Technology, Trade Costs, and the Pattern of Trade with Multistage Production*

Andreas Moxnes[†] Robert C. Johnson[‡]

November 2016

Download Most Recent Version

Abstract

We build a quantitative model of trade with multistage manufacturing production chains, which features iceberg trade costs and technology differences across both goods and production stages. We estimate technology and trade costs via the simulated method of moments, matching bilateral shipments of final goods and inputs. Applying the model, we investigate how comparative advantage and trade costs shape the structure of global production chains and trade flows. As the level of trade costs falls, we show that the elasticity of bilateral trade to trade costs increases, because the endogenous reorganization of production chains (increased export platform production) raises the sensitivity of input trade to trade costs. Surprisingly however, for modest declines in trade costs, the general equilibrium elasticity of world trade is not magnified relative to a benchmark "single-stage" Ricardian model with input-output linkages.

*We thank Lorenzo Caliendo, Thibault Fally, Kjetil Storesletten, and Kei-Mu Yi for helpful comments, as well as seminar participants at Dartmouth College, the University of Colorado, the University of Michigan, the University of Oslo, the World Bank, the 2012 Philadelphia Federal Reserve Trade Workshop, the 2013 ASSA/AEA Meetings, the 2013 NBER ITI Spring Meetings, the 2013 Rocky Mountain Empirical Trade Conference, the 2013 Barcelona GSE Summer Forum, and the 2014 ECARES/CAGE/CEPR Conference on Global Fragmentation and Trade Policy.

†University of Oslo, CEPR & NBER, andreas.moxnes@econ.uio.no

‡Dartmouth College & NBER, robert.c.johnson@dartmouth.edu

Many manufactured goods are produced via multistage production chains, in which individual production stages must be performed in sequence to produce the final good. In a global production chain, these sequential production stages are 'sliced up' and allocated across countries to minimize production costs, giving rise to trade in inputs and final goods.

While this sequential, multistage description of global production chains is straightforward, it is not the standard approach to incorporating input trade into quantitative trade models. Rather, the default approach is to assume that there is a roundabout input loop in the production process, whereby output may be either consumed as a final good or used as an intermediate input. Though tractable and useful for many purposes, models that feature roundabout production fail to capture two basic forces that govern the structure of trade via global production chains.

First, countries differ in the cost at which they can perform individual production stages [Dixit and Grossman (1982); Sanyal (1983); Yi (2003)].[1] Some countries have comparative advantage in downstream stages (e.g., manufacturing assembly in China), while others have comparative advantage in upstream stages (e.g., electronic components in Japan). Therefore, within-sector comparative advantage across stages influences production and trade patterns, in addition to traditional comparative advantage across goods and sectors.

Second, trade costs are particularly burdensome when production takes the sequential, multistage form. As inputs are shipped from country to country through the chain, trade costs are paid multiple times. Further, ad valorem trade costs (proportional to the gross value of goods shipped) are higher in absolute terms for the output of downstream stages, since the value of output accumulates along the production chain. These trade costs have a big impact on decisions about where to locate downstream production stages, because they are large relative to the cost savings from locating downstream stages in low wage or high productivity locations. As emphasized by Yi (2003, 2010), this aspect of multistage production may magnify the elasticity of trade flows to frictions.

In this paper, we build a quantitative model of trade with multistage production to study the role of comparative advantage and trade costs in shaping global production chains, trade patterns, and the elasticity of trade to frictions. The model features two sectors (manufacturing and non-manufacturing), with many goods in each sector. Production of each manufactured good requires a discrete number of sequential stages, while non-manufactured goods are produced via a conventional (single-stage) Ricardian production process.[2] In

[1] Dixit and Grossman (1982) emphasize comparative advantage based on differences in factor endowments within a multistage production process. As in Sanyal (1983) and Yi (2003), we focus on technological differences (Ricardian comparative advantage).

[2] Our model is mostly closely related to the multistage models developed in Yi (2003, 2010), which feature a discrete number of stages (see also Markusen and Venables (2007) and Baldwin and Venables (2013) for

both sectors, manufactured and non-manufactured goods from home and abroad are also used as inputs in production, via a roundabout production loop. Thus, the model features input linkages across both countries and sectors, with both sequential and non-sequential production chains side by side.

To study the quantitative properties of the model, we estimate technology parameters and bilateral trade costs using a simulated method of moments procedure, matching bilateral trade flows of final goods and inputs in the model and data.[3] We then assess the response of production chains and trade flows to changes in technology and trade costs via counterfactual experiments.

One barrier to this estimation and counterfactual analysis is that multistage models are computationally burdensome in high dimensional environments (e.g., with many countries). The basic problem is that equilibrium outcomes are discontinuous in parameters, because the number of number of goods and/or production stages is discrete. This is a familiar problem in Ricardian models, which is compounded in our multistage context.[4]

To overcome this challenge, we borrow a smoothing technique from the discrete choice literature, where similar issues arise in simulating choice probabilities [McFadden (1989)]. Specifically, we solve for an approximate equilibrium in the model, in which discrete (binary) sourcing choices are approximated by continuous (logit-type) functions of prices. With this technique, we are able to solve the model using standard, gradient-based optimization procedures in a multi-country environment, which in turn facilitates simulated method of moments estimation of the parameters. This new procedure allows a tighter mapping between theory and data than previous approaches to quantifying multistage models, and so paves the way for use of the multistage models in future applications.[5]

related models). Costinot et al. (2013) develop a model with a continuum of production stages, building on Dixit and Grossman (1982) and Sanyal (1983), but omit trade costs. Fally and Hillberry (2015) introduce (stylized) trade costs into a model with a continuum of stages, in addition transaction costs that pin down the allocation of stages to firms. Arkolakis and Ramanarayanan (2009) and Bridgman (2008, 2012) study models with more than one production stage that lack the the sequential stage allocation problem that gives rise to trade cost magnification.

[3]Remaining model parameters are calibrated to match various data targets, including final expenditure by sector and country, value-added to output ratios by sector, total GDP by country, and so on.

[4]In a two-country Ricardian model, Dornbusch et al. (1977) solve this problem by assuming that there is a continuum of goods. Because the continuum assumption alone is not enough in multi-country models, Eaton and Kortum (2002) develop a probabilistic approach to analyzing Ricardian models. Unfortunately, their procedure is not directly applicable in multistage models, without additional assumptions. Antràs and de Gortari (2016) adopt an information assumption – whereby agents learn downstream productivity only after locating upstream stages (or vice versa) – to apply the Eaton-Kortum idea in the multistage context. This alternative approach is a complement to the technique we develop in this paper. Our approach does not require any particular information or functional form assumptions, so it is applicable to variants of multistage models where Eaton-Kortum style aggregation fails.

[5]The closest antecedent is Yi (2010), who calibrates a related multistage model for the US and two Canadian regions using a mixture of data (on production, labor allocations, income, etc.) and parameter

Applying this procedure, we estimate technology and trade cost parameters for 15 industrial and emerging market countries, plus a composite rest-of-the-world region, using data from the World Input-Output Database. Our estimates show that there are substantial cross-country differences in relative productivity (comparative advantage) across stages. For example, we find that Mexico has a strong comparative advantage in downstream production, whereas Russia has a comparative advantage in upstream production. These differences in comparative advantage induce specialization, such that the export composition of countries with downstream comparative advantage is tilted toward final goods relative to inputs. We also find that estimated international trade costs are large in our multistage model (on the order of 240%), comparable in magnitude to estimated trade costs in standard gravity models.

Turning to trade elasticities, we start by characterizing bilateral trade elasticities for manufacturing in the model. Though our multistage model does not admit an exact gravity representation of trade flows, we interpret simulated data from our model through the lens of the gravity equation in order to compute elasticities that are comparable to estimates in the literature. Unlike Eaton-Kortum style Ricardian models (or other CES-gravity models), our model generates endogenous, heterogeneous elasticities, which vary across country pairs, for final goods versus inputs, and as the level of trade costs change.

As the level of trade costs falls (holding relative bilateral trade costs constant), we show that the average bilateral trade elasticity rises. The reason is that the bilateral elasticity of input trade rises, due to the endogenous reorganization of production chains in response to falling trade frictions.[6] Specifically, the importance of export platform production – in which inputs sold by country i to country j are re-exported (embedded in finished goods) to third destinations – rises as trade frictions fall. Moreover, the choice of where to set up export platforms is highly sensitive to trade costs, in that export platforms for production chains originating in country i are disproportionately located in countries j for which iceberg trade costs are low. Thus, as the importance of export platform production rises, the sensitivity of input trade to trade costs rises.

To evaluate global trade elasticities, we compare general equilibrium trade elasticities in the multistage model to those from a multi-sector Eaton-Kortum model with input-output structure, similar to Caliendo and Parro (2015).[7] We show that the elasticity of world trade

restrictions (e.g., equal productivity levels across stages). Yi (2010) also measures trade costs from auxiliary data, while we estimate trade costs to match trade shares.

[6] In the cross-section, gravity elasticities for inputs are endogenously higher than for final goods, despite the fact that we assume primitive trade costs are identical for final goods and inputs. Further, gravity elasticities are higher for country pairs with lower bilateral trade costs.

[7] We calibrate this benchmark model to match the equilibrium in our multistage model exactly, so we compute elasticities using different models that fit the exact same set of (simulated) data.

to changes in the level of trade frictions (an untargeted moment in both models) is higher in the benchmark model than in the multistage model at baseline. As the level of trade costs falls, the elasticities cross, and the multistage model generates a higher global trade elasticity when trade costs are low. This result is prima facie consistent with the elasticity magnification argument advanced by Yi (2003).

Nonetheless, we also show that the ratio of trade to GDP behaves in a nearly identical way in the multistage versus benchmark models as trade costs fall. This second result seems to contradict the argument in Yi (2003). Investigating this discrepancy, we argue that the choice of benchmark model is the crucial difference between these results.[8] While we compare our multistage model to a benchmark that allows for (roundabout) input trade, Yi (2003) compares a multistage model to a benchmark model without input trade. By comparing two models that both allow for trade in inputs, but differ in the microeconomics of input trade, we isolate the role of multistage versus roundabout production.

The paper proceeds as follows. We describe the model in Section 1. In Section 2, we discuss how we calibrate and estimate model parameters, and we describe the technology and trade cost estimates. We analyze trade elasticities in Section 3, and Section 4 concludes.

1 Framework

We start this section by laying out the basic elements of our multistage model, describing the economic environment first and defining the model equilibrium. We then discuss how we solve the model numerically, since the model does not admit an analytic solution.

1.1 Economic Environment

Consider a world economy with many countries and two sectors. Countries are indexed by $i, j, k \in \{1, \ldots, C\}$ and sectors are denoted by m and n, standing for manufacturing and non-manufacturing (including agriculture, natural resources, and services) respectively. Within each sector, there is a unit continuum of goods indexed by z. By way of notation, we put country labels in the superscript and good and sector labels in parentheses.

[8]A second important difference is that we estimate trade costs, and estimated trade costs are high. In contrast, Yi (2003) parameterizes trade costs based on tariffs, which are only a small component of total trade costs and thus cannot alone replicate observed home bias in trade. This distinction is important, because the model features non-linear responses to trade costs, where magnification effects kick in strongly only a low levels of trade costs. Thus, marginal changes in trade costs induce smaller trade responses near our estimated equilibrium than they would do if trade costs were (counterfactually) lower.

Manufacturing The manufacturing sector features a discrete multistage production process, as in Yi (2003, 2010). Each good requires $s \in \{1, \ldots, S\}$ production stages to be completed sequentially, and subscripts on each variable index the production stage.

Production in stage 1 uses labor and a composite input, and we assume the production function for good z in sector m is:

$$q_1^i(z,m) = T_1^i(z,m)\Theta_1(m)X^i(z,m)^{\theta_1(m)}l_1^i(z,m)^{1-\theta_1(m)}, \tag{1}$$

where $T_1^i(z,m)$ is the good-specific productivity of country i in manufacturing stage 1, $l_1^i(z,m)$ and $X^i(z,m)$ are the quantities of labor and the composite input used in production, $\theta_1(m)$ is the share of the composite input in production in stage 1, and $\Theta_1(m) = (1-\theta_1(m))^{1-\theta_1(m)}\theta_1(m)^{\theta_1}$ is a normalization.

Production at stages $s > 1$ requires labor and output from stage $s-1$ as in intermediate input, and the production function is given by:

$$q_s^i(z,m) = T_s^i(z,m)\Theta_s(m)x_{s-1}^i(z,m)^{\theta_s(m)}l_s^i(z,m)^{1-\theta_s(m)}, \tag{2}$$

where $T_s^i(z,m)$ is productivity in stage s, $x_{s-1}^i(z,m)$ is the quantity of the stage $s-1$ input used, $l_s^i(z,m)$ is labor used, $\theta_s(m)$ is the cost share attached to the stage $s-1$ input, and $\Theta_s(m)$ is again a parameter normalization.[9]

Output in each stage may be produced in any location, but every time output is shipped between countries it incurs a bilateral, sector-specific, ad valorem iceberg trade cost $\tau^{ij}(m)$.[10]

Non-manufacturing The non-manufacturing sector features Ricardian production and trade, as in Eaton and Kortum (2002). Production of good z in sector n requires labor and the composite intermediate input:

$$q^i(z,n) = T^i(z,n)\Theta(n)X^i(z,n)^{\theta(n)}l^i(z,n)^{1-\theta(n)}. \tag{3}$$

[9]Following Yi (2010), we adopt Cobb-Douglas production functions for stage output. This functional form facilitates calibration, but alternative functional forms are feasible. Further, we do not explicitly include capital in the model. Including endogenous capital stocks, as inYi (2003), would be a straightforward extension of the model.

[10]Extensions in which trade costs differ for final versus intermediate goods are feasible. This exenstion would allow one to consider the effects of input tariff liberalization in the model. Further, extensions with per unit trade costs, as in Irarrazabal et al. (2015), are also interesting. In particular, our iceberg formulation assumes that the proportional burden of trade costs is constant for all stages, but that the absolute value of trade costs incurred is higher for latter stages because the value of output is larger. In constrast, per unit trade costs would imply that the proportional burden of trade costs is lower for later stages.

where $T^i(z,n)$ is productivity, $l^i(z,n)$ and $X^i(z,n)$ are the quantities of labor and the composite input used in production, $\theta(n)$ is the share of the composite input in production, and $\Theta(n)$ is a parameter normalization. Each non-manufacturing good can be produced in any location, and shipping from source to destination incurs bilateral, ad valorem iceberg trade cost $\tau^{ij}(n)$, which may differ from the trade cost for manufactured goods.

Aggregation Within each sector, goods are aggregated to form non-traded composites $Q^i(m)$ and $Q^i(n)$, which are sold to final consumers and used to form the composite input.[11] Each sector-level composite is a Cobb-Douglas combination individual goods:

$$Q^i(m) = \exp\left(\int_0^1 \log(\tilde{q}^i(z,m))dz\right) \tag{4}$$

$$Q^i(n) = \exp\left(\int_0^1 \log(\tilde{q}^i(z,n))dz\right), \tag{5}$$

where $\tilde{q}^i(z,m)$ and $\tilde{q}^i(z,n)$ are the quantities of each good purchased (from low cost sources at home or abroad) by country i. For manufacturing, $\tilde{q}^i(z,m)$ represents purchases of stage S goods.

These sector-level composite goods are combined to form an aggregate final good and the composite input. The aggregate final good is given by $F^i = A^i F^i(m)^{\alpha_i} F^i(n)^{1-\alpha_i}$, where $F^i(m)$ and $F^i(n)$ denote the amount of the sector-level composite good that is sold to final consumers, α_i is a country-specific manufacturing expenditure share, and $A^i = (1-\alpha_i)^{1-\alpha_i}\,\alpha_i^{\alpha_i}$. The composite input is given by $X^i = BX^i(m)^{\beta}X^i(n)^{1-\beta}$, with $X^i = \int_0^1 X^i(z,m)dz + \int_0^1 X^i(z,n)dz$ and $B = (1-\beta)^{1-\beta}\,\beta^{\beta}$. Finally, adding up requires that $Q^i(m) = F^i(m) + X^i(m)$ and $Q^i(n) = F^i(n) + X^i(n)$.

Households Consumers supply labor inelastically to firms and consume the composite final good F_i. The consumer budget constraint is: $w^i L^i = P_F^i F^i + TB^i$, where w^i is the wage, L^i is the labor endowment, P_F^i is the price of the final composite, and TB^i is the nominal trade balance. The trade balance appears here in the budget constraint, since we treat it as an exogenous nominal transfer necessary to equate income and expenditure for each country.

[11] One can think of this aggregation step as an additional production stage with zero value added, or one can embed the aggregation directly into preferences and production functions. We choose the former route, but this choice has no consequences.

1.2 Model Equilibrium

We assume that all goods and factor markets are perfectly competitive. To define and solve for an equilibrium, we need to describe the optimal sourcing decisions. For non-manufacturing, this amounts to determining who the low cost suppliers are for each good to each destination. If $p^j(z,n)$ is the potential factory gate price that country j could supply non-manufacturing good z, and $p^{jk}(z,n) = \tau^{jk}(m)p^j(z,n)$ is the delivered price in country k, then realized price of good (z,n) in destination k is:

$$\tilde{p}^k(z,n) = \min_j p^{jk}(z,n). \tag{6}$$

The potential price at which j can supply non-manufacturing good z is itself given by $p^j(z,n) = \frac{(w^j)^{1-\theta(n)}(P_X^j)^{\theta(n)}}{T^j(z,n)}$, where P_X^j is the price of the composite input (defined below).

For manufacturing, we need to solve for the optimal assignment of stages to countries for production of all goods purchased by each destination, where the assignment of stages to countries depends on the destination in which that good is consumed. If $p_s^j(z,m)$ is the potential factory gate price that country j could supply stage-s of manufactured good z, and $p_s^{jk}(z,m) = \tau^{jk}(m)p_s^j(z,m)$ is the delivered price in destination k inclusive of trade costs, then optimal sourcing implies that the realized price in destination k is:

$$\tilde{p}_s^k(z,m) = \min_j p_s^{jk}(z,m). \tag{7}$$

For stages $s > 1$, the potential factory gate price at which country i can supply stage s output is:

$$p_s^j(z,m) = \frac{(w^j)^{1-\theta_s(m)}(\tilde{p}_{s-1}^j(z,m))^{\theta_s(m)}}{T_s^j(z,m)} \quad \text{with} \quad \tilde{p}_{s-1}^j(z,m) = \min_i p_{s-1}^{ij}(z,m). \tag{8}$$

At stage $s = 1$, the potential factory gate price of output from country i is:

$$p_1^i(z,m) = \frac{(w^i)^{1-\theta_1(m)}(P_X^i)^{\theta_1(m)}}{T_1^i(z,m)}, \tag{9}$$

where P_X^i is again the composite input price.

This sourcing problem has a recursive structure. The price at which k actually purchases stage s output for good (z,m) is the minimum of the set of possible prices at which each country j could deliver that output, conditional on country j choosing the minimum cost source its own purchases of stage $s-1$ output. Potential supply prices for stage $s-1$ inputs in turn depend on optimal sourcing further upstream, at stage $s-2$. And so on. Finally, at stage 1, input supply prices depend on the composite input price in country in each country, which itself is a function of the realized prices (given optimal sourcing) of stage S output.

Given Equation 4, the price of the manufacturing and non-manufacturing composite goods are $P^i(m) = \exp\left(\int_0^1 \log(\tilde{p}^i(z,m))dz\right)$ and $P^i(n) = \exp\left(\int_0^1 \log(\tilde{p}^i(z,n))dz\right)$. Then, the price indexes for F^i and X^i are $P_F^i = (P^i(m))^{\alpha_i}(P^i(n))^{1-\alpha_i}$ and $P_X^i = P^i(m)^{\beta}P^i(n)^{1-\beta}$.

The market clearing conditions for output are:

$$q_S^i(z,m) = \sum_j \tau^{ij}(m)\tilde{q}^j(z,m)\mathbf{1}(p_S^{ij}(z,m) \leq p_S^{kj}(z,m)\,\forall\, k \neq i), \tag{10}$$

$$q_s^i(z,m) = \sum_j \tau^{ij}(m)x_s^j(z,m)\mathbf{1}(p_s^{ij}(z,m) \leq p_s^{kj}(z,m)\,\forall\, k \neq i) \text{ for } 1 \leq s < S, \tag{11}$$

$$q^i(z,n) = \sum_j \tau^{ij}(n)\tilde{q}^j(z,n)\mathbf{1}(p^{ij}(z,n) \leq p^{kj}(z,n)\,\forall\, k \neq i). \tag{12}$$

Market clearing conditions for the composite input and the labor market are:

$$X^i = \int_0^1 X^i(z,n)dz + \int_0^1 X^i(z,m)dz, \tag{13}$$

$$L^i = \int_0^1 l^i(z,n)dz + \int_0^1 l_s^i(z,m)dz. \tag{14}$$

Given parameters $\{\alpha_i, \theta(s), \beta, T_1^i(z,m), T_2^i(z,m), T^i(z,n), \tau^{ij}(m), \tau^{ij}(n), L^i, TB^i\}$, an equilibrium is a collection of prices $\{w^i, \tilde{p}_1^i(z,m), \tilde{p}_2^i(z,m), \tilde{p}^i(z,n), P^i(m), P^i(n), P_X^i, P_F^i\}$, aggregate quantities $\{F^i, X^i, Q^i(m), Q^i(n), F^i(m), F^i(n), X^i(m), X^i(n)\}$, and production, sourcing, and input use decisions $\{X^i(z,n), X^i(z,m), l^i(z,n), l_s^i(z,m), q^i(z,n), \tilde{q}^i(z,n), q_s^i(z,m), x_s^i(z,m), \tilde{q}_2^i(z,m)\}$ such that producers maximize profits, consumers maximize real final expenditure subject to their budget constraint, and product and labor markets clear.

1.3 Discussion

Prior to discussing how we translate this model into a quantitative framework for analysis, we comment on two aspects of the model.

First, the model features both sequential multistage production and roundabout production. Roundabout production introduces a loop in the production process, which amplifies the ratio of gross output to value added. That is, gross output will exceed value added both because multistage production implies that inputs are produced and used up in the production process, but also because production in each sector uses its own output as inputs. Roundabout production also gives rise to input linkages across sectors in the model. We have assumed that sequential production is confined to the manufacturing sector, and that all cross-sector input flows are non-sequential in nature.[12] Both these aspects of the model are important for how we calibrate the model to match the data, discussed further below.

Second, the multistage component of the model is essential to understanding the behavior of the elasticity of trade flows to trade costs. One useful piece of intuition is that stages are more often co-located in the same country when trade costs are high. This implies that the model behaves more like a standard single-stage (multi-sector) Ricardian model as trade costs rise. Since the single-stage model features a constant partial elasticity of trade to changes in trade costs, the multi-stage model will also feature a near constant partial trade elasticity at high levels of trade costs. Further, it will also generate changes in trade and welfare that are similar to the Ricardian benchmark in response to marginal changes in trade costs when trade costs are initially high.

As trade costs fall, it is increasingly attractive to exploit cost differences and break up production stages across countries. The ability to substitute over the location of individual stages, rather than simply over the location of production for entire goods, tends to amplify the sensitivity of trade to trade costs. The key mechanism is that trade costs are paid on the full value of stage output, while cost savings of shifting the the location of a single stage of the production process apply only to the value added at that stage. For downstream production stages, the value of gross output – and thus the trade cost paid for exporting downstream output – is large relative to the marginal value added at that stage. This deters the formation of production chains in which inputs from home are used abroad and the final good is re-exported, either back home or to third countries. Yi (2010) refers to this as the "effective rate of protection" force.

At intermediate levels of trade costs, the model economy features both standard Ricardian trade, where consumers substitute across entire goods, and trade through multistage production chains in which agents substitute over production locations for each stage. Therefore, the aggregate model elasticity of trade to trade costs depends on the mix of Ricardian versus multistage trade. As trade costs fall, the share of trade via multistage production chains rises, so we expect the elasticity of trade to trade costs to rise as well.

1.4 Solving the Model

With an eye toward quantitative implementation of the model, we impose restrictions on technology parameters $\{T_s^i(z,m), T^i(z,n)\}$ to reduce the dimensionality of the parameter space. We assume that productivity parameters $T_s^i(z,m)$ and $T^i(z,n)$ are independent draws

[12] The model could be easily extended to include more than two sectors, possibly including multiple multistage production sectors.

from Fréchet distributions, which have a common shape parameter shape parameter κ, and country,- stage-, and sector-specific location parameters $\{T_s^i(m), T^i(n)\}$.

With these assumptions, we can solve for trade shares and price indexes in closed form for the non-manufacturing sector, as in Eaton and Kortum (2002):

$$\pi^{ji}(n) = \frac{T^j(n)\left(\tau^{ji}(n)(w^j)^{1-\theta(n)}(P_X^j)^{\theta(n)}\right)^{-\kappa}}{\sum_j T^j(n)\left(\tau^{ji}(n)(w^j)^{1-\theta(n)}(P_X^j)^{\theta(n)}\right)^{-\kappa}}, \tag{15}$$

$$P^i(n) = \exp(\gamma/\kappa)\left(\sum_j T^j(n)\left(\tau^{ji}(n)(w^j)^{1-\theta(n)}(P_X^j)^{\theta(n)}\right)^{-\kappa}\right)^{-1/\kappa}. \tag{16}$$

The ability to solve for the non-manufacturing equilibrium in closed form is useful, because it facilitates computation of the full model equilibrium.

It is not possible to solve for the model equilibrium with a continuum of goods in manufacturing, even with Fréchet technology assumptions.[13] Therefore, we introduce a discrete approximation to the continuum of goods in the manufacturing sector. We assume that there are a large, finite number of manufactured goods, and let $r = \{1, \ldots, R\}$ index individual goods. The equilibrium of the model is essentially the same as described above, with r rather than z indexing goods in manufacturing and summations over this set of goods replacing integrals where appropriate.

As in Ricardian models with a finite number of goods, the equilibrium of this discretized model is not continuous in the underlying parameters. This makes the model challenging to solve numerically, and hence also to estimate the model parameters via simulated method of moments. The standard approach to dealing with this complication would be to approximate the continuum with a "large" number of goods, where "large" means a high enough value for R so that the discretized model equilibrium conditions are sufficiently smooth to be accurately solved with standard (derivative-based) numerical methods. In our multi-country setting, this standard approach is computationally burdensome. Therefore, we adopt a different procedure that allows us to use a "small" value for R.

Noting the similarity between sourcing decisions in the model and consumer optimization in discrete choice models, we borrow a smoothing technique developed to facilitate simulation of choice probabilities in the discrete choice literature. Specifically, we draw on the logit-smoothed accept-reject (AR) simulator, developed by McFadden (1989).[14] The key

[13] We follow Yi (2010) in assuming that productivity is drawn from Fréchet distributions in manufacturing, but this assumption can be easily relaxed given the approximate solution technique that we use to solve the model. One important difference relative to Yi (2010) is that we allow countries to have comparative advantage across sectors, while Yi instead restricts the location of the productivity to be the same in all stages, as in $T_s^i(m) = T^i(m)$.

[14] See Train (2009) for a lucid presentation of accept-reject simulators.

observation is that – like simulated choice probabilities – sourcing decisions and hence trade shares are discontinuous. These discontinuities are associated with the presence of indicator functions in the market clearing conditions of the model. The logit-smoothed AR simulator approximates the indicator function with a continuous logit function, as in:

$$\mathbf{1}(p_s^{ij}(r,m) \leq p_s^{kj}(r,m)\,\forall\, k \neq i) \approx \frac{e^{-p_s^{ij}(r,m)/\lambda}}{\sum_k e^{-p_s^{kj}(r,m)/\lambda}}, \tag{17}$$

where $\lambda > 0$ is a smoothing parameter that determines the accuracy of the approximation.[15] Intuitively, when country i is a relatively high cost supplier to j (near the max of the set $\{p_s^{kj}(r,m)\}$), the logit function takes on a value near zero. In contrast, as country i's price falls relative to its competitors, the logit function smoothly converges to one.

With this assumption, the market clearing conditions for manufacturing in the smoothed, discretized model become:

$$q_S^i(r,m) = \sum_j \tau^{ij}(m)\tilde{q}^j(z,m)\left(\frac{e^{-p_S^{ij}(r,m)/\lambda}}{\sum_k e^{-p_S^{kj}(r,m)/\lambda}}\right), \tag{18}$$

$$q_s^i(r,m) = \sum_j \tau^{ij}(m)x_s^j(z,m)\left(\frac{e^{-p_s^{ij}(r,m)/\lambda}}{\sum_k e^{-p_s^{kj}(r,m)/\lambda}}\right) \text{ for } 1 \leq s < S. \tag{19}$$

A brief overview of the procedure we use to solve the model is as follows. Given parameters $\{\alpha_i, \theta(s), \beta, T^i(n), T_s^i(m), \kappa, \tau^{ij}(m), \tau^{ij}(n)\}$, data $\{L^i, TB^i\}$, productivity draws $\{T_s^i(r,m)\}$, and an initial guess for the vector of wages, we can solve for the optimal assignment of stages to countries and equilibrium prices in manufacturing, along with equilibrium prices in non-manufacturing. Given this, we then construct manufacturing and non-manufacturing production, and thus labor demanded. An equilibrium vector of wages equates labor demand and labor supply, as in Equation (14). We describe in detail an algorithm to solve the model in Appendix A.

2 Model to Data

In this section, we describe how we fit the model in Section 1 to the data. We first describe how we calibrate a subset of the parameters of the model and estimate the remainder via simulated method of moments. We then describe our data source. We conclude with a

[15] As $\lambda \to 0$, the logit function converges to the indicator function and the smoothed trade shares approach the exact trade shares in the discrete model. The choice of λ is guided by a trade-off between accuracy and computational speed, and there is little guidance on the appropriate level of λ in general. By trial and error, we find that $\lambda = 0.02$ yields a very good approximation to the exact trade shares.

description of the estimated values for technology and trade cost parameters.

2.1 Fitting the Model

There are a number of free parameters, including technology parameters $\{T_s^i(m), T^i(n), \kappa\}$, trade costs $\{\tau^{ij}(m), \tau^{ij}(n)\}$, and share parameters in production functions and preferences $\{\theta_s(m), \theta_s(n), \beta, \alpha_i\}$. We mix calibration and estimation in pinning down these parameters.

Thus far, we have not specified the number of production stages (S) in the model, but we must take a stand on the number of stages to calibrate/estimate the remaining parameters. In the baseline version of our model, we assume there are two production stages in manufacturing ($S = 2$), following Yi (2003, 2010).[16] We also need to choose a value for the number of manufacturing goods (R), which yields an acceptable trade-off between simulation accuracy and computation time. In Monte Carlo simulations, we have found that our estimation procedure is able to recover the true parameters of the model when $R = 20,000$, so we use this value.

2.1.1 Calibrated Parameters

We calibrate $\{\theta_s(m), \theta(n), \beta, \alpha_i\}$ to match production and expenditure data. The parameter α_i is set to match the share of manufacturing in final expenditure in each country. The median value of α_i is 0.17, with values that range from 0.14 to 0.23 across countries. In the baseline calibration, we assume that $\theta(m) \equiv \theta_1(m) = \theta_2(m)$ and set $\theta(m) = 0.69$ and $\theta(n) = 0.43$ to match the ratio of value-added to output for the world as a whole in the manufacturing and non-manufacturing, respectively.[17] Given this, we calibrate β match input flows across manufacturing and non-manufacturing sectors for the world.[18] This yields $\beta = 0.17$, which means that the composite input is composed primarily of non-manufacturing output.

[16] As work in progress, we are examining the robustness of our results to adding a third production stage.

[17] For each country, value added in manufacturing is equal to the wage bill: $va^i(m) = \sum_{r=1}^{R} w^i(l_1^i(r,m) + l_2^i(r,m)) = (1-\theta(m))go^i(m)$, where $go^i(m) = \sum_{r=1}^{R}[p_1(r,m)q_1^i(r,m) + p_2(r,m)q_2^i(r,m)]$. Then $1-\theta(m) = \sum_i va^i(r,m)/\sum_i go^i(m)$, so we set $\theta(m)$ to match the ratio of value-added to output in manufacturing for the world as a whole. While we could allow $\theta(m)$ to vary across countries, we have chosen not to do so for two reasons. First, while differences in value-added to output ratios differ a lot between manufacturing and services, differences across countries within sectors are more muted. Second, imposing a common value for $\theta(m)$ across countries facilitates calibration of β.

[18] Let $y(m) = y_1(m) + y_2(m)$ be the value of total world output in manufacturing, where$y_1(m)$ and $y_2(m)$ denote stage 1 and stage 2 output. Whereas $y(m)$ is directly observable, $y_1(m)$ and $y_2(m)$ are not. In the model, $y_1(m) = \theta(m)y_2(m)$, so $y(m) = (1+\theta(m))y_2(m)$. Further, $y_2(m) = \sum_i \alpha_i P_F^i F^i + \beta \sum_i P_X^i X^i$, with $\sum_i P_X^i X^i = \theta(m)y_1(m) + \theta(n)y(n)$, where $y(n)$ is total world non-manufacturing output. Combining these observations, $(1+\theta(1))^{-1}y(m) = \sum_i \alpha_i P_F^i F^i + \beta\theta(m)\left[\frac{\theta(m)}{1+\theta(m)}\right]y(m) + \beta\theta(n)y(n)$. Given parameters $\alpha_i, \theta(m), \theta(n)$ and data $y(m), y(n), P_F^i F^i$, this equation can be solved for β.

2.1.2 Estimation of Technology and Trade Cost Parameters

In Section 1.4, we assumed that $\{T_s^i(z,m), T^i(z,n)\}$ are independent draws from Fréchet distributions. We set the common shape parameter in these distributions to $\kappa = 4.12$, guided by Simonovska and Waugh (2014). We leave technology levels $\{T_s^i(m), T^i(n)\}$ as parameters to be estimated, and we normalize $T_1^1(s) = T_2^1(s) = T^i(n) = 1$ so that technology levels are measured relative to country 1.

We parameterize trade costs by assuming that bilateral trade costs are a power function of distance: $\tau^{ij}(m) = \tau^j \left(d^{ij}\right)^{\rho(m)}$ and $\tau^{ij}(n) = \tau \left(d^{ij}\right)^{\rho(n)}$, where d^{ij} is the distance between country i and country j, $\{\tau^j, \tau\}$ are a level parameters for trade costs, and $\rho(m)$ and $\rho(n)$ are sector-specific elasticities of trade costs to distance.[19] We set trade costs on domestic shipments to one in all countries ($\tau^{ii}(m) = \tau^{ii}(n) = 1$).

We estimate the parameters $\Theta = \{T_1^i(m), T_2^i(m), T^i(n), \tau^j, \tau, \rho(m), \rho(n)\}$ by minimizing the distance between trade flows in the model and data, given data on expenditure in each market $\{P_F^i F^i\}$, trade balances $\{TB^i\}$, and labor endowments $\{L^i\}$. Together, these data allow us to compute wages as $w^i = \left(P_F^i F^i + TB^i\right)/L^i$, and we set $w^1 = 1$ as our price normalization.[20] Given wages and a candidate parameter vector $\tilde{\Theta}$, we draw $\{T_s^i(r,m)\}$, compute the model equilibrium, and form a vector of moments based on trade shares.

The first set of moments are based on bilateral trade shares for final manufactured goods. In the model, the share of final goods purchased by destination j from source i as a share of final expenditure in country j is equal to the probability that destination j sources stage 2 goods from country i, because expenditure per good is constant. Employing the same smoothing procedure described in Section 1.4, we compute these shares as:

$$\pi_F^{ij}(m) = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{e^{-p_2^{ij}(r,m)/\lambda}}{\sum_k e^{-p_2^{kj}(r,m)/\lambda}}\right). \tag{20}$$

The second set of moments are based on trade shares for manufactured inputs. Input shipments from country i to j include both stage 1 goods and stage 2 goods destined for the

[19] We adopt a parsimonious specification for trade costs to limit the number of parameters that need to be estimated. While the specification we choose is sufficient to fit the data well, a more flexible specification would be feasible. The assumption that trade costs have a destination-specific component can be motivated in a number of ways. The most obvious is that it captures differences in multilateral import protection. In relation to prior work, Eaton and Kortum (2002) also assume that there is a destination-specific, but not source-specific, component of bilateral trade costs. We omit source-specific effects (in part) to reduce the dimensionality of the parameter space for estimation.

[20] Since we build these values into the estimation, our model equilibrium will match data on aggregate expenditure and GDP exactly.

composite input. Input shipments from i to j of sector m goods are:

$$Inputs^{ij}(m) = \sum_{r=1}^{R} \left(\frac{e^{-p_1^{ij}(r,m)/\lambda}}{\sum_k e^{-p_1^{kj}(r,m)/\lambda}} \right) [\theta(m) p_2^j(r,m) q_2^j(r,m)] + \frac{1}{R} \sum_{r=1}^{R} \left(\frac{e^{-p_2^{ij}(r,m)/\lambda}}{\sum_k e^{-p_2^{kj}(r,m)/\lambda}} \right) [\beta P_M^j M^j]. \quad (21)$$

Then the share of inputs from source i in country j's total purchases of manufactured inputs is: $\pi_I^{ij}(m) = \frac{Inputs^{ij}(m)}{\sum_k Inputs^{ij}(m)}$. The third set of moments are based on trade shares in the non-manufacturing sector, which can be computed in closed form as in Equation (15).

Since the trade shares sum to one for each importer, we only use off diagonal trade shares $(i \neq j)$ for estimation. This gives us $3\,(N^2 - N)$ moments to estimate $(3N-3)+N+3$ unknown parameters. Letting π^{ij} denote the vector of trade shares for pair ij, we stack log differences between actual and simulated trade shares $\pi^{ij}(\Theta) = \ln \pi^{ij} - \ln \hat{\pi}^{ij}(\Theta)$ in a column vector $\pi(\Theta)$. The moment condition is then $E[\pi(\Theta_0)] = 0$, where Θ_0 is the true value of Θ, so we estimate a $\hat{\Theta}$ that satisfies:

$$\hat{\Theta} = \arg\min \left\{ \pi(\Theta)' \pi(\Theta) \right\} \quad (22)$$

One point worth noting here is that the algorithm we use to solve this minimization problem exploits the existence of closed forms in non-manufacturing to speed computation, by estimating non-manufacturing parameters via regression within the simulated MoM procedure for manufacturing parameters.

2.1.3 Data

We draw all data from the World Input-Output Database (WIOD) [Dietzenbacher et al. (2013); http://www.wiod.org/]. This data covers 40 countries and a composite rest of the world region at an annual frequency from 1995-2011. To limit the computational burden, we extract data for 15 countries separately, and aggregate the remainder into a composite rest of the world region. The countries included individually are Australia, Brazil, Canada, China, France, Germany, India, Italy, Japan, Mexico, Russia, South Korea, Spain, the United Kingdom, and the United States. Further, we aggregate the data to the two-sector level, defining manufacturing as sectors 4-16 in the WIOD nomenclature.

Consistent with the discussion above, we use data on aggregate final expenditure and the trade balance for each country, and we measure aggregate labor endowments using aggregate hours worked from WIOD's auxiliary Socio-Economic Accounts. We also use data on bilateral shipments of final goods and inputs to form trade shares, including each country's purchases from itself. In order to calibrate $\{\theta_s(m), \theta(n), \beta\}$, we aggregate across countries (including the rest of the world) to compute world-level data on sector-level value added, sector-level gross output, and cross-sector input shipments. Finally, in estimating the trade costs, we use bilateral distance data from CEPII.

2.2 Technology and Trade Cost Estimates

We present estimates for technology levels by stage for the 15 countries and the composite region in Table 1. The first two columns present the geometric means of $T_1^i(z,m)$ and $T_2^i(z,m)$ in each country, expressed relative to relative to the United States.[21] The estimates indicate that all countries have technology levels lower than the U.S. level. Productivity levels are also correlated across stages: countries with high absolute productivity in stage 1 tend to also have high absolute productivity in stage 2. Despite this correlation, there are sizable cross-country differences in relative productivity across stages.

The final column in Table 1 reports the ratio of mean technology in stage 2 relative to stage 1 in each country, where numbers greater than one indicate that a country has a comparative advantage in stage 2 (downstream) production relative to the U.S. Scanning the table, most countries have comparative advantage in downstream production relative to the US. This inferred comparative advantage is naturally related to export composition in the data. In Figure 1, we plot the share of final goods in exports against relative stage 2 productivity. There is an evident positive correlation, wherein countries with high productivity in stage 2 relative to stage 1 have higher shares of final goods in their exports. Since this is not a moment that we have directly targeted, this is a useful plausibility check on the technology estimates. Further, while comparative advantage has an obvious influence on export composition, it is not the only determinant of it. For example, Austalia has comparative advantage in stage 2 production relative to the US, but has a significantly lower share of final goods in exports than does the US.

Our estimate of the elasticity of trade costs to distance for the manufacturing sector is $\rho(m) = 0.21$. Combining this distance elasticity with estimates of τ^j, we compute implied iceberg frictions for manufacturing and plot the distribution of trade costs across bilateral pairs in Figure 2. Clearly, estimated iceberg frictions are both large on average (with mean value of 3.4) and heterogeneous across partners (ranging from roughly 2 to 8) . The high

[21] Since $T_s^i(z,m)$ is drawn from a Fréchet, the unweighted geometric mean is given by $\exp(\gamma/\kappa)T_s^i(m)^{1/\kappa}$, where γ here is the Euler-Mascheroni constant. Since we report means for each country relative to the U.S., the numbers reported in the table are effectively $\left(T_s^i(m)/T_s^{US}(m)\right)^{1/\kappa}$.

level of these trade cost estimates is broadly in line with findings in the previous literature.[22]

As one would expect, the level and heterogeneity in bilateral trade costs is tightly linked to variation in sourcing shares. For one, the level of trade costs is highly correlated with the home bias in each country's expenditure. Second, bilateral frictions are needed to match the wide variation in import shares across partners. To illustrate this variation in the data and model, we plot intermediate and final goods trade shares in Figure 3, with true trade shares on the x-axis and simulated trade shares on the y-axis (both log scales).[23] The model clearly fits both sets of bilateral trade flows well.

3 Trade Elasticities

In this section, we use counterfactual simulations to study how trade elasticities behave in the model. We start with a discussion about how bilateral elasticities vary with the level of trade costs, and then we discuss global trade elasticities.

3.1 Bilateral Trade Elasticities

To compute bilateral trade elasticities in the multistage model that are comparable to existing empirical estimates, we interpret our simulated data through the lens of a general gravity equation. Following Head and Mayer (2014), let bilateral gross exports be given by $X^{ij} = GS^iM^j\phi^{ij}$, where G is a (gravitational) constant, S^i captures the supply capabilities of the exporter, and M^j captures characteristics of the import market. The parameter $\phi^{ij} = (\tau^{ij})^{\zeta^{ij}}$ is the level of bilateral (iceberg) trade frictions raised to the power ζ^{ij}, which measures the elasticity of trade flows to frictions.

In standard applications of the gravity equation, ζ^{ij} is a constant – as in $\zeta^{ij} = \zeta$ for all i and j – pinned down by the elasticity of substitution in the CES-Armington model or the Fréchet shape parameter in an Eaton-Kortum model. We relax this assumption by allowing the elasticity to vary by country pair, while maintaining the assumption that it is symmetric within each country pair ($\zeta^{ij} = \zeta^{ji}$). Then, we solve for the bilateral trade elasticity as follows:

$$\zeta^{ij} = \ln\left(\frac{X^{ij}X^{ji}}{X^{ii}X^{jj}}\right) - \ln\left(\tau^{ij}\tau^{ji}\right). \tag{23}$$

[22]The review by Anderson and van Wincoop (2004) argues that existing estimates implied that trade costs for a representative industrialized country are near 170% ($\tau = 2.7$). Our sample includes emerging markets as well, where trade costs are plausibly higher. Eaton and Kortum (2002) report estimated distance costs of roughly 300% for country pairs in the 3000 to 6000 mile distance range.

[23]The cluster of points in the upper right corner is the share of each country's purchases from itself. Not surprisingly, these own shares are uniformly large.

This approach to computing bilateral elasticities is the analog to the ratio-type estimation of trade costs, developed by Head and Ries (2001).

We compute elasticities using simulated trade data for a sequence of counterfactual equilibria in the multistage model, which differ only in terms of the level of trade costs in each equilibrium. Starting from baseline trade cost estimates $\{\hat{\tau}^{ij}(m), \hat{\tau}^{ij}(n)\}$, we consider counterfactual equilibria with trade costs equal to $\{\delta\hat{\tau}^{ij}(m), \delta\hat{\tau}^{ij}(n)\}$, where $\delta \lesseqgtr 1$ is a scaling factor that raises or lowers the level of trade costs relative to the baseline estimates.[24] For each equilibrium, we use simulated trade flows and trade costs $\{\delta\hat{\tau}^{ij}(m), \delta\hat{\tau}^{ij}(n)\}$ to compute bilateral elasticities, using Equation (23). This procedure is analogous to using measured barriers to trade (e.g., tariffs) and observed trade flows to estimate trade elasticities, as in Caliendo and Parro (2015) for example.

In Figure 4, we plot mean values of ζ^{ij} for the manufacturing sector in the set of counterfactual equilibria, where the mean level of trade costs in each equilibrium is on the x-axis.[25] In the baseline equilibrium (with mean trade costs near 240%), the mean trade elasticity is near 4, close to standard values in the literature. As the level of trade costs falls (rises) from their estimated level, the mean elasticity rises (falls) in the model.[26] Quantitatively, a decline in trade costs from 350% ($\tau = 4.5$) to 150% ($\tau = 2.5$) raises the mean elasticity from about 3.85 to 4.15. This inverse correlation between the trade elasticity and the level of trade costs is the multistage magnification effect in action: bilateral trade appears to become more sensitive to trade frictions as the level of trade costs falls and the extent of production chain fragmentation rises.

Underlying this average elasticity result, there is heterogeneity in elasticities across final goods versus inputs. We estimate elasticities using Equation (23) for final goods and inputs separately, and then plot the the mean values in Figure 5. The first point to note is that the elasticity for inputs is endogenously larger than for final goods in the model. This reflects the fact that most input trade dominated by stage 1 goods, and the elasticity for stage 1 goods is roughly 5 while the elasticity for stage 2 goods is near 4. Given that trade frictions increase with distance, this implies that input trade is concentrated among geographically proximate trade partners, relative to trade in final goods.

[24] Prior to changing trade costs, we close the exogenous aggregate trade imbalances and re-compute the baseline equilibrium with balanced trade (holding all other parameters at their baseline values). We then maintain balanced trade as we compute counterfactual equilibria and compare each of these counterfactuals to this initial balanced trade equilibrium.

[25] The gravity elasticity for non-manufacturing is pinned down by the Fréchet shape parameter, so there is no reason to discuss the trade elasticity for non-manufacturing.

[26] Bilateral elasticities also vary across trade partners in the cross-section. In the baseline equilibrium, the 10th and 90th percentiles for bilateral gravity elasticities are roughly 3.7 and 4.2. Reinforcing the idea that gravity elasticities are a function of trade costs, bilateral elasticities are higher for country pairs with lower bilateral trade costs.

The second point to note is that the input trade elasticity rises as the level of trade costs declines, while the final trade elasticity changes little. Referring back to Figure 4, virtually all the increase in the overall mean bilateral gravity elasticity in the model is explained by this rise in the input elasticity (driven by an increase in the stage 1 trade elasticity), rather than by changes in trade composition. The fact that the elasticity of trade in inputs rises as the level of trade costs falls means that the model can rationalize the "distance puzzle" – the empirical fact that estimated distance elasticities have not declined over time, despite seemingly obvious declines in trade costs [Disdier and Head (2008)]. In fact, the model generates a "distance puzzle" *because* trade costs have fallen.

The rise in the bilateral elasticity of input trade is driven by changes in the structure of global production chains as trade costs fall in the model. To understand these changes, it is useful to distinguish between two types of multistage input trade. On the one hand, some stage 1 goods are shipped to foreign countries and then embodied in stage 2 goods that are directly absorbed there. In this type of production chain, stages are fragmented across countries, but stage 1 imports are not used to produce exports. On the other hand, some stage 1 goods are shipped to foreign countries and then embodied in stage 2 goods that are exported, either back home or to third countries.

Building on this distinction between different types of global production chains, we can decompose stage 1 exports from country i to j as follows:

$$EX_1^{ij}(m) = EX_1^{ij}(m,j) + EX_1^{ij}(m,k \neq j) \tag{24}$$

The first term $EX_1^{ij}(m,j)$ is the value of stage 1 exports from i that are used to produce stage 2 goods that are absorbed directly in j, and is given by:

$$EX_1^{ij}(m,j) = \sum_{r=1}^{R} \left(\frac{e^{-p_1^{ij}(r,m)/\lambda}}{\sum_l e^{-p_1^{lj}(r,m)/\lambda}} \right) \left[\theta(m) p_2^j(r,m) \tilde{q}^j(z,m) \left(\frac{e^{-p_2^{jj}(r,m)/\lambda}}{\sum_l e^{-p_S^{lj}(r,m)/\lambda}} \right) \right].$$

The second term $EX_1^{ij}(m,k \neq j)$ is the value of stage 1 exports from i that are used to produce stage 2 goods that are exported by j:

$$EX_1^{ij}(m,k\neq j) = \sum_{r=1}^{R} \sum_{k\neq j} \left(\frac{e^{-p_1^{ij}(r,m)/\lambda}}{\sum_l e^{-p_1^{lj}(r,m)/\lambda}} \right) \left[\theta(m) \tau^{jk}(m) p_2^j(r,m) \tilde{q}^k(z,m) \left(\frac{e^{-p_2^{jk}(r,m)/\lambda}}{\sum_l e^{-p_S^{lk}(r,m)/\lambda}} \right) \right].$$

In plain English, this term captures how much "round-trip trade" occurs between i and j – i selling stage 1 inputs to j that are embedded in stage 2 goods sold back to i – as well as the role of j as an "export platform" for production chains that originate from country i.

Reflecting this breakdown, we can write $EX_1^{ij}(m, k \neq j) = EX_1^{ij}(m, i) + EX_1^{ij}(m, k \neq i, j)$, where first term is round-trip trade and the second is export platform trade.

As discussed in Section 1.3, trade costs deter organizing the production chain in a way that involves stage 2 being performed in a country where neither stage 1 is performed, nor where stage 2 output is consumed. The reason is that the cost savings of locating stage 2 abroad, rather than at home or where stage 2 output is consumed, will tend to be small relative to the gross trade costs incurred in exporting stage 2 output. Lower trade costs make it more profitable to exploit cost differences in locating stage 2 production. As a result, we should expect to see an increase in both round-trip and export platform trade in stage 1 goods as trade costs fall.

Building on these ideas, we plot the share of stage 1 exports that are absorbed in the destination, embedded in round-trip trade, or embedded in platform exports against the mean level of trade costs in each counterfactual equilibrium in Figure 6. When trade costs are high, stage 1 trade is dominated by input shipments that are absorbed in the destination. As trade costs fall, a larger share of stage 1 goods are dedicated to round-trip trade and platform exporting.[27] In particular, the rise in stage 1 inputs dedicated to platform exporting comes at the expense of direct absorption of stage 1 inputs.

This rise in platform exporting is important for understanding how the elasticity of input trade behaves in the model. Across bilateral export destinations in the cross-section, stage 1 inputs dedicated to platform exports fall off quickly with respect to bilateral trade costs. That is, export platforms for production chains originating in country i tend to be located in countries j for which $\tau^{ij}(m)$ is low (e.g., countries that are close to country i itself). As the level of trade costs falls and platform exporting increases, the composition of stage 1 exports thus shifts toward the type of trade that is more sensitive to trade costs. The result is that the measured elasticity of trade increases.

To illustrate these mechanics, we turn to Figure 7. For each equilibrium in our model, we bin country pairs based on the quintiles of the distribution of bilateral trade costs. Because relative trade costs are fixed in the counterfactuals, the mapping from country pairs to quintile is stable across equilibria. For each quintile q, we compute log stage 1 trade broken down by destination use, given by $\ln(\sum_{ij\in q} EX_1^{ij}(m, j))$ and $\ln(\sum_{ij\in q} EX_1^{ij}(m, k \neq j))$, where this second summation pools round-trip and platform exports. We plot these values against the log of the mean trade cost in each quintile in Figure 7, for both the highest and lowest trade cost equilibria.

[27] Focusing on the levels, round-trip trade is relatively uncommon in the model. In our data, this type of trade only occurs in substantial amounts among the lowest trade cost pairs (e.g., France-Germany, US-Canada, China-Korea, etc.). If we were to consider a different set of countries (e.g., European countries only), we would expect to see round-trip trade playing a larger role.

The slope of the line that connects the quintile values for each series corresponds (roughly speaking) to the magnitude of the bilateral elasticity for each type of trade. In both equilibria, stage 1 exports associated with round-trip trade and platform exporting fall off faster with respect to distance than do stage 1 exports that are absorbed in the destination. Further, comparing the right figure to the left, as trade costs fall, stage 1 exports dedicated to round-trip and export platform trade rise relative to those absorbed in the destination. The average response of stage 1 exports with respect to distance (again, roughly speaking) is the weighted mean slope of the two series. The mean slope is evidently higher in the low trade cost equilibrium, primarily because round-trip and platform trade becomes more important for low trade cost partners as trade costs fall. In the end, this endogenous supply chain re-organization drives the rise in the overall bilateral trade elasticity as trade costs fall.

3.2 World Trade Elasticities

We now turn to the elasticity of trade at the global level. Whereas the bilateral gravity elasticities discussed above were interpretable as partial equilibrium elasticities, we describe world trade elasticities in general equilibrium. Due to this shift in emphasis, we now need to specify a particular non-multistage model to serve as a benchmark against which we can evaluate elasticities in the multistage model.

The benchmark we use is a multi-sector Eaton-Kortum model with an input-output structure. Henceforth, we refer to this benchmark model as the EK-CP model, because it is closely related to the Eaton-Kortum model developed by Caliendo and Parro (2015). Following Dekle et al. (2008), we compute exact changes in the model's endogenous variables relative to initial equilibrium values. We parameterize the initial equilibrium so that it exactly matches simulated data on bilateral trade shares, income, sector-level production and expenditure, and sector-level input cost shares from our multistage model.[28] We set the Fréchet technology shape parameters to 3.85 for manufacturing and 4.08 in non-manufacturing, based on sector-level regressions of simulated bilateral trade from the multi-stage model on estimated log bilateral trade costs. Additional specification, calibration, and solution details are discussed in Appendix B.

We compute equilibria in the multistage and EK-CP models for different levels of trade costs, as in Section 3.1. Then, we define the world trade elasticity in each equilibrium as $\Delta\ln(EX)/\Delta\ln(\tau)$, where $\Delta\ln(EX)$ is the log change in total world manufacturing exports from one equilibrium the next and $\Delta\ln(\tau) = \ln(\delta)$ is pinned down by the scaling factor for trade costs across equilibria.

[28] We discard data on shipments between production stages generated by our model in calibrating the EK-CP model, since the meaning of a production stage is undefined in that model.

In Figure 8, we plot the world trade elasticity against the level of trade costs in each counterfactual equilibrium. The first point to note in the figure is that the world trade elasticity declines in both models as the level of the trade costs falls. It is useful to appeal to the EK-CP model to interpret this result. While the partial equilibrium elasticity of trade is constant in the EK-CP model, the general equilibrium elasticity is not. The most important reason is that the general equilibrium elasticity is computed allowing for the aggregate price level to respond to changes in frictions. As trade costs fall, the importance of imports in determining aggregate prices rises, and thus the aggregate price level becomes more sensitive to marginal changes in trade costs. This pulls down the general equilibrium trade elasticity as trade costs fall in the EK-CP model, and this force is active in the multistage model as well.[29]

The second point to note is that the general equilibrium trade elasticity is actually lower in the multistage model than in the EK-CP model in the baseline equilibrium, with $\tau \approx 3.4$.[30] As the level of trade costs falls, world trade elasticities converge in the alternative models and then cross when mean $\tau \approx 2.8$.. This crossing reflects the role of multistage production. Multistage production alone tends to raise trade elasticities as trade costs fall, as production fragmentation is more sensitive to marginal changes in trade costs when the level of trade costs is low. This multistage production effect competes with general equilibrium forces that pull down the trade elasticity as trade costs fall, attenuating the fall in the general equilibrium elasticity in the multistage model relative to the EK-CP model.

Quantitatively, the value of the world trade elasticities is similar across models. Given this, the ratio of manufacturing trade to GDP also behaves similarly across models. We plot the ratio of manufacturing exports to GDP for both models in Figure 9. The two series match up at the baseline (estimated) level of trade costs by construction, and then track each other closely as the level of trade costs rises or falls from baseline. Further, large changes in trade costs are needed to generate substantial changes in the ratio of trade to GDP in both models – a decline in trade costs from 250% to 150% roughly doubles the trade to GDP ratio. Finally, note that the relationship is convex in both models, such that marginal changes in trade costs have larger impacts on the trade to GDP ratio as the level of trade costs falls.

Discussion These results allow us to revisit the conclusions of Yi (2003, 2010) on the quantitative importance of multistage production in explaining the response of trade to frictions. While our results are generally consistent with the idea that multistage production magnifies trade elasticities, the quantitative strength of these magnification forces is relatively

[29] See also related discussion of partial versus general equilibrium elasticities in Head and Mayer (2014).

[30] While the partial equilibrium elasticities in the EK-CP model are set to match simulated bilateral trade flows given frictions, the global elasticity is an untargeted moment.

weak in our model. For one, global trade responds similarly to changes in trade frictions in the multistage and the EK-CP models at the global level. This result is surprising, given the argument advanced by Yi (2003), who argues that fragmentation leads to large trade effects for small tariff changes. In our appraisal, there are two principal reasons why our conclusions about the importance of multistage production differ from Yi's results.

The first reason is that the benchmark against which we evaluate the importance of multi-stage production is different. Whereas Yi (2003) compares a multistage model to a single-stage model *without input trade*, we instead compare our multistage model to a single-stage model *with input trade* (the EK-CP model). This distinction is important. In a model without input trade, there is no scope for double counting in gross trade data, because goods (at most) cross borders once en route from source to destination. In contrast, models with input trade – either the multistage model or the EK-CP model – allow for goods to cross borders multiple times as they move through the production process, which inflates the value of gross trade relative to GDP in the model.

One way to read the results in Yi (2003) is that they point out that the double counting generated by input trade is important in explaining trade growth. That is, the multistage model in Yi (2003) generates larger increases in trade relative to GDP than the single-stage model (without input trade) precisely because it generates double counting in trade. While we agree on this point, it does not follow that multistage production itself – as opposed to other ways of modeling input trade (e.g., roundabout production) – is essential for explaining increases in trade relative to GDP. In our analysis, we compare two models that both allow for double counting, and thus both generate inflated values for gross trade relative to GDP as trade costs fall. The similarity in trade-to-GDP ratios in the multistage and EK-CP models indicates that these models yield additional double counting in trade at approximately the same rate, so both generate non-linear responses in the trade-to-GDP ratio as trade costs decline.

The second reason our results differ relative to Yi (2003) is that we focus on the behavior of the model around equilibria with high trade costs, while both Yi (2003) and Yi (2010) examine multistage models with low levels of trade costs. We choose to focus on these high trade cost equilibria precisely because we estimate trade costs in the multistage model to match trade shares; the model requires a high level of trade costs to match the high degree of home bias in expenditure observed in the data. The level of trade costs matters for comparative statistics, because the multistage model features non-linear responses of production chain fragmentation to trade costs. That is, fragmentation and trade flows are less sensitive to marginal changes in trade frictions at high levels of trade costs than at low levels of trade costs. Referring back to Yi (2003), we point out that Yi examines a liberalization scenario in his model with a very low initial level of trade costs (tariffs under 15%).[31] If one introduces the same tariff reduction (in either his model or ours) from a higher initial level of trade costs, the model implies a smaller increase in the trade-to-GDP ratio.

4 Conclusion

Despite substantial academic and policy interest in the rise of global supply chains, few quantitative models incorporate multistage production chains. In contrast, this paper puts the decision to collocate or fragment production stages at center stage. This allows us to quantify the role of technology and trade costs in driving fragmentation and the role of supply chain structure in shaping trade elasticities.

We found that there are sizable differences in upstream versus downstream comparative advantage across countries, and that these are an important driver of the final versus input composition of exports. We also found that the response of supply chain structure to frictions plays an important role in determining the endogenous bilateral elasticity of input trade. As trade costs decline, the input trade elasticity rises, leading the bilateral trade elasticity to increase as well. Nonetheless, the multistage model generates aggregate world trade elasticities that are quantitatively similar to more standard Ricardian models with input trade, at least for modest changes in trade costs from their current level.

Digging beneath these aggregate results, global production chains play an interesting conceptual role in explaining how trade costs map into trade flows. Changes in trade frictions induce supply chain re-organization, leading to the growth of more complex export platform sourcing strategies. This points to a broader message: multistage models generate a variety of predictions regarding the micro-structure of supply chains that have yet to be explored fully. We expect that combining the type of model we have written down here with micro-data on supply chain structure would be a fruitful path for future empirical work.

[31] A subtle point is that Yi focus entirely on growth in trade rather than trade levels in his analysis, by normalizing simulated trade flows to match the levels in data (see footnote 32 of Yi (2003)). That is, his model cannot match the import share given this low level of trade costs.

References

Anderson, J. E. and E. van Wincoop (2004). Trade costs. *Journal of Economic Literature XLII*, 691–751.

Antràs, P. and A. de Gortari (2016). On the geography of global value chains. Unpublished Manuscript, Harvard University.

Arkolakis, C. and A. Ramanarayanan (2009). Vertical specialization and international business cycle synchronization. *Scandinavian Journal of Economics 111*(4), 655–680.

Baldwin, R. and A. Venables (2013). Spiders and snakes: Offshoring and agglomeration in the global economy. *Journal of International Economics 90*(2), 245–254.

Bridgman, B. (2008). Energy prices and the expansion of world trade. *Review of Economic Dynamics 11*, 904–916.

Bridgman, B. (2012). The rise of vertical specialization. *Journal of International Economics 86*, 133–140.

Caliendo, L. and F. Parro (2015). Estimates of the trade and welfare effects of NAFTA. *Review of Economic Studies 82*(1), 1–44.

Costinot, A., J. Vogel, and S. Wang (2013). An elementary theory of global supply chains. *Review of Economic Studies 80*(1), 109–144.

Dekle, R., J. Eaton, and S. Kortum (2008). Global rebalancing with gravity: Measuring the burden of adjustment. *IMF Staff Papers 55*(3), 511–540.

Dietzenbacher, E., B. Los, R. Stehrer, M. Timmer, and G. de Vries (2013). The construction of world input-output tables in the WIOD project. *Economic Systems Research 25*(1), 71–98.

Disdier, A.-C. and K. Head (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics 90*(1), 37–48.

Dixit, A. K. and G. M. Grossman (1982). Trade and protection with multistage production. *Review of Economic Studies 49*(2), 583–594.

Dornbusch, R., S. Fischer, and P. A. Samuelson (1977). Comparative advantage, trade, and payments in a Ricardian model with a continuum of goods. *The American Economic Review 67*(5), 823–839.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Fally, T. and R. Hillberry (2015). A Coasian model of international production chains. NBER Working Paper No. 21520.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4, Chapter 3, pp. 131–195. Elsevier B.V.

Head, K. and J. Ries (2001). Increasing returns versus national product differentiation as an explanation for the pattern of U.S.-Canada trade. *The American Economic Review 91*(4), 858–876.

Irarrazabal, A., A. Moxnes, and L. D. Opromolla (2015). The tip of the iceberg: A quantitative framework for estimating trade costs. *Review of Economics and Statistics 97*(4), 777–792.

Markusen, J. R. and A. J. Venables (2007). Interacting factor endowments and trade costs: A multi-country, multi-good approach to trade theory. *Journal of International Economics 73*(2), 333–354.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995–1026.

Sanyal, K. K. (1983). Vertical specialization in a Ricardian model with a continuum of stages of production. *Economica 50*(197), 71–78.

Simonovska, I. and M. Waugh (2014). The elasticity of trade: Estimates and evidence. *Journal of International Economics 92*(1), 34–50.

Train, K. (2009). *Discrete Choice Methods with Simulation* (Second ed.). New York: Cambridge University Press.

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

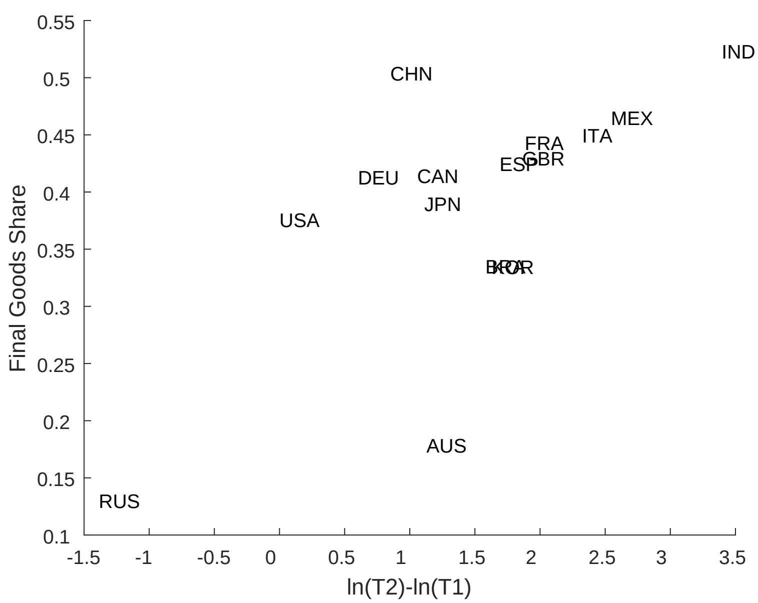
Yi, K.-M. (2010). Can multistage production explain the home bias in trade? *The American Economic Review 100*(1), 364–393.

Table 1: Estimated Manufacturing Technology

	Stage 1	Stage 2	$\frac{Stage2}{Stage1}$
Russia	0.32	0.23	0.71
United States	1.00	1.00	1.00
Germany	0.84	0.97	1.16
China	0.33	0.41	1.23
Canada	0.54	0.70	1.29
Japan	0.75	0.98	1.31
Australia	0.48	0.63	1.32
Brazil	0.27	0.40	1.47
South Korea	0.46	0.68	1.49
Spain	0.50	0.75	1.51
United Kingdom	0.56	0.88	1.57
France	0.62	0.98	1.58
Italy	0.55	0.96	1.76
Mexico	0.27	0.50	1.85
India	0.13	0.29	2.28

Note: Columns labeled Stage 1 and Stage 2 report the geometric mean of the Fréchet distribution for each manufacturing stage, relative to the geometric mean of the US distribution. Relative productivity (Stage 2/Stage 1) measures comparative advantage across stages relative to US comparative advantage.

Figure 1: Comparative Advantage across Stages and Export Composition



Note: The y-axis reports the share of final goods in manufacturing exports in WIOD data. The x-axis reports the log difference between estimated stage 2 and stage 1 Fréchet location parameters for manufacturing in the multistage model.

Figure 2: Distribution of Estimated Bilateral Iceberg Trade Costs for Manufacturing

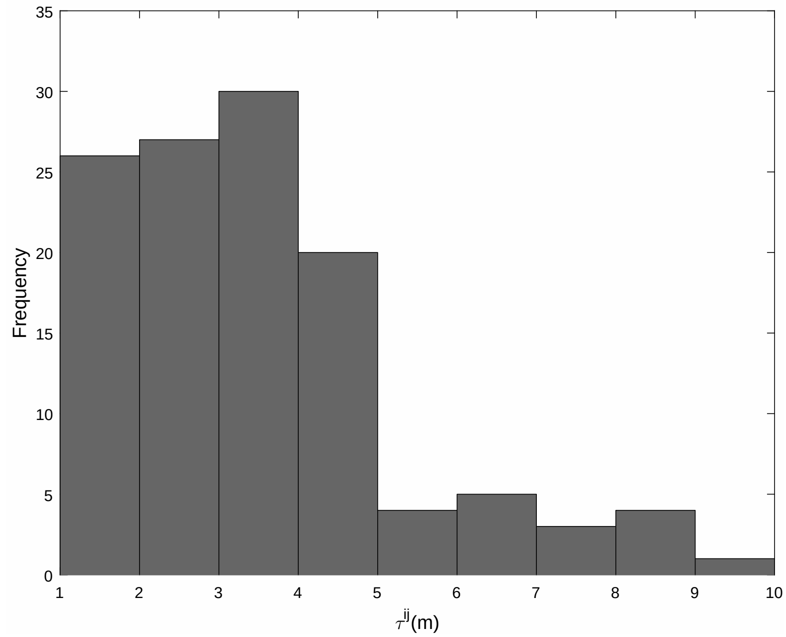
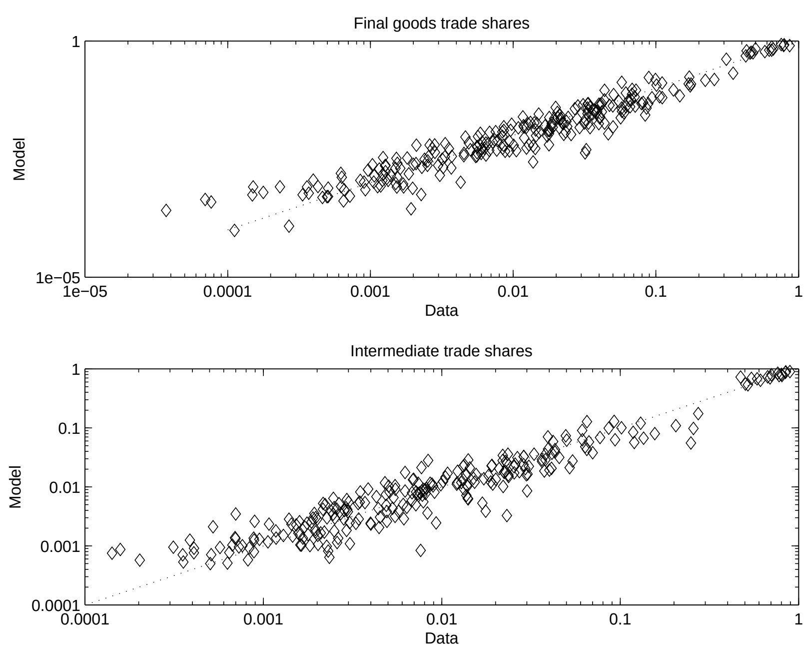
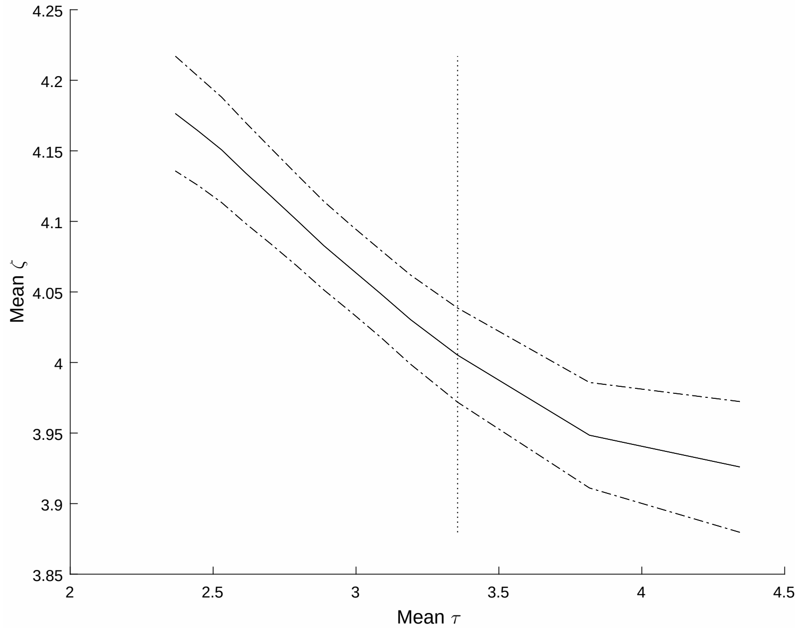


Figure 3: Manufacturing Trade Shares in Model and Data



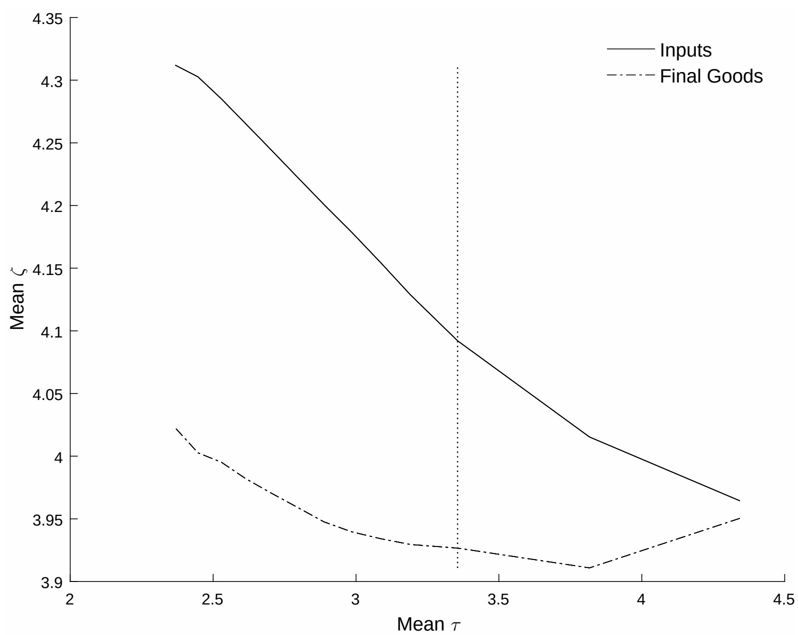
Note: Fitted trade shares from the estimated model are on the y-axis, and target shares from WIOD data are on the x-axis. Axes are log scale, because the objective function minimizes log differences in trade shares in model versus data.

Figure 4: Mean Gravity Trade Elasticity for Multistage Manufacturing



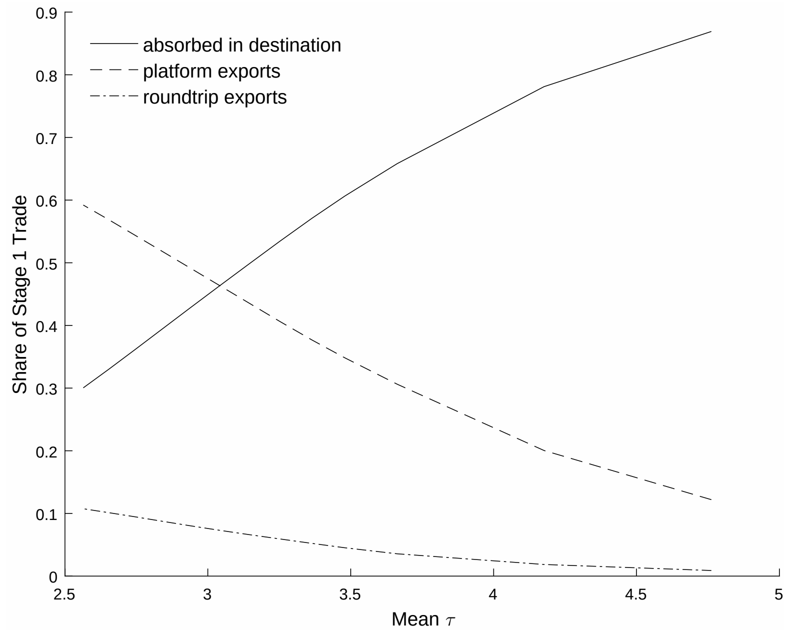
Note: Figure reports the simple mean (with 95% confidence interval) of the bilateral gravity elasticities ζ^{ij}, defined in Equation (23), for equilibria of the multistage model with different levels of trade costs. The mean level of trade costs in each equilibrium is on the x-axis. Vertical dashed line indicates baseline equilibrium with estimated trade costs.

Figure 5: Mean Gravity Trade Elasticity for Final Goods vs. Inputs



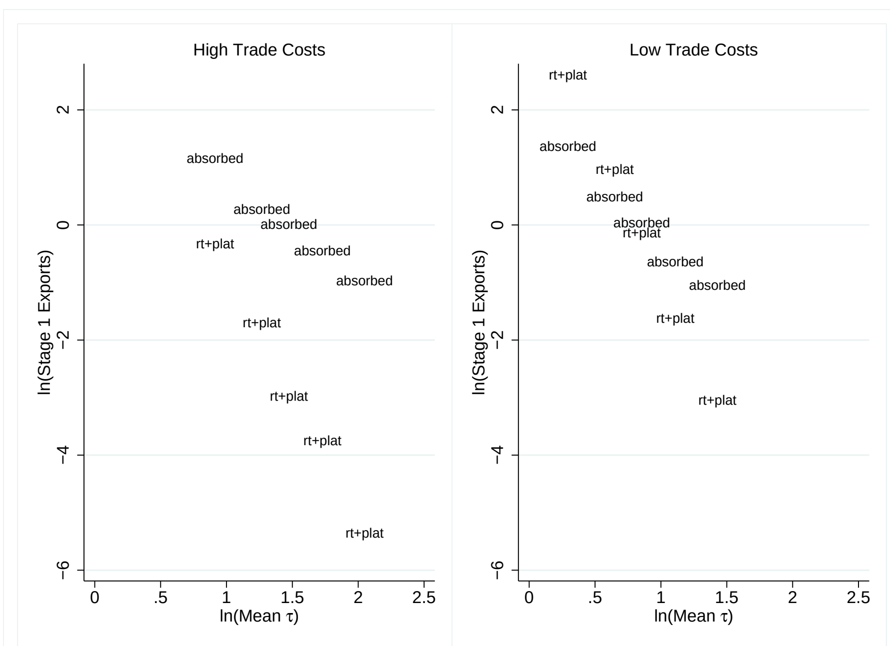
Note: Figure reports the simple mean (with 95% confidence interval) of the bilateral gravity elasticities ζ^{ij} computed for final goods and inputs separately. The mean level of trade costs in each equilibrium is on the x-axis. Vertical dashed line indicates baseline equilibrium with estimated trade costs.

Figure 6: Decomposition of Stage 1 Exports by Destination Use



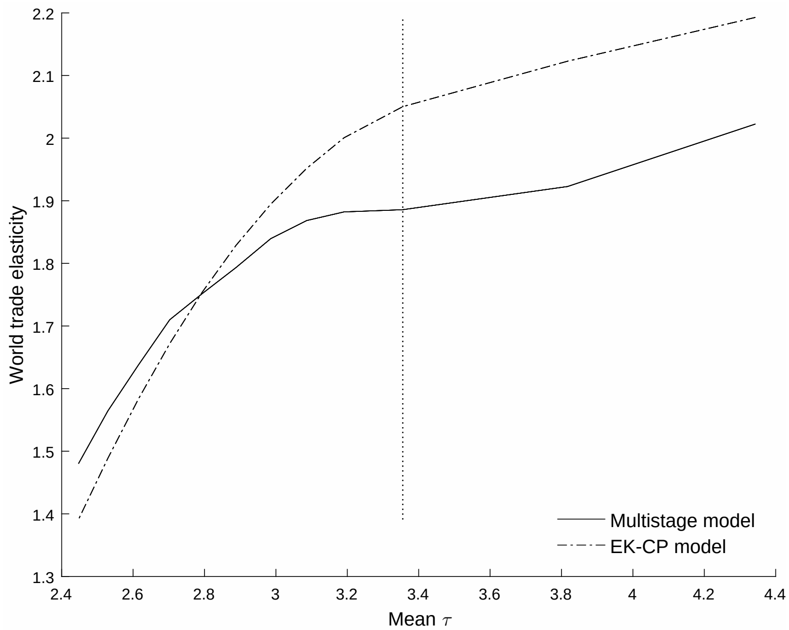
Note: "Absorbed in destination" is the share of bilateral stage 1 exports that are embodied in stage 2 goods absorbed in the destination. "Platform exports" is the share of bilateral stage 1 exports that are embodied in stage 2 goods that are exported to third destinations. "Roundtrip exports" is the share of bilateral stage 1 exports that are embodied in stage 2 goods that are exported back to the stage 1 source. The mean level of trade costs in each equilibrium is on the x-axis.

Figure 7: Decomposing Stage 1 Exports by Destination Use and Bilateral Trade Costs



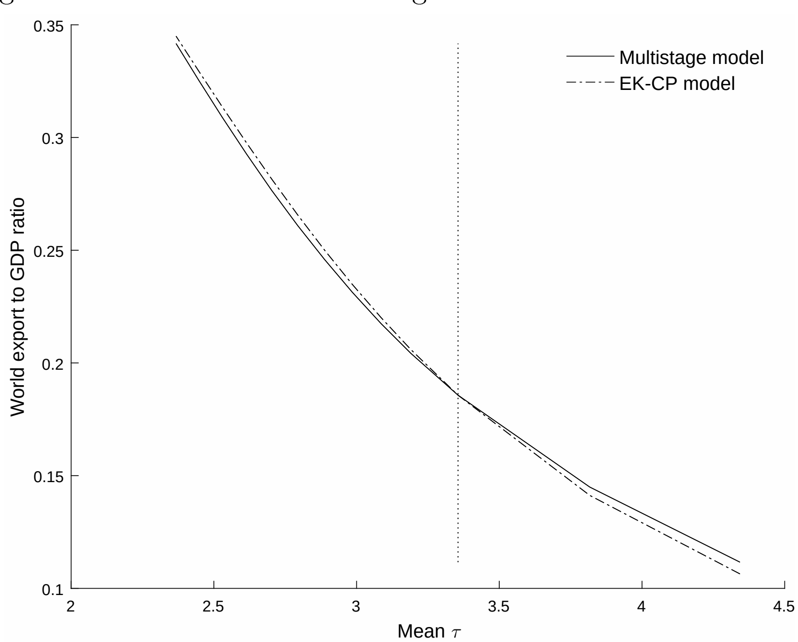
Note: Each point in the figure represents the mean trade cost and log value of stage 1 exports broken down by destination use for country pairs in a given quintile of the distribution of bilateral trade costs. Formally, for quintile q the data are absorbed $= \ln(\sum_{ij \in q} EX_1^{ij}(m,j))$ and rt+plat $= \ln(\sum_{ij \in q} EX_1^{ij}(m,k \neq j))$ on the y-axis. The left figure is for data from an equilibrium with "high trade costs" and the right reports results from an equilibrium with "low trade costs."

Figure 8: The World Manufacturing Trade Elasticity



Note: World trade elasticity is the log change in world trade with respect to the log change in the level of world trade costs. Elasticities for manufacturing trade are reported for the multistage model and Ricardian model with input-output linkages (EK-CP model). The mean level of trade costs in each equilibrium is on the x-axis. Vertical dashed line indicates baseline equilibrium with estimated trade costs.

Figure 9: Ratio of Manufacturing Trade to GDP for the World



Note: World export to GDP ratio is the ratio of world manufacturing trade to world manufacturing GDP. The ratio is reported for the multistage model and Ricardian model with input-output linkages (EK-CP model). The mean level of trade costs in each equilibrium is on the x-axis. Vertical dashed line indicates baseline equilibrium with estimated trade costs.

A Solving the Model

In this appendix, we provide details concerning the algorithms we use to solve the model.

Picking up on the discussion in Section 1.4, we solve a discrete approximation to the model with a continuum of manufacturing goods, using a smoothing technique borrowed from the discrete choice literature in the process. We proceed here assuming that we know the model parameters $\{\alpha_i, \theta(s), \beta, T^i(n), T^i_s(m), \kappa, \tau^{ij}(m), \tau^{ij}(n)\}$, have realized productivity draws $\{T^i_s(r,m)\}$ for $r = 1, \ldots, R$ in hand, and data on exogenous factor endowments and trade balances $\{L^i, TB^i\}$.

For a given an initial value for the vector of wages $\{w^i\}$, we solve for the optimal assignment of stages to countries and hence prices of manufactured goods $\{\tilde{p}^k_2(r,m)\}$. One complication in doing so is that the composite input price P^i_X is a function of these prices, and simultaneously the cost of producing manufactured goods depends on P^i_X itself. In addition, P^i_X is also a function $P^i(n)$, which depends on production costs in the non-manufacturing sector and P^i_X due to the input loop. In the end, this problem has a fixed point structure.

Starting with a guess for the vector of composite input prices $\{\dot{P}^i_X\}$, we calculate the the optimized stage 1 input price that would prevail in each country j if it were to produce stage 2 output:

$$\tilde{p}^j_1(r,m) = \min_i \tau^{ij}(m) p^i_1(r,m), \quad \text{with} \quad p^i_1(r,m) = \frac{\left(w^i\right)^{1-\theta(m)} \left(\dot{P}^i_X\right)^{\theta(m)}}{T^i_1(r,m)}.$$

Then we compute the optimized price at which country k purchases stage 2 goods, given decisions optimized decisions at stage 1:

$$\tilde{p}^k_2(r,m) = \min_j \tau^{jk}(m) p^j_2(r,m), \quad \text{with} \quad p^j_2(r,m) = \frac{\left(w^j\right)^{1-\theta(m)} \left(\tilde{p}^j_1(r,m)\right)^{\theta(m)}}{T^j_2(r,m)}.$$

This yields $\{\tilde{p}^k_2(r,m)\}$, assuming that $\dot{P}^i_X$ is the composite input price, and tracing backwards the optimal location of stage 2 and stage 1 production for serving each destination k.

With these prices, we construct the composite price of manufactured goods as $P^i(m) = \exp\left(\frac{1}{R}\sum_r \log(\tilde{p}^k_2(r,m))\right)$, and use Equation (16) to compute $P^i(n)$. With these, we construct an updated value for the composite input price: $\ddot{P}^i_X = P^i(m)^{\beta} P^i(n)^{1-\beta}$. We iterate on these steps until $\ddot{P}^i_X = \dot{P}^i_X$.

Having converged on a value for P^i_X, we can easily compute the solution for all equilibrium prices, as well as the allocation of stages to countries for manufactured goods. The next step is to compute equilibrium quantities. Total demand for the sector-level composite goods is given by:

$$P^k(m)Q^k(m) = \alpha_k P_F^k F^k + \beta P_X^k X^k$$
$$P^k(n)Q^k(n) = (1-\alpha_k) P_F^k F^k + (1-\beta) P_X^k X^k.$$

Since we have taken the wage as given and treat the trade balance as an exogenous parameter, we can compute final demand as $P_F^k F^k = w^k L^k - TB^i$. We cannot immediately compute expenditure on the composite input, because we do not yet know X^k. However, we can solve for it as follows. Given a guess for $\dot{X}^k$, we can compute $P^k(m)Q^k(m)$ and$P^k(n)Q^k(n)$. Demand for stage 2 goods in manufacturing in destination k is then:

$$\tilde{q}^k(r,m) = \frac{\frac{1}{R} P^k(m)Q^k(m)}{\tilde{p}_2^k(r,m)}.$$

Tracing these demands back to the countries that supply those goods, the quantity of stage 2 goods produced in each source j is:

$$q_2^j(r,m) = \sum_k \tau^{jk}(m)\tilde{q}^k(r,m)\left(\frac{e^{-p_2^{jk}(r,m)/\lambda}}{\sum_k e^{-p_2^{jk}(r,m)/\lambda}}\right).$$

Given this stage 2 production in country j, demand for stage 1 inputs in country j is:

$$x_1^j(r,m) = \frac{\theta(s)p_2^j(r,m)q_2^j(r,m)}{\tilde{p}_1^j(r,m)}.$$

These input demands allow us to then solve for the quantity of each stage 1 good supplied by country i as:

$$q_1^i(r,m) = \sum_j \tau^{ij}(m)x_1^j(r,m)\left(\frac{e^{-p_1^{ij}(r,m)/\lambda}}{\sum_k e^{-p_1^{kj}(r,m)/\lambda}}\right).$$

Finally, we can compute an updated value for demand for the composite input:

$$\ddot{X}^i = \frac{1}{P_X^i}\left[\sum_r \theta(m)p_1^i(r,m)q_1^i(r,m) + \theta(n)\sum_j \pi^{ij}(n)P^j(n)Q^j(n)\right],$$

where $\pi^{ij}(n)$ is computed as in Equation (15). We iterate on this fixed point problem until $\ddot{X}^i = \dot{X}^i$. Upon convergence, we have the entire equilibrium for a given wage vector.

Lastly, we need to check whether the wage vector clears the labor market. We can calculate labor demand as:

$$
\begin{aligned}
L_D^i(\mathbf{w}) &= \int_0^1 l^i(z,n)dz + \sum_r \left[l_1^i(r,m) + l_2^i(r,m)\right] \\
&= \frac{1}{w^i}(1-\theta(n)) \sum_j \pi^{ij}(n) P^j(n) Q^j(n) \\
&\quad + \frac{1}{w^i}(1-\theta(m)) \sum_r \left(p_1^i(r,m)\, q_1^i(r,m) + p_2^i(r,m)\, q_2^i(r,m)\right)
\end{aligned}
$$

The equilibrium wage vector then sets labor demand equal to labor supply: $L_D^i(\mathbf{w}) = L^i$ for $i = 2, .., N$ (where market 1 is dropped appealing to Walras' law).

B Benchmark Ricardian Trade Model with Input-Output Linkages

This appendix describes the benchmark two-sector Ricardian model against which we evaluate the multistage model. We present the key equilibrium conditions here and refer the reader to Caliendo and Parro (2015) for details regarding this class of models.

We define $E^i(m)$ and $E^i(n)$ to be total spending on final goods plus intermediates goods from the manufacturing and non-manufacturing sectors. Otherwise, the notation used here matches that use in the main text, with slight (obvious) modifications in the meaning of variables as necessary. For example, $c^i(s)$ denotes unit costs and $P^i(s)$ denotes an aggregate price level of an aggregate of sector s goods, but the functional forms are different here than in the main text reflecting differences between this model and the multi-stage model. The equilibrium of the model can be written as:

$$c^i(m) = (w^i)^{1-\theta(m)}\left[P^i(m)^{\gamma^i(m)}P^i(n)^{1-\gamma^i(m)}\right]^{\theta(m)} \tag{25}$$

$$c^i(n) = (w^i)^{1-\theta(n)}\left[P^i(m)^{\gamma(n)}P^i(n)^{1-\gamma(n)}\right]^{\theta(n)} \tag{26}$$

$$P^i(m) = \left[\sum_j T^j(m)\left(c^j(m)\tau^{ji}(m)\right)^{-\bar{\kappa}(m)}\right]^{-1/\bar{\kappa}(m)} \tag{27}$$

$$P^i(n) = \left[\sum_j T^j(n)\left(c^j(n)\tau^{ji}(n)\right)^{-\bar{\kappa}(n)}\right]^{-1/\bar{\kappa}(n)} \tag{28}$$

$$\pi^{ij}(m) = T^i(m)\left[\frac{c^i(m)\tau^{ij}(m)}{P^j(m)}\right]^{-\bar{\kappa}(m)} \tag{29}$$

$$\pi^{ij}(n) = T^i(n)\left[\frac{c^i(n)\tau^{ij}(n)}{P^j(n)}\right]^{-\bar{\kappa}(n)} \tag{30}$$

$$E^i(m) = \sum_s \gamma^i(s)\theta(s)\left[E^i(s) + TB^i(s)\right] + \alpha_i P_F^i F^i \tag{31}$$

$$E^i(n) = \sum_s (1-\gamma^i(s))\theta(s)\left[E^i(s) + TB^i(s)\right] + (1-\alpha_i) P_F^i F^i \tag{32}$$

$$E^i(m) = \sum_j \pi^{ij}(m)E^j(m) \tag{33}$$

$$E^i(n) = \sum_j \pi^{ij}(n)E^j(n) \tag{34}$$

$$P_F^i F^i = w^i L^i \tag{35}$$

There are several new parameters here. The parameters $\{\gamma^i(m), \gamma(n)\}$ are Cobb-Douglas input shares, equal to the share of input expenditure that each sector dedicates to inputs from sector m. The parameter $\bar{\kappa}(s)$ is a sector-specific trade elasticity. The parameters $\theta(s)$ and α_i are defined as in the main text. Lastly, note that the equilibrium above features balanced trade, to be consistent with the balanced trade assumption imposed in simulations of the multistage model.

Following Dekle et al. (2008), the equilibrium system of equations can be re-written in terms of changes relative to an initial equilibrium. Defining $\hat{x} \equiv \frac{x'}{x}$, where x' is the value of a variable in the new equilibrium and x is the value in the initial equilibrium, the equilibrium in changes is:

$$\hat{c}^i(m) = (\hat{w}^i)^{1-\theta(m)} \left[\hat{P}^i(m)^{\gamma^i(m)} \hat{P}^i(n)^{1-\gamma^i(m)}\right]^{\theta(m)} \tag{36}$$

$$\hat{c}^i(n) = (\hat{w}^i)^{1-\theta(n)} \left[\hat{P}^i(m)^{\gamma(n)} \hat{P}^i(n)^{1-\gamma(n)}\right]^{\theta(n)} \tag{37}$$

$$\hat{P}^i(m) = \left[\sum_j \pi^{ji}(m)\hat{T}^j(m) \left(\hat{c}^j(m)\hat{\tau}^{ji}(m)\right)^{-\bar{\kappa}(m)}\right]^{-1/\bar{\kappa}(m)} \tag{38}$$

$$\hat{P}^i(n) = \left[\sum_j \pi^{ji}(n)\hat{T}^j(n) \left(\hat{c}^j(n)\hat{\tau}^{ji}(n)\right)^{-\bar{\kappa}(n)}\right]^{-1/\bar{\kappa}(n)} \tag{39}$$

$$\hat{\pi}^{ij}(m) = \hat{T}^i(m) \left[\frac{\hat{c}^i(m)\hat{\tau}^{ij}(m)}{\hat{P}^j(m)}\right]^{-\bar{\kappa}(m)} \tag{40}$$

$$\hat{\pi}^{ij}(n) = \hat{T}^i(n) \left[\frac{\hat{c}^i(n)\hat{\tau}^{ij}(n)}{\hat{P}^j(n)}\right]^{-\bar{\kappa}(n)} \tag{41}$$

$$E^i(m)\hat{E}^i(m) = \sum_s \gamma^i(s)\theta(s) \left[E^i(s)\hat{E}^i(s) + TB^i(s)\hat{TB}^i(s)\right] + \alpha_i P_F^i F^i \widehat{P_F^i F^i} \tag{42}$$

$$E^i(n)\hat{E}^i(n) = \sum_s (1-\gamma^i(s))\theta(s) \left[E^i(s)\hat{E}^i(s) + TB^i(s)\hat{TB}^i(s)\right] + (1-\alpha_i) P_F^i F^i \widehat{P_F^i F^i} \tag{43}$$

$$E^i(m)\hat{E}^i(m) = \sum_j \pi^{ij}(m)E^j(m)\hat{\pi}^{ij}(m)\hat{E}^j(m) \tag{44}$$

$$E^i(n)\hat{E}^i(n) = \sum_j \pi^{ij}(n)E^j(n)\hat{\pi}^{ij}(n)\hat{E}^j(n) \tag{45}$$

$$P_F^i F^i \widehat{P_F^i F^i} = w^i L^i \hat{w}^i \hat{L}^i. \tag{46}$$

In all simulations, we assume that labor input is fixed in all countries $\hat{L}^i = 1$, and changes in trade costs $(\hat{\tau}^{ij}(s))$ and technology $(\hat{T}^j(s))$ are exogenous forcing variables. This leaves $10+2N^2$ endogenous variables $\{\hat{w}^i, \widehat{P_F^i F^i}, \hat{c}^i(s), \hat{P}^i(s), \hat{E}^i(s), \pi^{ij}(s)\}$ and $10+2N^2$ equations, before choosing a normalization.[32]

To solve for these endogenous variables, we need parameters $\{\alpha, \bar{\kappa}(s), \gamma^i(m), \gamma(n), \theta(s)\}$ and values for $\{w^i L^i, P_F^i F^i, TB^i, E(s), \pi^{ij}(s)\}$ in an initial equilibrium. We set these parameters based on simulated data generated by the multistage model – i.e., we treat equilibrium values from our estimated model as data, which implies that we start simulations from an "observationally equivalent" equilibrium in the multistage model and Ricardian models. We

[32]Note that we treat nominal final expenditure $P_F^i F^i$ as one variable, hence the wide-hat notation on $\widehat{P_F^i F^i}$. We do not need to separate the final price level and real final expenditure to compute the counterfactuals that interest us.

need $\{P_F^i F^i, E^i(s), \pi^{ij}(s)\}$ in the balanced trade equilibrium of the multistage model, plus the structural parameters $\{\alpha, \bar{\kappa}(s), \gamma^i(m), \gamma(n), \theta(s)\}$, to compute changes in equilibrium variables. We now describe the details regarding how we obtain values for these parameters.

Reading values for $\{P_F^i F^i, E^i(s), \pi^{ij}(s)\}$ from our simulated data is completely straightforward. Further, $\{\theta(s), \alpha_i\}$ are set to the same values as in the multi-stage model. The parameter $\gamma^i(m)$ for sector m (the manufacturing sector) is equal to the value of inputs from sector m used by sector m as a share of total input use by sector m. Denoting the value of stage s output produced by country i in sector m as $y_s^i(m)$, then total input use by sector m is equal to use of stage 1 inputs by stage 2, which are equal to $\theta(m)y_2^i(m)$, plus use of the composite input, which is equal to $\theta(m)y_1^i(m)$. Then all stage 1 inputs used by stage 2 in sector m originate from sector m, but only a fraction (β) of the composite input originates from sector m. This implies that:

$$\gamma^i(m) = \frac{\theta(m)y_2^i(m) + \beta\theta(m)y_1^i(m)}{\theta(m)y_2^i(m) + \theta(m)y_1^i(m)}$$

The value of this parameter varies across countries to the extent that the mix of stage 1 versus stage 2 output varies across countries. Turning to $\gamma(n)$, sector n uses sector m inputs only embodied in the composite input, and the composite input itself is the only input in production. Therefore, $\gamma(n) = \beta$ from the multistage model.

Finally, turning to the values for $\bar{\kappa}(s)$, we note that these values correspond to the elasticity of log bilateral trade to log bilateral trade costs in this model. Therefore, we obtain them by regressing simulated sector-level bilateral trade from the multistage model on log bilateral trade costs. For reference, this returns estimates $\bar{\kappa}(m) = 3.85$ and $\bar{\kappa}(s) = 4.08$.