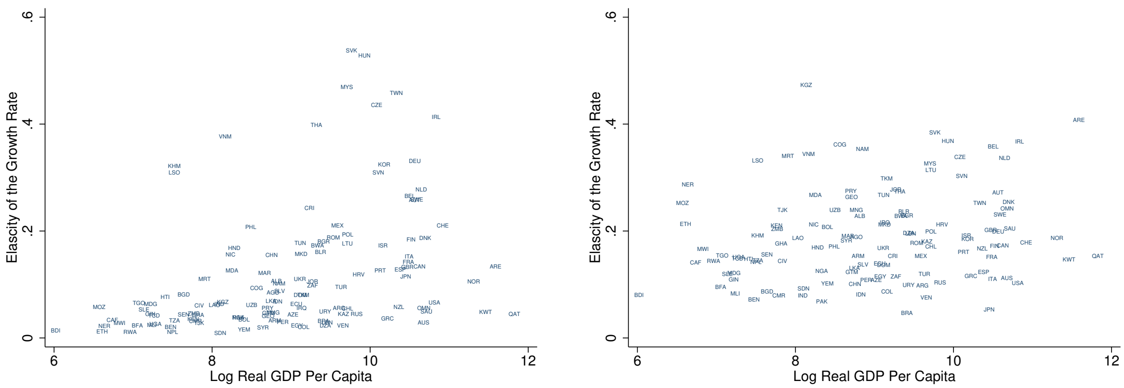
$$\frac{d\ln y_{i,t}}{d\ln CMA} = \sum_{g\in G} \widehat{\delta}_g^{im} \sum_{k\in G} \lambda_{ik,t}^{im}.$$

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 $\lambda_{ik,t}$s. Countries that have high export shares in clusters with a high estimated growth impact will have more positive growth response, all else equal.

The resulting elasticities calculated based on the 2015 import and export shares are plotted in Figure 3 against log PPP-adjusted income per capita. There is indeed a great deal of heterogeneity in the country impact of foreign shocks. The growth elasticity with respect to foreign demand shocks (left panel) ranges from essentially zero for countries chiefly in Sub-Saharan Africa, to 0.4-0.5 for some Central European and East Asian countries such as Hungary, Slovakia, Malaysia, and Taiwan. There is a similar level, and a similar amount of heterogeneity in the elasticity of growth with respect to foreign supply shocks (right panel). Here, the relationship with per capita income is not apparent, as countries in virtually all income groups experiencing about the same range.

Having illustrated the impact of heterogeneity in countries' comparative advantage, our next counterfactual is designed to illustrate the role of geography. Even though there is only one importer fixed effect for each country in each sector, the same vector of wor-

Figure 3: Elasticity of the Growth Rate with Respect to Foreign Shocks



(a) Foreign Demand Shocks

(b) Foreign Supply Shocks

Notes: This figure presents the scatterplot of elasticity of growth rate with respect to the foreign demand shocks (FMA) (left panel), and foreign supply shocks (CMA) (right panel) against real GDP per capita. Elasticity of growth rate is calculated using the baseline estimates of coefficients in estimating equation (14) and the sectoral export and import shares in 2015.

ldwide importer effects is experienced differently by each exporter due to its geographic position. As an example, there is only one change in the demand for capital goods in Germany, and one for China. Suppose that in a particular period, the importer effects reveal that China is having 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. What we would like to understand is how large is this type of heterogeneity. We thus construct counterfactual growth rates that would occur if Belgium experienced Vietnam's market access shocks. This counterfactual answers the question: how much would Belgium's growth change if, in a particular time period 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 2 reports the results for a set of prominent countries, namely the G7 and the BRICS. The first column reports

Table 2: Predicted Annual Growth Difference, 2005-2015

	Growth difference, actual vs:		
G7	Median	25th pctile	75th pctile
Canada	-1.55	-1.95	-1.07
France	-0.89	-1.18	-0.51
Germany	-1.31	-1.68	-0.76
Italy	-0.56	-0.75	-0.29
Japan	0.80	0.66	0.96
UK	-1.43	-1.75	-1.04
US	0.02	-0.13	0.20
BRICS			
Brazil	0.03	-0.06	0.16
China	-1.63	-1.86	-1.26
India	0.37	0.21	0.48
Russia	-0.03	-0.26	0.33
South Africa	0.32	-0.05	0.69

the difference between the country's actual growth 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 per annum faster 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.

A few features of the table stand out. First, the numbers are large and heterogeneous. In this period, most of the G7 countries actually grew substantially *slower* than they could have in an alternative geographic position, and some of these growth differentials are substantial, between 0.5 and 1.5 percent annually. The exception to this pattern is Japan, which grew 0.8 percentage points faster than it would have in the median geographic position. The picture for the BRICS is less clear, with medians closer to zero, with the exception of China, which would have been better off locating in the median position.

Table 3 reports the summary statistics by region and period. The two regions at the extremes are East Asia & Pacific and Western Europe/North America. The median country in East Asia has reaped a substantial and increasing benefit of geographic location. In the most recent decade, its growth has been 0.8 percentage points per annum higher than it would have been had it been located at the median geographic location in the world. This benefit of East Asian location has been consistently positive across 5 decades, and if anything increasing over time. On the opposite end, the typical country in the Western Europe/North America region has for the most part grown slower than it would have had it been moved to the median location. This may first appear surprising, as these are some of the richest and most open countries in the world. However, 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 countries having high market access *levels*. What these results reveal is that these wealthy countries are located next to relatively slow-growing countries, and thus foreign demand and supply have 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 overall growth impact of geographic location is quite a bit smaller overall, 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 of their geographic position, whereas for Latin America and Eastern Europe/Central Asia, their location has had a modest cost.

Finally, we may ask a finer question of 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 happened if it were moved to a particular region. Table 4 presents the results for the period 2005-2015. It reports the per annum change in growth for the median country in the row region 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. 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 anywhere else. 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.

Table 3: Predicted Annual Growth Difference Relative to Median Geographic Location, Medians by Region and Time Period

Region	1975	1985	1995	2005	2015
East Asia & Pacific	0.45	0.55	0.12	0.78	0.80
	[0.16, 0.60]	*[0.23,0.68]*	*[-0.12,0.27]*	*[0.43,2.14]*	*[0.30,1.92]*
	10	14	14	14	14
Eastern Europe & Central Asia	0.00	-0.39	0.09	-0.12	-0.28
	[-0.01,0.00]	*[-0.72,-0.22]*	*[0.04,0.33]*	*[-0.32,0.23]*	*[-0.74,-0.02]*
	2	6	6	24	24
Latin America & Caribbean	-0.25	-0.14	0.00	-0.07	-0.11
	[-0.39,0.02]	*[-0.27,0.03]*	*[-0.19,0.24]*	*[-0.15,0.06]*	*[-0.39,-0.07]*
	18	18	18	18	18
Middle East & North Africa	0.01	-0.09	-0.02	-0.05	0.21
	[-0.02,0.11]	*[-0.63,0.12]*	*[-0.17,0.14]*	*[-0.30,0.12]*	*[-0.08,0.45]*
	7	14	14	15	15
South Asia & 0.03	0.07	0.06	0.04	0.34	
	[-0.01,0.16]	*[0.04,0.20]*	*[0.01,0.07]*	*[-0.01,0.22]*	*[0.31,0.37]*
	4	5	5	5	5
Sub-Saharan Africa	-0.04	0.07	-0.22	0.13	0.23
	[-0.20,0.05]	*[-0.01,0.23]*	*[-0.43,-0.11]*	*[0.06,0.37]*	*[0.04,0.35]*
	28	30	30	30	33
West Europe/North America	-0.06	-0.88	0.58	-0.39	-0.88
	[-0.23,0.04]	*[-1.39,-0.54]*	*[0.34,1.03]*	*[-1.13,-0.12]*	*[-1.55,-0.63]*
	18	18	18	18	18

Notes: This table reports the region- and period-specific differences in economic growth, in percent per annum, 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 the region and time period.

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

	Countefactual 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	0.11	0.87	0.95	0.48	-0.05	0.58	1.40
& Pacific (N = 14)	*[-0.07,0.60]*	*[0.57,2.09]*	*[0.56,2.13]*	*[0.11,1.49]*	*[-0.15,0.55]*	*[0.10,1.71]*	*[1.07,3.05]*
Eastern Europe	-1.14	0.07	0.00	-0.46	-1.14	-0.54	0.81
& Central Asia (N = 24)	*[-2.22,-0.69]*	*[-0.38,0.60]*	*[-0.36,0.55]*	*[-0.98,-0.01]*	*[-2.32,-0.68]*	*[-1.12,-0.22]*	*[0.51,1.36]*
Latin America	-0.55	0.08	0.02	-0.30	-0.65	-0.28	0.37
& Caribbean (N = 18)	*[-0.94,-0.43]*	*[-0.11,0.14]*	*[-0.14,0.13]*	*[-0.56,-0.18]*	*[-1.00,-0.51]*	*[-0.50,-0.19]*	*[0.25,0.62]*
Middle East	-0.44	0.59	0.47	0.19	-0.61	-0.02	1.15
& North Africa (N = 15)	*[-0.95,-0.12]*	*[0.15,0.95]*	*[0.00, 0.76]*	*[-0.18, 0.32]*	*[-0.98, -0.27]*	*[-0.26,0.15]*	*[0.29,1.70]*
South Asia (N = 5)	0.10	0.47	0.50	0.26	0.03	0.22	0.69
	[0.04,0.10]	*[0.45,0.54]*	*[0.33,0.51]*	*[0.15,0.36]*	*[0.03,0.12]*	*[0.19,0.29]*	*[0.68, 0.86]*
Sub-Saharan	-0.30	0.55	0.41	0.05	-0.46	0.04	1.10
Africa (N = 33)	*[-0.65,-0.20]*	*[0.28,0.77]*	*[0.15,0.58]*	*[-0.03,0.17]*	*[-0.73,-0.28]*	*[-0.11,0.13]*	*[0.81,1.38]*
West Europe/	-1.47	-0.59	-0.73	-1.00	-1.69	-1.06	-0.02
North America (N = 18)	*[-2.26,-1.20]*	*[-1.17,-0.30]*	*[-1.26,-0.42]*	*[-1.66,-0.70]*	*[-2.54,-1.41]*	*[-1.81,-0.89]*	*[-0.58,0.48]*

Notes: This table reports the region pair-specific differences in economic growth, in percent per annum, 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.

7 Conclusion

Using a theoretically grounded approach and employing new empirical techniques, we have shown that there is significant heterogeneity in the impact of foreign shocks in different sectors. Positive foreign demand shocks in sectors producing complex intermediate and capital goods generate significantly higher growth than shocks in other sectors, while positive supply shocks to capital goods are especially beneficial to developing countries. Our quantitative results imply that the interaction between initial comparative advantage and the pattern of foreign shocks is important for understanding the variety of growth experiences across countries.

Our results do not have immediate implications for policy, except perhaps that countries should pursue increased market access more vigorously in some sectors relative to others. However, questions surrounding the effect of the external environment on economic development for developing countries 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 growth and validate a focus on the sectoral dimensions of policy. A fuller understanding of optimal sectoral policy requires considering domestic policies as well (Bartelme et al., 2018), along with the every-mysterious drivers of productivity growth.

References

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

Aguiar, Angel, Badri Narayanan, and Robert McDougall, "An overview of the GTAP 9 data base," *Journal of Global Economic Analysis*, 2016, *1* (1), 181–208.

Allen, Treb, Costas Arkolakis, and Yuta Takahashi, "Universal gravity," *Forthcoming, Journal of Political Economy*, 2019.

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

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.

Bartelme, Dominick, "Trade costs and economic geography: evidence from the US," *manuscript*, 2018.

_ , Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare, "External economies of scale and industrial policy: A view from trade," Technical Report, Mimeo, University of Michigan 2018.

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.

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.

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

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

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

—, Samuel S. Kortum, and Sebastian Sotelo, "International Trade: Linking Micro and Macro," February 2012. NBER WP 17864.

Feyrer, James, "Trade and Income – Exploiting Time Series in Geography," 2019.

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

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

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

Hidalgo, Cesar, 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.

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.

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

Kucheryavyy, Konstantin, Gary Lyn, and Andrés Rodríguez-Clare, "Grounded by gravity: A well-behaved trade model with industry-level economies of scale," August 2016. NBER WP No. 22484.

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

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

Nunn, Nathan, "Relationship-specificity, incomplete contracts, and the pattern of trade," *The 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.

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.

Rodriguez, 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.

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.

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

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

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

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:

$$w_j L_{j,k} = \mu_{j,k} \cdot Y_k, \forall j \in J,\ k \in K \tag{25}$$

$$\sum_{k \in K} L_{j,k} = \bar{L}_j,\ \ \forall j \in J \tag{26}$$

$$E = \sum_k \sum_j w_j \cdot L_{j,k} \tag{27}$$

$$P_k^{1-\sigma_k} = z_k c_k^{1-\sigma_k} + CMA_k,\ \ \forall k \in K \tag{28}$$

$$Y_k = c_k^{1-\sigma_k} \left(z_k \frac{e_k \cdot E + \sum_{l \in K} \alpha_{l,k} Y_l}{P_k^{1-\sigma_k}} + FMA_k \right),\ \ \forall k \in K \tag{29}$$

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 ((25)) are the industry factor demand equations, which can be summed to generate aggregate factor demand. The second set of conditions ((26)) 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 ((28)) defines the price index, while the fifth set of equations ((29)) 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 exogeneous 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 ((25)) given a set of factor prices $\mathbf{w}$, then clearly we can reduce this system to a system of J equating factor demand and factor supply.

In a closed economy, the J equations 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

$$\begin{aligned} d\ln y = & \sum_{k\in K} \lambda_k^f d\ln FMA_k + \sum_{k\in K} \left(\lambda_k^d \theta_k^f - \frac{e_k \theta_k^f}{1-\sigma_k} \right) d\ln CMA_k \\ & + d\ln\alpha + \sum_{k\in K} \lambda^d d\ln E_k + \sum_{k\in K} ((1-\sigma_k)(\lambda_k - \lambda_k^d \theta_k^d) - e_k \theta_k^d) d\ln c_k \end{aligned} \tag{30}$$

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

Since α, $d\ln c_k$ and $d\ln E_k$ are all ultimately functions of the exogenous variables $d\ln FMA_k$, $d\ln CMA_k$ and $d\ln T_k$, we can substitute in for these variables to derive the expression in the main text.

Planner's Problem

We denote the quantity of goods in each industry consumed by domestic consumers by $q_k^{c,d}$ for domestic goods and foreign goods $q_{n,k}^{c,f}$ respectively, and use an i subscript to indicate the corresponding intermediate consumption. We denote the quantity exported by $q_{n,k}^{ex}$ and the production function in each sector by F_k. We denote $E_{n,k}/P_{n,k}^{1-\sigma} \equiv D_{n,k}$.

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

$$\max_{q_k^{c,d}, q_{n,k}^{c,f}, q_k^{i,d}, q_{n,k}^{i,f}, q_{n,k}^{ex}, L_{j,k}} \quad \ln U(\{q_k^{c,d}\}, \{q_{n,k}^{c,f}\}) \tag{31}$$

$$s.t. \;\; F_k\left(\{L_{j,k}\}, \{q_k^{i,d}\}, \{q_{n,k}^{i,f}\}\right) = q_k^{c,d} + q_{n,k}^{i,d} + \sum_{n \in N} q_{n,k}^{ex}, \;\; \forall k \tag{32}$$

$$\sum_k L_{j,k} = \bar{L}_j, \;\; \forall j \tag{33}$$

$$\sum_k \sum_n p_{n,k}^f \left(q_{n,k}^{c,f} + q_{n,k}^{i,f}\right) = \sum_{k \in K} \sum_{n \in N} (q_{n,k}^{ex})^{\frac{\sigma_k - 1}{\sigma_k}} \cdot D_{n,k}^{\frac{1}{\sigma}} \tag{34}$$

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

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

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

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

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

$$\max_{q_k^{c,d}, q_k^{c,f}, q_k^{i,d}, q_k^{i,f}, q_k^{ex}, L_{j,k}} \quad \ln U(\{q_k^{c,d}\}, \{q_k^{c,f}\}) \tag{35}$$

$$s.t. \;\; F_k\left(\{L_{j,k}\}, \{q_k^{i,d}\}, \{q_k^{i,f}\}\right) = q_k^{c,d} + q_{n,k}^{i,d} + q_k^{ex}, \;\; \forall k \tag{36}$$

$$\sum_k L_{j,k} = \bar{L}_j, \;\; \forall j \tag{37}$$

$$\sum_k \left(q_k^{c,f} + q_k^{i,f}\right) CMA_k^{\frac{1}{1 - \sigma_k}} = \sum_{k \in K} q_k^{ex} FMA_k^{\frac{1}{\sigma_k}} \tag{38}$$

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

$$\delta_k^{ex} = \mu \cdot \frac{1}{\sigma_k}, \;\; \delta_k^{im} = \mu \cdot \frac{1}{\sigma_k - 1} \tag{39}$$

where μ is the multiplier on the trade balance constraint (and is constant across countries). Our assumption of homotheticity allows us to normalize this constant to equal 1.

Single Factor Economy

We assume upper tier Cobb-Douglas preferences with constant expenditure share e_k. The equilibrium conditions in this case specialize to

$$w\bar{L} = \sum_{k \in K} \left(\frac{w}{T_k}\right)^{1-\sigma_k} \cdot \left(z_k \frac{e_k \cdot w\bar{L}}{z_k \left(\frac{w}{T_k}\right)^{1-\sigma_k} + CMA_k} + FMA_k \right). \tag{40}$$

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

$$\begin{aligned} d\ln w \approx \sum_k & (1-\sigma_k)(\lambda_k^d + \lambda_k^{ex}) d\ln w + \lambda_k^{ex} \cdot d\ln FMA_k \\ & + \lambda_k^d d\ln w - \lambda_k^d (1-\sigma_k)(\bar{\theta}_k d\ln w + \frac{1-\bar{\theta}_k}{1-\sigma_k} \cdot d\ln CMA_k) \end{aligned}$$

where λ_k^d is the domestic sales share and and $\bar{\theta}_k$ is the domestic expenditure share *within* industry k, i.e. $\bar{\theta}_k = \frac{\lambda_k^d}{\lambda_k^d + \lambda_k^{im}}$. The first term captures the effect of changes in wages on domestic costs through both foreign and domestic sales. The second term is the direct effect of changes in export market access. The third term captures the domestic expenditure channel of increases wages. The fourth term captures the effect of changing prices, both domestic and foreign, on nominal income.

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

$$d\ln w \approx \sum_{k \in K} \frac{\lambda_k^{ex} d\ln FMA_k - \lambda_k^d (1-\bar{\theta}_k) d\ln CMA_k}{1 - \left[\sum_{j \in K} (1-\sigma_j)(\lambda_j^{ex} + (1-\bar{\theta}_j)\lambda_j^d) + \lambda_j^d\right]} \tag{41}$$

To solve for the changes in real income, we need to consider the effect on the overal price index $\mathbb{P} = \prod_{k \in K} P_k^{e_k}$. Using the Cobb-Douglas assumption and the results above, we can write

$$d \ln \mathbb{P} \approx \sum_{k \in K} e_k (\bar{\theta}_k d \ln w + \frac{1 - \bar{\theta}_k}{1 - \sigma_k} \cdot d \ln CMA_k) \tag{42}$$

Putting the two results together, we get

$$d \ln y \approx d \ln w - \left[\sum_{k \in K} (1 - \lambda_k^{im}) d \ln w + \sum_{k \in K} \lambda_k^{im} \frac{d \ln CMA_k}{1 - \sigma_k} \right] \tag{43}$$

$$= \lambda_k^{im} d \ln w - \sum_{k \in K} \lambda_k^{im} \frac{d \ln CMA_k}{1 - \sigma_k} \tag{44}$$

$$= \lambda^{im} \cdot \frac{\sum_{k \in K} \lambda_k^{ex} d \ln CMA_k - \lambda_k^d (1 - \bar{\theta}_k) d \ln CMA_k}{\sum_{j \in K} 1 - \left[\sum_{j \in K} \lambda_k^d + (1 - \sigma_k)(\lambda_j^{ex} + (1 - \bar{\theta}_k) \lambda_k^d) \right]} - \sum_{k \in K} \lambda_k^{im} \frac{d \ln CMA_k}{1 - \sigma_k} \tag{45}$$

$$= \kappa \cdot \left[\sum_{k \in K} \lambda_k^{ex} d \ln FMA_k - \frac{\lambda_k^d (1 - \bar{\theta}_k)}{\lambda_k^{im}} \lambda_k^{im} d \ln CMA_k \right] - \sum_{k \in K} \lambda_k^{im} \frac{d \ln CMA_k}{1 - \sigma_k} \tag{46}$$

$$= \kappa \cdot \sum_{k \in K} \lambda_k^{ex} d \ln FMA_k + \sum_{k \in K} \left(\frac{1}{\sigma_k - 1} - \kappa \bar{\theta}_k \right) \lambda_k^{im} d \ln CMA_k, \tag{47}$$

$$\kappa = \frac{\lambda^{im}}{\sum_{j \in K} 1 - \left[\sum_{j \in K} \lambda_k^d + (1 - \sigma_k)(\lambda_j^{ex} + (1 - \bar{\theta}_k) \lambda_k^d) \right]} \tag{48}$$

where $\lambda^{im} = \sum_{k \in K} \lambda_k^{im}$.

This expression simplifies to the following when we set the domestic sales share in each industry, $\bar{\theta}_k$, equal to zero:

$$d \ln y \approx \kappa \cdot \left[\sum_{k \in K} \lambda_k^{ex} d \ln FMA_k - \frac{1}{1 - \sigma_k} \lambda_k^{im} d \ln CMA_k \right], \tag{49}$$

$$\kappa = \frac{1}{\sum_{j \in K} 1 - \left[\sum_{j \in K} (1 - \sigma_k) \lambda_j^{ex} \right]} \tag{50}$$

External Economies

We now consider a single factor economy with upper tier Cobb-Douglas preferences (as above), but with external economies of scale as in Kucheryavyy et al. (2016). The cost function in each industry is given by $c_k = \frac{w}{T_k L_k^{\gamma_k}}$. We specialize their model to the case with zero domestic sales in any industry. The equilibrium conditions can be expressed as

$$w\bar{L} = \sum_{k \in K} \left(\frac{w}{T_k L_k^{\gamma_k}} \right)^{1-\sigma_k} \cdot FMA_k \tag{51}$$

$$wL_k = \left(\frac{w}{T_k L_k^{\gamma_k}} \right)^{1-\sigma_k} \cdot FMA_k, \ \ \forall k \in K. \tag{52}$$

We assume that, for all industries, $\gamma_k(\sigma_k - 1) < 1$ 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. Since all consumption is imported, CMA only matters for welfare through its direct impact on the consumption prices, in exactly the same manner as in the case with no spillovers. Hence we focus on production.

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

$$w\bar{L} = \sum_{k \in K} w^{\frac{(1+\gamma_k)(1-\sigma_k)}{1-\gamma_k(\sigma_k-1)}} \cdot FMA_k^{\frac{1}{1-\gamma_k(\sigma_k-1)}} \cdot T_k^{\frac{\sigma_k-1}{1-\gamma_k(\sigma_k-1)}} \tag{53}$$

Using this expression, it is easy to see that

$$d\ln w \approx \kappa \sum_{k \in K} \left(\frac{1}{1-\gamma_k(\sigma_k - 1)} \right) \lambda_k^{ex} d\ln FMA_k \tag{54}$$

where

$$\kappa = \frac{1}{\sum_{j \in K} 1 - \left[\sum_{j \in K} \frac{(1+\gamma_k)(1-\sigma_k)}{1-\gamma_k(\sigma_k-1)}) \lambda_j^{ex} \right]} \tag{55}$$

B Data 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 1997 NAICS industry classifications, 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 industries, 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 31621 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 industries. We matched them to the corresponding 4-digit NAICS industries.

Overall, the 784 4-digit SITC items are matched to 268 industries. Among them, 233 industries are from the manufacturing sector, 26 from agriculture, and 9 from 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 clustering evaluation. 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 its own cluster with the ones in neighboring clusters (separation).

The silhouette width s_i is measured as follows:

1. (Measuring the cohesion) Measuring the average distance between point i and all other points in the same cluster. Denote it as a_i.

2. (Measuring the separation) Measuring the average distance between i and all points in the nearest cluster. Denote it as b_i.

3. The silhouette width of the observation i is measured as $s_i = \frac{b_i - a_i}{max(a_i, b_i)}$

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.

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 good clustering. The average silhouette method computes the average silhouette of observations for different number of clusters J. The optimal number of clusters J is the one that maximizes the average silhouette over a range of possible values for J.

Appendix Figure A1 plots the silhouette width for industries in each cluster and Appendix Figure A2 plots the average silhouette over the possible cluster number range. The silhouette analysis suggests that either 4 or 5 are good values for number of clusters. While the average silhouette value indicates the 5 clusters is a good clustering, the silhouette analysis suggests that 4-clusters clustering has less industries near the boundary.

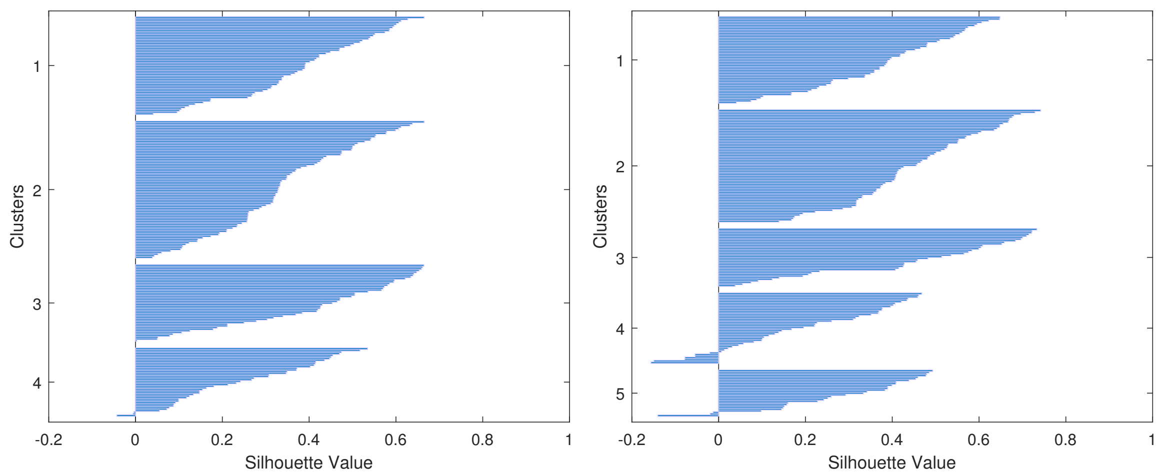
B.2.2 Representative sectors in each cluster

The 233 manufacturing sectors are grouped into 4 clusters using the k-means algorithm. Table A1 lists the 3 most representative sectors in each clusters. The representative sectors are those that are closest to the cluster centroid.

B.2.3 K-means clustering using a subset of characteristic variables

The average silhouette value of 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

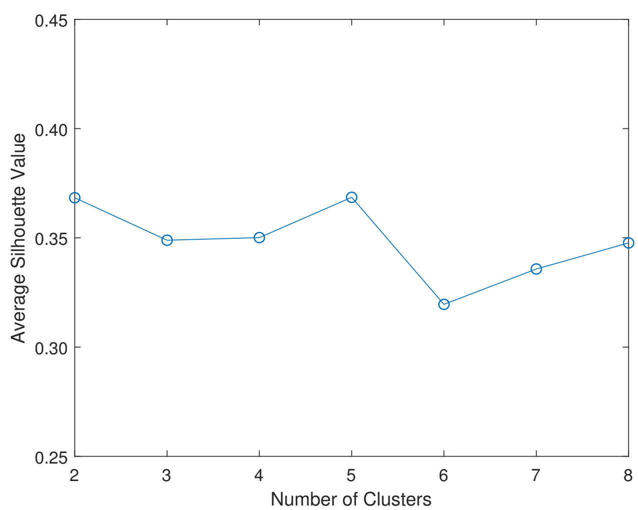
Figure A1: Silhouette Analysis



(a) Cluster Number Equals 4

(b) Cluster Number Equals 5

Figure A2: Average Silhouette Value



Tibshirani (2010). The 4 clusters based on these three characteristics closely replicate the baseline cluster structure; see Table A2. The average silhouette value is now about 0.65, suggesting a strong cluster structure.

Table A1: The 3 Most Representative Sectors in Each Cluster

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

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

Equation (4) and (5) features the relationship between external Firm Market Access (FMA) and external Consumer Market Access (CMA) with the gravity equation. The $FMA_{ik,t}$ and $CMA_{nk,t}$ are denoted as follows,

$$FMA_{ik,t} = \sum_{n \in N} \frac{E_{n,k}}{P_{n,k}^{1-\sigma_k}} \cdot \tau_{in,k}^{1-\sigma_k},$$
$$CMA_{nk,t} = \sum_{i \in N} c_{i,k}^{1-\sigma_k} \cdot \tau_{in,k}^{1-\sigma_k},$$

where i is exporter and n is importer. The foreign shocks are estimated by using

Figure A3: Average Silhouette Value

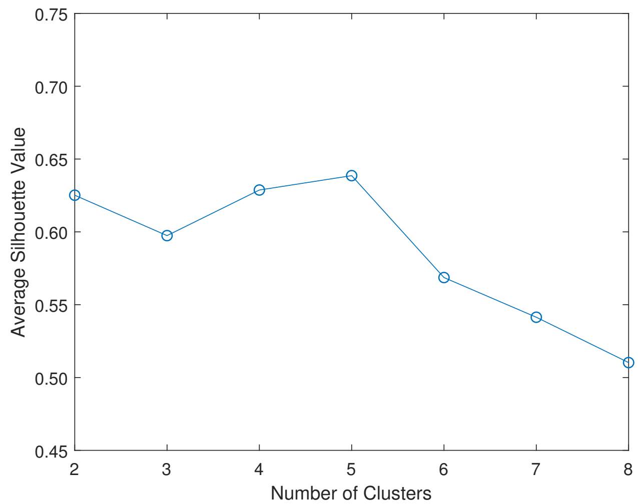


Table A2: Summary Statistics of Clusters: K-means Clustering Using a Subset of Characteristic Variables

	cluster					
	1	2	3	4	Mean	Std. Dev.
Inv. Share	0.01	0.07	0.56	0.05	0.13	0.22
Int. Using	0.7	0.63	0.65	0.63	0.66	0.16
Int. Sales	0.83	0.78	0.28	0.25	0.57	0.31
Conc. Ratio	0.41	0.3	0.34	0.48	0.4	0.21
Sk. Share	0.3	0.31	0.35	0.33	0.32	0.13
Cap. Int.	0.64	0.55	0.55	0.64	0.61	0.1
Con. Int.	0.29	0.65	0.72	0.57	0.51	0.22
Num of ind.	87	45	42	59		
Trade share	0.38	0.16	0.2	0.19		

sectoral bilateral trade flow data and a structural gravity equation.

B.3.1 Gravity regression

Gravity equation (2) can be rewritten as

$$E_{in,kt} \equiv p_{in,k} \cdot q_{in,k} = c_{i,kt}^{1-\sigma_{kt}} \cdot \frac{E_{n,kt}}{P_{n,kt}^{1-\sigma}} \cdot \tau_{in,kt}^{1-\sigma_{kt}}, \tag{56}$$

where $E_{in,kt}$ is denoted as country n's sectoral expenditure on imported consumption and intermediate goods from country i. We do not observe the domestic trade flows, instead we estimate the share version of this equation *a la* Eaton et al. (2012). Dividing both sides by the total imports of country n, we get

$$\frac{E_{in,kt}}{\sum_{i \neq n} E_{in,kt}} = c_{i,kt}^{1-\sigma_{kt}} \cdot \frac{E_{n,kt}}{P_{n,kt}^{1-\sigma} \cdot \sum_{i \neq n} E_{in,kt}} \cdot \tau_{in,kt}^{1-\sigma_{kt}}.$$

It can be estimated by regressing bilateral trade flows on exporter and importer fixed effects and bilateral trade distance. The estimation equation is

$$log\left(\frac{E_{in,kt}}{\sum_{i \neq n} E_{in,kt}}\right) = \kappa_{ikt}^{ex} + \kappa_{nkt}^{im} + \zeta_{kt} \ln Distance_{in} + \xi_{kt} Contig_{in} + \epsilon_{ikt}. \tag{57}$$

where $\frac{E_{in,kt}}{\sum_{i \neq n} E_{in,kt}}$ is the share of total imports from country i to n in sector k at time t, κ_{ikt}^{ex} is the exporter fixed effect, κ_{nkt}^{im} is the importer fixed effect, ζ_{kt} and ξ_{kt} are the distance and common border coefficients. $Distance_{in}$ measures the geographic distance between country i and n, and $Contig_{in}$ indicates whether country i and n are spatially adjacent.

Importer and exporter fixed effects (κ_{ikt}^{ex} and κ_{nkt}^{im}) and the bilateral distance coefficients (ζ_{kt} and ξ_{kt}) are estimated from the above gravity equation. Larger countries may affect their trading partners' estimated importer and exporter effects, thus we implement a leave-one-out estimation algorithm to obtain the country-specific importer and exporter fixed effects. For example, take the estimation of US-specific importer and exporter fixed effects. We drop the US from both the exporter and importer side in our gravity estimation to obtain the corresponding exporter and importer fixed effects for all possible trading partners of the US, as well as the estimates of the distance coefficients. That is to say, in estimating the foreign shocks faced by the US, we drop the US from the gravity sample, and estimate the exporter and importer fixed effects for all the other countries. In this way, we obtain a set of US-specific exporter and importer fixed effects, as well as distance coefficients by sector and time $\{\kappa_{ikt}^{ex}(US)\ \ \kappa_{nkt}^{im}(US)\ \ \zeta_{kt}(US)\ \ \xi_{kt}(US)\}$. In practice this does not affect any of our conclusions. The results are very similar if we run the simple gravity regression with all countries included, and take the importer and exporter fixed effects from that. This reflects the fundamental fact that most countries are small in foreign markets.

The sectoral-level leave-one-out gravity equations are estimated through Poisson pseudo-maximum likelihood approach following Silva and Tenreyro (2006). We estimate the country-specific fixed effects and distance coefficients for 127 countries and 268 sectors over the period 1965-2015. The fixed effects $\kappa^{im}_{nkt}(\omega)$ and $\kappa^{ex}_{ikt}(\omega)$ are identified only up to a sector-time-specific multiplicative constant, and we renormalize the estimated fixed effects by restricting the sum of the importer fixed effects to be zero:

$$\bar{\kappa}^{im}_{nkt}(\omega) = \kappa^{im}_{nkt}(\omega) - \frac{\sum_z \kappa^{im}_{zkt}(\omega)}{N_{kt}(\omega)}$$
$$\bar{\kappa}^{ex}_{ikt}(\omega) = \kappa^{ex}_{ikt}(\omega) + \frac{\sum_z \kappa^{im}_{zkt}(\omega)}{N_{kt}(\omega)},$$

where $\kappa^{im}_{nkt}(\omega)$ and $\kappa^{ex}_{ikt}(\omega)$ are the importer and exporter fixed effects when country ω is left out of the sample, respectively, and $N_{kt}(\omega)$ is the total number of countries with positive imports for industry k and time t when ω is 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.

B.3.2 $FMA_{ik,t}$ and $CMA_{nk,t}$

The $FMA_{ik,t}$ and $CMA_{nk,t}$ are estimated by using a structural gravity regression.

The (log) $c^{1-\sigma_{kt}}_{i,kt}$ and $\frac{E_{n,kt}}{P^{1-\sigma}_{n,kt} \cdot \sum_{i \neq n} E_{in,kt}}$ are estimated by using the importer and exporter fixed effects respectively. We denote the estimated $c^{1-\sigma_{kt}}_{i,kt}$ and $\frac{E_{n,kt}}{P^{1-\sigma}_{n,kt}}$ as ex_{ikt} and im_{nkt}

$$ex_{ikt}(\omega) = \exp\{\bar{\kappa}^{ex}_{ikt}(\omega)\}$$
$$im_{nkt}(\omega) = \left(\sum_{i \neq n} E_{in,kt}(\omega)\right) \cdot \exp\{\bar{\kappa}^{im}_{nkt}(\omega)\},$$

where $\sum_{i \neq n} E_{in,kt}(\omega)$ is total importer expenditure when leaving country ω out, $\bar{\kappa}^{ex}_{ikt}(\omega)\}$ and $\bar{\kappa}^{im}_{nkt}(\omega)\}$ are the renormalized exporter and importer fixed effects.

The iceberg bilateral component $\tau^{1-\sigma_{kt}}_{in,kt}$ are estimated by using the bilateral geographic distance and the common border dummy, as well as corresponding distance and common border coefficients. The estimated bilateral component is denoted as $distance_{in}^{\zeta_{kt}} \cdot exp\left(\xi_{kt} \cdot Contig_{in}\right)$.

The estimated $FMA_{ik,t}$ and $CMA_{nk,t}$ are written as

$$FMA_{ikt} = \sum_{n \neq i} im_{nkt}(i) \cdot distance_{in}^{\zeta_{kt}(i)} \cdot exp\left(\xi_{kt}(i) \cdot Contig_{in}\right)$$
$$CMA_{nkt} = \sum_{i \neq n} ex_{ikt}(i) \cdot distance_{in}^{\zeta_{kt}(i)} \cdot exp\left(\xi_{kt}(i) \cdot Contig_{in}\right).$$

B.4 Post-double-selection Method

B.4.1 Estimation specification

The growth estimation equation is specified as follows,

$$d\ln y_{i,t} = \sum_{g \in G} \delta_g^{ex} \cdot [d\ln FMA_{ig,t}] + \sum_{g \in G} \delta_g^{im} \cdot [d\ln CMA_{ig,t}] + \gamma w_{i,t} + \theta s_{i,t} + D_t + \varepsilon_{i,t},$$

where $d\ln FMA_{ig,t} = \sum_{k \in G} \lambda_{ik,t}^{ex} d\ln FMA_{ik,t}$ and $d\ln CMA_{ig,t} = \sum_{k \in G} \lambda_{ik,t}^{ex} d\ln CMA_{ik,t}$. D_t is the time dummy.

$w_{i,t}$ are the industry-level initial equilibrium variables such as initial import and export shares ($\lambda_{ik,t}^{im}$ and $\lambda_{ik,t}^{ex}$), weighted initial firm and consumer market access ($\lambda_{ik,t}^{ex} \cdot \ln FMA_{ik,t}$ and $\lambda_{ik,t}^{im} \cdot \ln CMA_{ik,t}$), the squares ($(\lambda_{ik,t}^{im})^2$, $(\lambda_{ik,t}^{ex})^2$, $(\lambda_{ik,t}^{ex} \cdot \ln FMA_{ik,t})^2$ and $(\lambda_{ik,t}^{im} \cdot \ln CMA_{ik,t})^2$) and the interactions ($(\lambda_{ik,t}^{ex})^2 \cdot \ln FMA_{ik,t}$ and $(\lambda_{ik,t}^{im})^2 \cdot \ln CMA_{ik,t}$).

$s_{i,t}$ denote the vector of interactions between the initial equilibrium variables and the industry-level foreign shocks, such as $(\lambda_{ik,t}^{ex})^2 \cdot d\ln FMA_{ik,t}$ and $(\lambda_{ik,t}^{im})^2 \cdot d\ln CMA_{ik,t}$.

Since our estimation equation has a large number of controls relative to the sample size, the OLS estimation is infeasible. In this case, dimension reduction is necessary. We estimate the above growth equation by implementing the "post-double-selection" method.

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 controls variables that are useful for predicting the dependent and independent variables respectively. In the post-selection step, coefficients are estimated via ordinary least squares regression by regression dependent variables on the independent variables and the selected controls.

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

$$d \ln y_{i,t} = d_{i,t}\delta + x_{i,t}\beta_y + \mu_{i,t},$$

where $d_{i,t}$ denote a vector of treatment variables $d \ln FMA_{ig,t}$ and $d \ln CMA_{ig,t}$, and $x_{i,t}$ are a vector of control variables.

Applying Lasso directly to our estimation equation above might lead to the omitted-variable bias if the Lasso procedure drop 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 reduce form equation

$$d_{i,t} = x_{i,t}\beta_d + v_{i,t}$$

Substituting the reduce form $d_{i,t}$ into the growth estimation equation. We get

$$\begin{aligned} d \ln y_{i,t} =& x_{i,t}(\beta_d\delta + \beta_y) + (v_{i,t}\delta + \mu_{i,t}) \\ d_{i,t} =& x_{i,t}\beta_d + v_{i,t}. \end{aligned}$$

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 the treatment variables $d_{i,t}$. The reduced form system could be further rewritten as

$$z_{i,t} = x_{i,t}\beta + \varepsilon_{i,t}$$

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

$$\min_{\beta} E(z_{i,t} - x_{i,t}\beta)^2 + \frac{\lambda}{n}||L\beta||_1$$

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 $d_{i,t}$. Then in the post-Lasso step, we estimate δ_g^{ex} and $\delta_g im$ by ordinary least squares regression of $d\ln y_{i,t}$ on $d_{i,t}$ and the union of the variables selected for predicting $d\ln y_{i,t}$ and $d_{i,t}$.

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), Belloni et al. (2014a) and Belloni et al. (2014b). We follow Belloni et al. (2014a) online appendix choosing λ by K-fold cross validation.

The K-fold cross-validation works as follows,

1. Randomly split the data $(y_{i,t}, x_{i,t}, d_{i,t})$ into K subsets of equal size, $S_1, S_2, \ldots, S_k$

2. Set the potential tuning parameter set to be $[\lambda^{RT} - 100 : grid : \lambda^{RT} + 100]$, $grid = 10$. λ^{RT} is the rule of thumb tuning parameter suggested in Belloni et al. (2012) and Belloni et al. (2014b). $\lambda^{RT} = 2.2\sqrt{n}\Phi(1 - \gamma/2p)$, where $\gamma = 0.1/log(p)$, n is the number of observations and p is the number of variables.

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

   (a) (Training on $(y_{i,t}, x_{i,t}, d_{i,t})$, $i \notin S_k$) Leave the kth subsect out, and implement the post-double-selection method with tunning parameter λ on the K-1 subsets. Denote the estimated coefficients as $\hat{\delta}^{-k}(\lambda)$ and $\hat{\beta}_y^{-k}(\lambda)$.

   (b) (Validating on $(y_{i,t}, x_{i,t}, d_{i,t})$, $i \in S_k$) Given, $\hat{\delta}^{-k}(\lambda)$ and $\hat{\beta}_y^{-k}(\lambda)$ computing the error in predicting the Kth part,

   $$e_k(\lambda) = \sum_{i \in S_k} (d\ln y_{i,t} - d_{i,t}\hat{\delta}^{-k}(\lambda) - x_{i,t}\hat{\beta}_y^{-k}(\lambda))^2$$

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 step 3-4 and choose the tuning parameter that minimizes the $CV(\lambda)$.

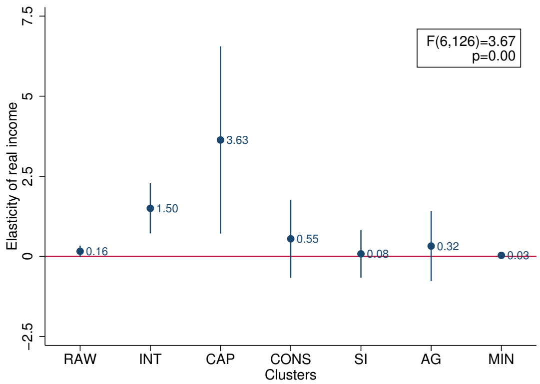
B.5 Robustness Checks

Table A3: Summary Statistics of Clusters: Grouping the Manufacturing Industries to 5 Clusters

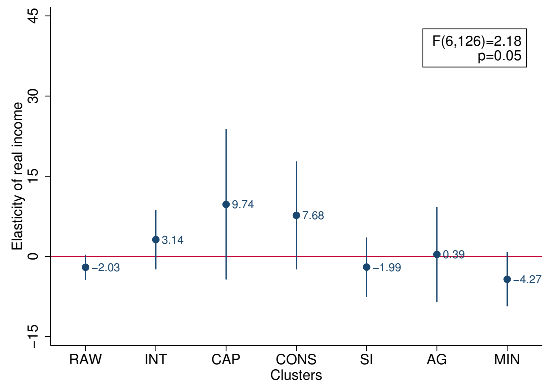
	cluster						
	1	2	3	4	5	Mean	Std. Dev.
Inv. Share	0	0.05	0.57	0.03	0.16	0.13	0.22
Int. Using	0.76	0.62	0.67	0.66	0.57	0.66	0.16
Int. Sales	0.85	0.71	0.26	0.31	0.52	0.57	0.31
Conc. Ratio	0.48	0.23	0.35	0.59	0.41	0.4	0.21
Sk. Share	0.33	0.23	0.3	0.32	0.54	0.32	0.13
Cap. Int.	0.69	0.55	0.54	0.69	0.55	0.61	0.1
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	0.2	0.15	0.07	0.2		
Label	Raw Materials Processing	Producer Non-durables	Capital Goods	Consumer Goods	Skill Intensive		
Abbreviation	RAW	INT	CAP	CONS	SI		

Figure A4: Cluster-Specific Coefficients and Confidence Intervals When Grouping the Manufacturing Industries to 5 Clusters

A. OLS Estimates

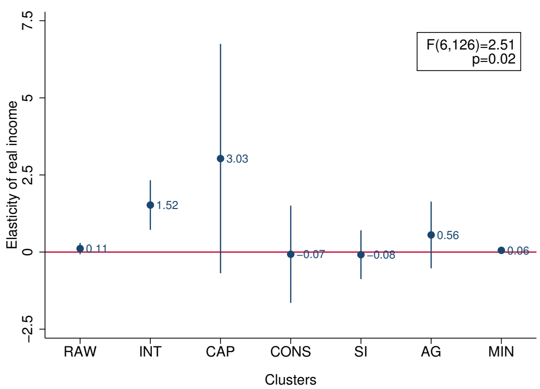


(a) Foreign Demand Shocks

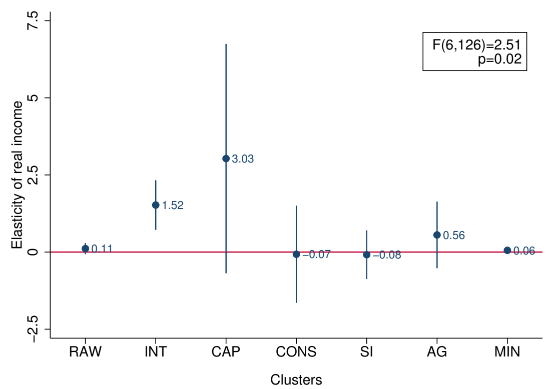


(b) Foreign Supply Shocks

B. LASSO Estimates



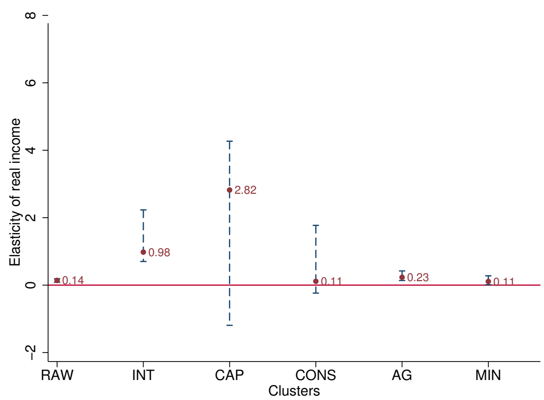
(c) Foreign Demand Shocks



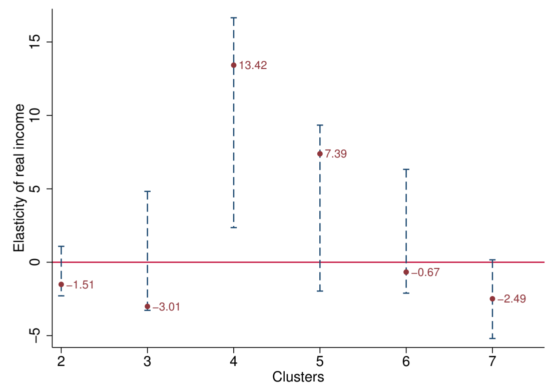
(d) Foreign Supply Shocks

Notes: This figure reports the coefficients in estimating Equation ((14)) when grouping the manufacturing industries to 5 clusters, for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The top panel displays the baseline OLS estimates. The specifications control for initial GDP per capita. The bottom panel displays the post double-LASSO estimates. 11 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country.

Figure A5: Cluster Measurement Error Simulation



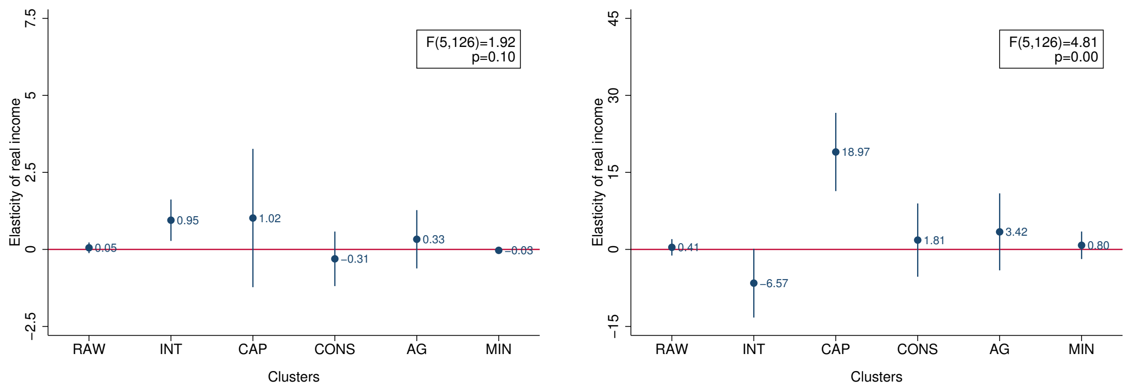
(a) Foreign Demand Shocks



(b) Foreign Supply Shocks

Notes: This figure reports the coefficients in estimating equation ((14)), for the foreign demand shocks (FMA) (left panel), and foreign supply shocks (CMA) (right panel), in the measurement error simulations. The vertical bars report the 95% range of coefficient estimates. The specifications control for initial GDP per capita.

Figure A6: Dropping Large Trading Partners: Cluster-Specific Coefficients and Confidence Intervals

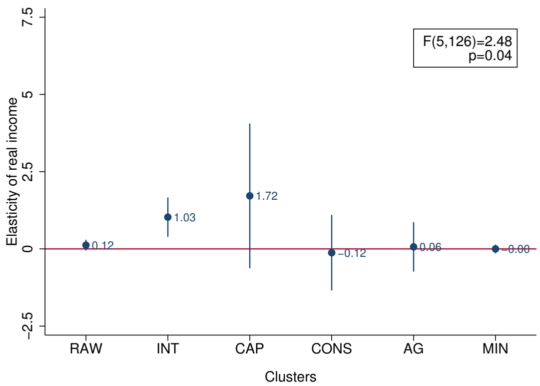


(a) Foreign Demand Shocks

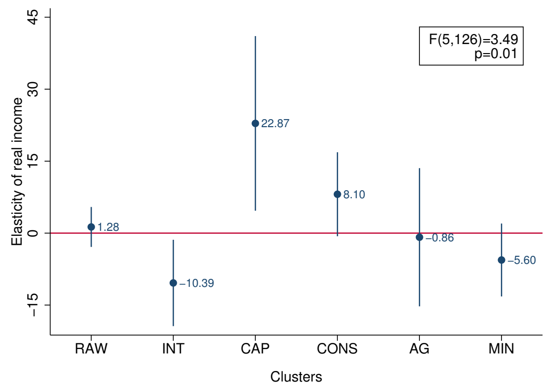
(b) Foreign Supply Shocks

Notes: This figure reports the coefficients in estimating Equation ((14)), for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The construction of the FMA and CMA terms omit foreign markets for which country i is a large trading partner. The figure displays the post double-LASSO estimates. 29 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

Figure A7: Dropping Contiguous Countries: Cluster-Specific Coefficients and Confidence Intervals



(a) Foreign Demand Shocks



(b) Foreign Supply Shocks

Notes: This figure reports the coefficients in estimating Equation ((14)), for the foreign demand shocks (FMA) (left panel), and foreign supply shocks (CMA) (right panel). The construction of the FMA and CMA terms omit contiguous countries. The figure displays the post double-LASSO estimates. 14 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

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

Controls Included	Controls Selected		
	Baseline	Developed Countries	Developing Countries
$\lambda^{ex}_{ik,t}$	$\lambda^{ex}_{i176,t}$ $\lambda^{ex}_{i182,t}$	$\lambda^{ex}_{i176,t}$	
$\lambda^{im}_{ik,t}$			
$\lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t}$	$\lambda^{ex}_{i143,t} \cdot \ln FMA_{i143,t}$	$\lambda^{ex}_{i143,t} \cdot \ln FMA_{i180,t}$	
$\lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t}$	$\lambda^{im}_{i21,t} \cdot \ln CMA_{i21,t}$ $\lambda^{im}_{i76,t} \cdot \ln CMA_{i76,t}$ $\lambda^{im}_{i157,t} \cdot \ln CMA_{i157,t}$ $\lambda^{im}_{i158,t} \cdot \ln CMA_{i158,t}$ $\lambda^{im}_{i166,t} \cdot \ln CMA_{i166,t}$ $\lambda^{im}_{i180,t} \cdot \ln CMA_{i180,t}$ $\lambda^{im}_{i221,t} \cdot \ln CMA_{i221,t}$	$\lambda^{im}_{i74,t} \cdot \ln CMA_{i74,t}$ $\lambda^{im}_{i180,t} \cdot \ln CMA_{i180,t}$ $\lambda^{im}_{i230,t} \cdot \ln CMA_{i230,t}$	
$(\lambda^{ex}_{ik,t})^2$	$(\lambda^{ex}_{i258,t})^2$		$(\lambda^{ex}_{i279,t})^2$
$(\lambda^{im}_{ik,t})^2$			
$(\lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t})^2$			
$(\lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t})^2$		$(\lambda^{im}_{i231,t} \cdot \ln CMA_{i231,t})^2$	
$(\lambda^{ex}_{ik,t})^2 \cdot \ln FMA_{ik,t}$			
$(\lambda^{im}_{ik,t})^2 \cdot \ln CMA_{ik,t}$			
$(\lambda^{ex}_{ik,t})^2 \cdot d\ln FMA_{ik,t}$	$(\lambda^{ex}_{i109,t})^2 \cdot d\ln FMA_{i109,t}$ $(\lambda^{ex}_{i224,t})^2 \cdot d\ln FMA_{i224,t}$ $(\lambda^{ex}_{i233,t})^2 \cdot d\ln FMA_{i233,t}$	$(\lambda^{ex}_{i267,t})^2 \cdot d\ln FMA_{i267,t}$	$(\lambda^{ex}_{i17,t})^2 \cdot d\ln FMA_{i17,t}$
$(\lambda^{im}_{ik,t})^2 \cdot d\ln CMA_{ik,t}$		$(\lambda^{im}_{i207,t})^2 \cdot d\ln CMA_{i207,t}$ $(\lambda^{im}_{i237,t})^2 \cdot d\ln CMA_{i237,t}$	
Number of Controls Selected	14	9	2
Estimates Figures	Figure 1	Figure 2	Figure 2

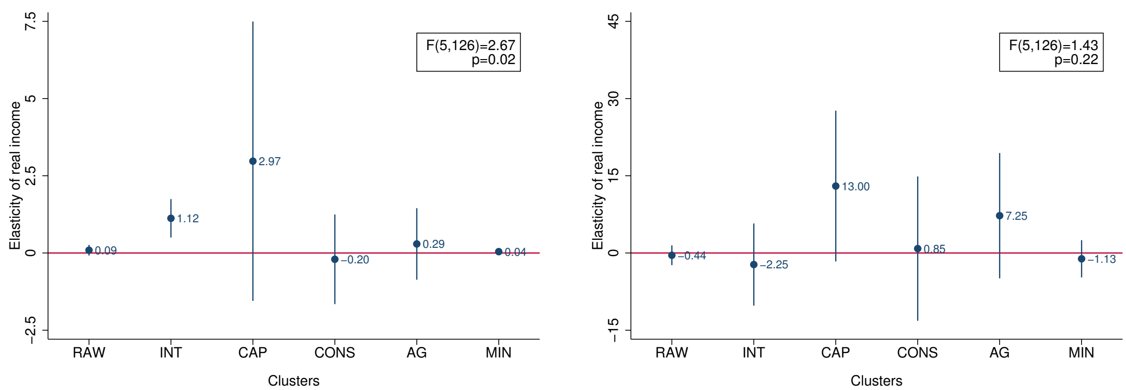
Notes: Industries in our sample are relabeled by number from 1 to 281 for coding purpose, i.e. $k = 1, 2, \ldots, 281$. The numbers in the subscripts refers to the corresponding industries.

Table A5: Control Variables Selected in the Double-Selection Procedure via LASSO: Robustness Checks

Controls included	Controls Selected	
	Dropping Large Trading Partners	Dropping Contiguous Countries
$\lambda^{ex}_{ik,t}$	$\lambda^{ex}_{i96,t}$	$\lambda^{ex}_{i143,t}$
	$\lambda^{ex}_{i104,t}$	$\lambda^{ex}_{i176,t}$
	$\lambda^{ex}_{i114,t}$	
	$\lambda^{ex}_{i152,t}$	
	$\lambda^{ex}_{i176,t}$	
	$\lambda^{ex}_{i182,t}$	
$\lambda^{im}_{ik,t}$		
$\lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t}$	$\lambda^{ex}_{i44,t} \cdot \ln FMA_{i44,t}$	$\lambda^{ex}_{i186,t} \cdot \ln FMA_{i186,t}$
	$\lambda^{ex}_{i103,t} \cdot \ln FMA_{i103,t}$	$\lambda^{ex}_{i203,t} \cdot \ln FMA_{i203,t}$
	$\lambda^{ex}_{i114,t} \cdot \ln FMA_{i114,t}$	
	$\lambda^{ex}_{i143,t} \cdot \ln FMA_{i143,t}$	
	$\lambda^{ex}_{i152,t} \cdot \ln FMA_{i152,t}$	
	$\lambda^{ex}_{i175,t} \cdot \ln FMA_{i175,t}$	
	$\lambda^{ex}_{i186,t} \cdot \ln FMA_{i186,t}$	
$\lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t}$	$\lambda^{im}_{i13,t} \cdot \ln CMA_{i13,t}$	$\lambda^{im}_{i166,t} \cdot \ln CMA_{i166,t}$
	$\lambda^{im}_{i47,t} \cdot \ln CMA_{i47,t}$	$\lambda^{im}_{i180,t} \cdot \ln CMA_{i180,t}$
	$\lambda^{im}_{i99,t} \cdot \ln CMA_{i99,t}$	$\lambda^{im}_{i236,t} \cdot \ln CMA_{i236,t}$
	$\lambda^{im}_{i180,t} \cdot \ln CMA_{i180,t}$	$\lambda^{im}_{i237,t} \cdot \ln CMA_{i237,t}$
		$\lambda^{im}_{i277,t} \cdot \ln CMA_{i277,t}$
$(\lambda^{ex}_{ik,t})^2$	$(\lambda^{ex}_{i224,t})^2$	$(\lambda^{ex}_{i223,t})^2$
$(\lambda^{im}_{ik,t})^2$		
$(\lambda^{ex}_{ik,t} \cdot \ln FMA_{ik,t})^2$		$(\lambda^{ex}_{i224,t} \cdot \ln FMA_{i224,t})^2$
$(\lambda^{im}_{ik,t} \cdot \ln CMA_{ik,t})^2$	$(\lambda^{im}_{i213,t} \cdot \ln CMA_{i213,t})^2$	$(\lambda^{im}_{i207,t} \cdot \ln CMA_{i207,t})^2$
$(\lambda^{ex}_{ik,t})^2 \cdot \ln FMA_{ik,t}$		
$(\lambda^{im}_{ik,t})^2 \cdot \ln CMA_{ik,t}$	$(\lambda^{im}_{i102,t})^2 \cdot \ln CMA_{i102,t}$	
$(\lambda^{ex}_{ik,t})^2 \cdot d\ln FMA_{ik,t}$	$(\lambda^{ex}_{i29,t})^2 \cdot d\ln FMA_{i29,t}$	$(\lambda^{ex}_{i176,t})^2 \cdot d\ln FMA_{i176,t}$
	$(\lambda^{ex}_{i184,t})^2 \cdot d\ln FMA_{i184,t}$	
	$(\lambda^{ex}_{i238,t})^2 \cdot d\ln FMA_{i238,t}$	
$(\lambda^{im}_{ik,t})^2 \cdot d\ln CMA_{ik,t}$	$(\lambda^{im}_{i76,t})^2 \cdot d\ln CMA_{i76,t}$	$(\lambda^{im}_{i127,t})^2 \cdot d\ln CMA_{i127,t}$
	$(\lambda^{im}_{i94,t})^2 \cdot d\ln CMA_{i94,t}$	
	$(\lambda^{im}_{i153,t})^2 \cdot d\ln CMA_{i153,t}$	
	$(\lambda^{im}_{i207,t})^2 \cdot d\ln CMA_{i207,t}$	
	$(\lambda^{im}_{i214,t})^2 \cdot d\ln CMA_{i214,t}$	
	$(\lambda^{im}_{i223,t})^2 \cdot d\ln CMA_{i223,t}$	
Number of Controls Selected	29	14
Estimates Figures	Figure A6	Figure A7

Notes: Industries in our sample are relabeled by number from 1 to 281 for coding purpose, i.e. $k = 1, 2, \ldots, 281$. The numbers in the subscripts refers to the corresponding industries.

Figure A8: Cluster-Specific Coefficients and Confidence Intervals With a Decreased Tuning Parameter



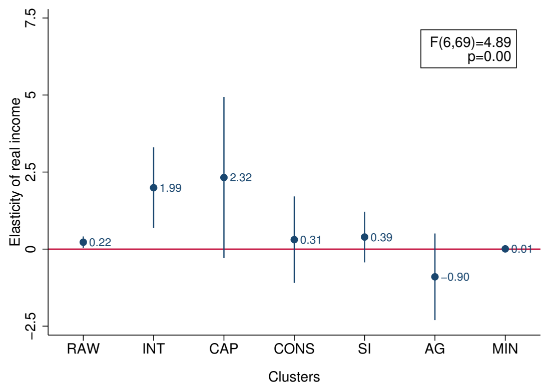
(a) Foreign Demand Shocks

(b) Foreign Supply Shocks

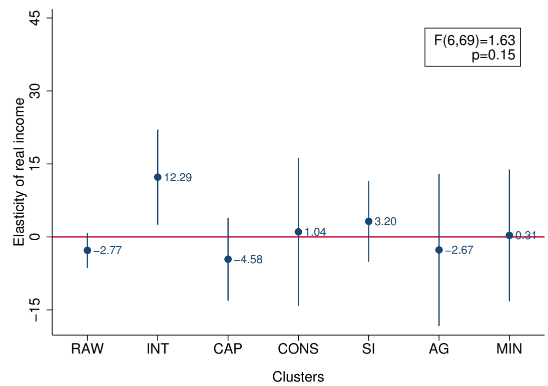
Notes: This figure reports the coefficients in estimating Equation ((14)) with a decreased tuning parameter, for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The figure displays the post double-LASSO estimates. 38 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

Figure A9: Developed vs. Developing Countries: Cluster-Specific Coefficients and Confidence Intervals When Grouping the Manufacturing Industries to 5 Clusters

A. Developed Countries

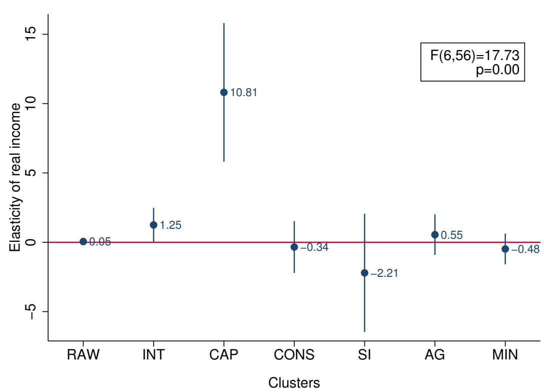


(a) Foreign Demand Shocks

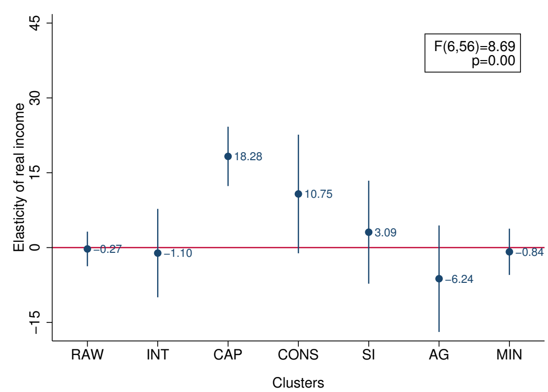


(b) Foreign Supply Shocks

B. Developing Countries



(c) Foreign Demand Shocks

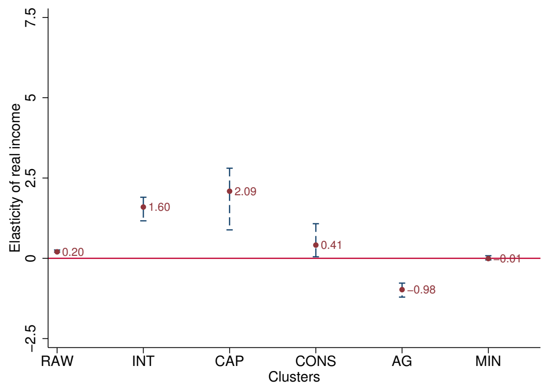


(d) Foreign Supply Shocks

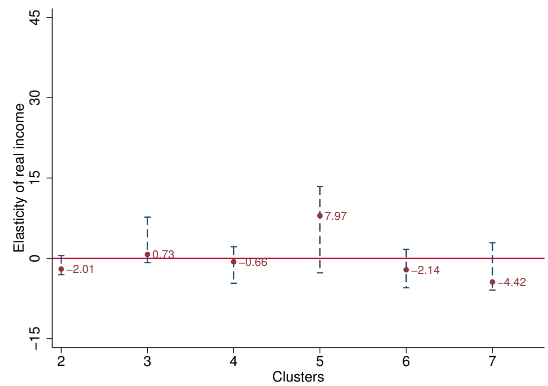
Notes: This figure reports the coefficients in estimating Equation ((14)) when grouping the manufacturing industries to 5 clusters, for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The top panel displays the results for the sample of developed countries. 14 control variables are selected in the double-selection step. The bottom panel displays the results for developing countries. 6 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

Figure A10: Developed vs. Developing Countries: Cluster Measurement Error Simulation

A. Developed Countries

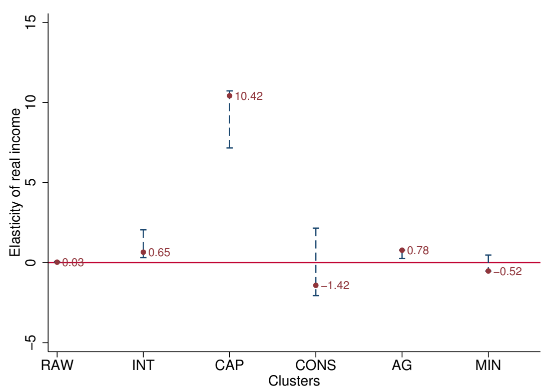


(a) Foreign Demand Shocks

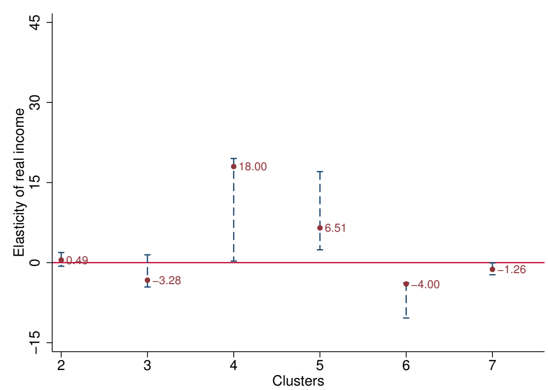


(b) Foreign Supply Shocks

B. Developing Countries



(c) Foreign Demand Shocks

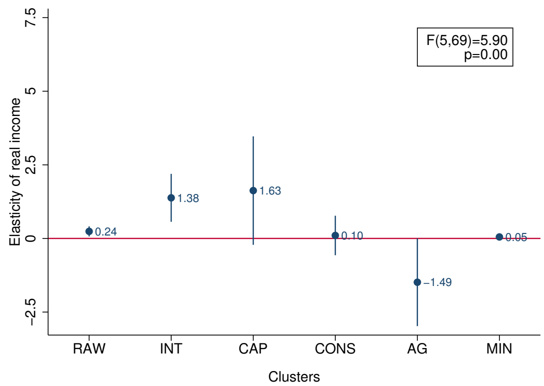


(d) Foreign Supply Shocks

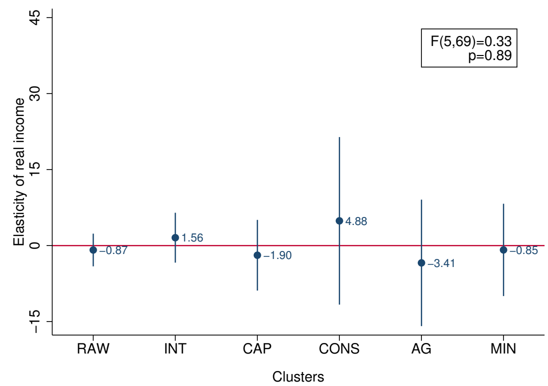
Notes: This figure reports the coefficients in estimating equation ((14)), for the foreign demand shocks (FMA) (left panel), and foreign supply shocks (CMA) (right panel), in the measurement error simulations. The top panel displays the results for the sample of developed countries. The bottom panel displays the results for developing countries. The vertical bars report the 95% range of coefficient estimates. The specifications control for initial GDP per capita.

Figure A11: Developed vs. Developing Countries: Dropping Large Trading Partners

A. Developed Countries

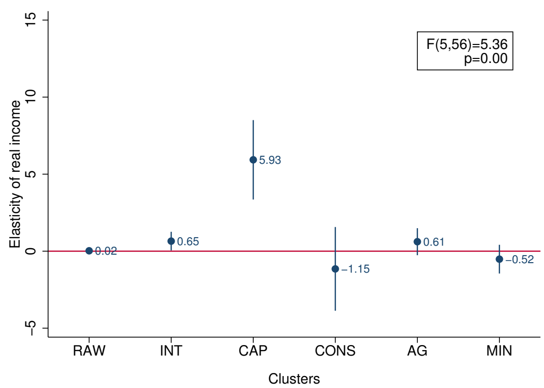


(a) Foreign Demand Shocks

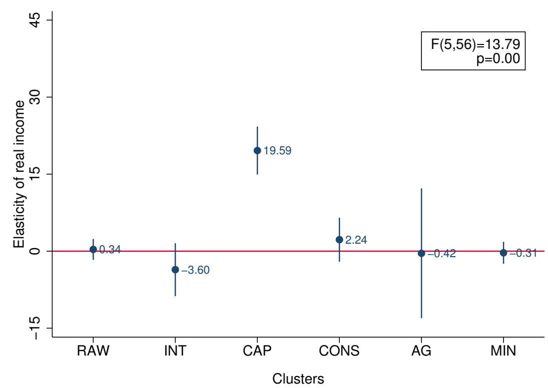


(b) Foreign Supply Shocks

B. Developing Countries



(c) Foreign Demand Shocks

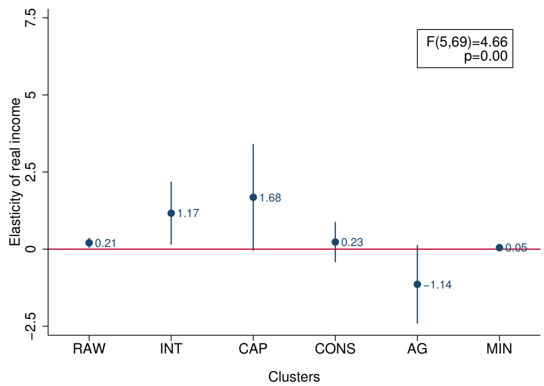


(d) Foreign Supply Shocks

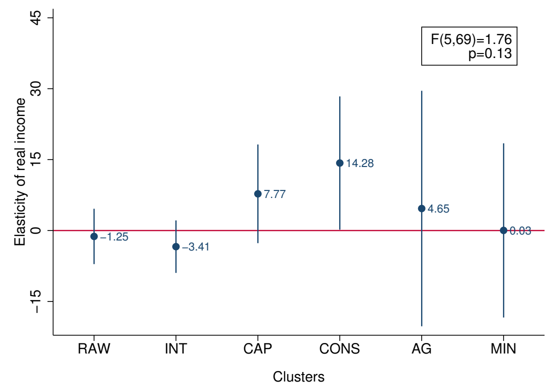
Notes: This figure reports the coefficients in estimating Equation ((14)), for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The construction of the FMA and CMA terms omit foreign markets for which country i is a large trading partner. The top panel displays the results for the sample of developed countries. 2 control variables are selected in the double-selection step. The bottom panel displays the results for developing countries. 2 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

Figure A12: Developed vs. Developing Countries: Dropping Contiguous Countries

A. Developed Countries

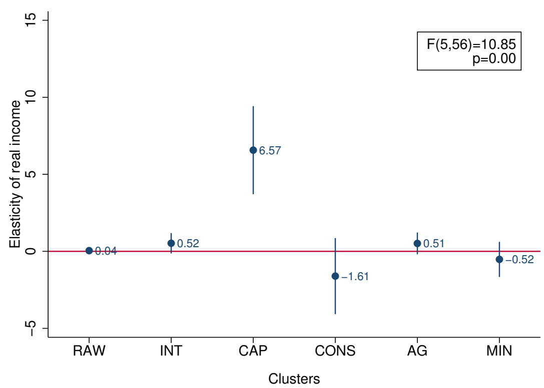


(a) Foreign Demand Shocks

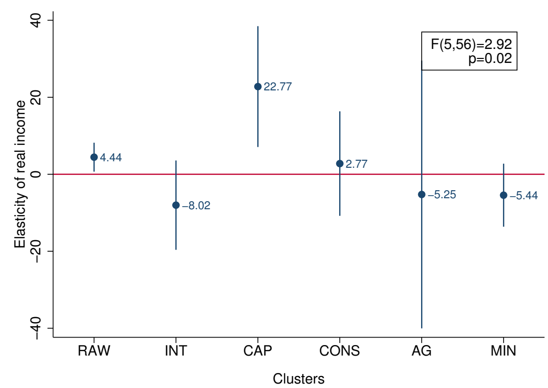


(b) Foreign Supply Shocks

B. Developing Countries



(c) Foreign Demand Shocks

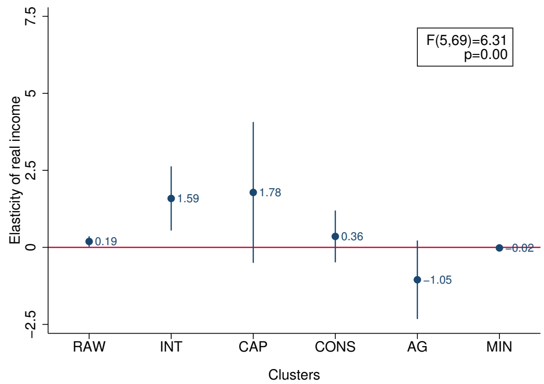


(d) Foreign Supply Shocks

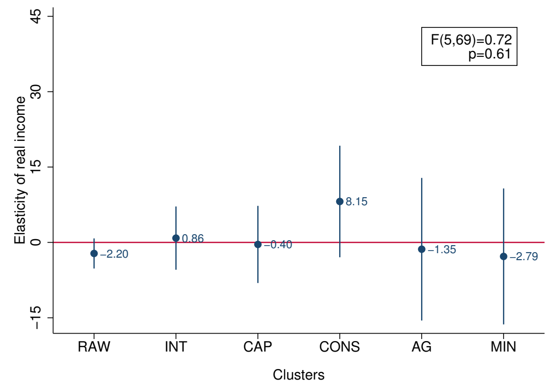
Notes: This figure reports the coefficients in estimating Equation ((14)), for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The construction of the FMA and CMA terms omit contiguous countries. The top panel displays the results for the sample of developed countries. 3 control variables are selected in the double-selection step. The bottom panel displays the results for developing countries. 1 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

Figure A13: Developed vs. Developing Countries: Cluster-Specific Coefficients and Confidence Intervals With a Decreased Tuning Parameter

A. Developed Countries

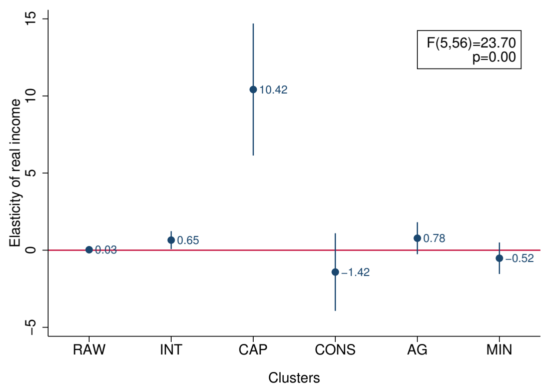


(a) Foreign Demand Shocks

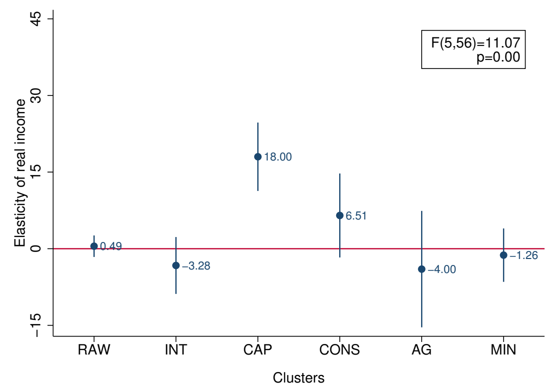


(b) Foreign Supply Shocks

B. Developing Countries



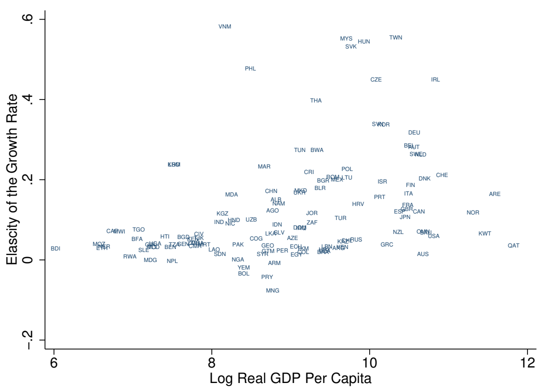
(c) Foreign Demand Shocks



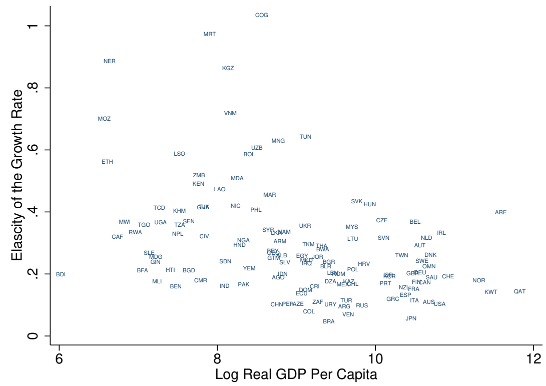
(d) Foreign Supply Shocks

Notes: This figure reports the coefficients in estimating Equation ((14)) with a decreased tuning parameter, for the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels). The top panel displays the results for the sample of developed countries. 9 control variables are selected in the double-selection step. The bottom panel displays the results for developing countries. 2 control variables are selected in the double-selection step. The bars display the 95% confidence bands, that use standard errors clustered by country. The specifications control for initial GDP per capita.

Figure A14: Developed vs. Developing Countries: Elasticity of the Growth Rate



(a) Foreign Demand Shocks



(b) Foreign Supply Shocks

Notes: This figure presents the scatterplot of elasticity of growth rate with respect to the foreign demand shocks (FMA) (left panels), and foreign supply shocks (CMA) (right panels) against real GDP per capita. Elasticity of growth rate is calculated using the developed and developing country-specific estimates of coefficients in estimating equation ((14)) and the sectoral export and import shares in 2015.