WORKING PAPERS
RESEARCH DEPARTMENT

WORKING PAPER NO. 08-12
CAN MULTI-STAGE PRODUCTION EXPLAIN THE HOME BIAS IN TRADE?

Kei-Mu Yi
Federal Reserve Bank of Philadelphia

June 2008

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA
Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/econ/index.html

Can Multi-Stage Production Explain the Home Bias in Trade?

Kei-Mu Yi[1]

June 2008

[1]Research Department, Federal Reserve Bank of Philadelphia, 10 Independence Mall, Philadelphia, PA 19106; Kei-Mu.Yi@phil.frb.org. The previous version of this paper was called "Vertical Specialization and the Border Effect." I thank George Alessandria, Scott Baier, Gadi Barlevy, Chad Bown, Satyajit Chatterjee, Zhiqui Chen, Ed Dozier, Jonathan Eaton, Caroline Freund, Gordon Hanson, Rebecca Hellerstein, Russell Hillberry, Tom Holmes, David Hummels, Aubhik Khan, Narayana Kocherlakota, Sam Kortum, Ayhan Kose, Jim MacGee, Philippe Martin, Thierry Mayer, Esteban Rossi-Hansberg, Stanislaw Rzeznik, Eric van Wincoop, Neil Wallace, John Whalley, and participants at the SED Meetings, AEA Meetings, Pittsburgh, Dartmouth Summer Trade Camp, San Francisco Fed, Stanford, Western Ontario, Carleton, Clemson, Duke, Georgia Tech, Emory, Texas, Penn State, NBER ITI workshop, Vanderbilt, Iowa, Midwest Macro Meetings, Chicago Fed, Philadelphia Fed, Villanova, and Michigan for very useful comments. I also thank Keith Head for providing his CUSFTA staging data and John Helliwell and Eric van Wincoop for providing me with their data. Special thanks go to Ron Rioux of Statistics Canada for going to great lengths to prepare data and answer questions. Mohan Anand, Mychal Campos, James Gillard, Matthew Kondratowicz, and especially Edith Ostapik provided excellent research assistance.

Abstract

A large empirical literature finds that there is too little international trade, and too much intra-national trade to be rationalized by observed international trade costs such as tariffs and transport costs. The literature uses frameworks in which goods are assumed to be produced in just one stage. This paper investigates whether the multi-stage nature of production helps explain the home bias in trade. I show that multi-stage production magnifies the effects of trade costs. I then calibrate a multi-stage production model to the U.S. and Canada. I solve the model with measures of trade costs constructed from data on tariffs, transport costs, and wholesale distribution margins. The model can explain about 3/8 of the Canada border effect; this is three times more than what a calibrated one-stage model can explain. The model also explains a good deal of Canada's vertical specialization trade. Finally, a reverse engineering exercise suggests that the unknown or unobserved component of trade costs is smaller than observed trade costs.

JEL Classification code: F1, F4

Keywords: border effect, multi-stage production, trade costs, U.S.-Canada trade, vertical specialization, calibration, home bias in international trade

1 Introduction

How integrated are goods markets across countries? There is a simple integration benchmark that arises from several well known models of international trade. Consider a frictionless world – international trade costs are zero – in which complete specialization occurs. Then, the fraction of output that is exported by a country equals one minus the country's share of world output.[1] For example, Canada's share of the G7 countries' merchandise production in 1999 was 4 percent. The frictionless benchmark would predict 96 percent of Canada's production should be exported and 4 percent should be sold domestically. In fact, 52 percent was exported and 48 percent was sold domestically. This pattern of international and intra-national trade suggests that Canada may not be highly integrated; there may be large frictions or costs inhibiting Canada's international trade.

Indeed, there is a large empirical literature that examines the extent of integration. This research has studied many sets of countries and regions and many time periods, and has repeatedly found a large "border effect". There is too little international trade and too much intra-national trade to be rationalized by the standard, observed measures of trade costs – tariffs and transport costs, in particular – which add up to about 5 to 10 percent for high-income countries. Anderson and van Wincoop (2004, 2003) is the most theoretically consistent research in this literature; they find that the trade patterns between and within the U.S. and Canada in 1993 can be rationalized only by international trade costs of 91 percent. Obstfeld and Rogoff (2001) summarize these findings as the "home bias in trade" puzzle.[2]

The empirical research mentioned above is based on frameworks in which goods are produced in one stage. All of the value-added occurs in a single country. In reality, however, we know that most goods are produced in multiple, *sequential* stages. Semiconductors, for example, are produced in at least three stages. First, there is the wafer manufacturing stage; then there is the fabrication stage in which circuits are etched onto the wafers; finally, there is the packaging (wafers are cut apart into individual semiconductors) and testing stage. Automobiles are also produced in at least three sequential stages. Are there any manufactured goods that are not produced in multiple stages?

[1] Under incomplete specialization, as long as the share of a country's total purchases of, for example, steel, that are from a particular steel producing country equals the steel producing country's share of world steel output, the result still holds. Consumers are assumed to have identical, homothetic preferences. See Deardorff (1998).

[2] The estimate from Anderson and van Wincoop is for an elasticity of substitution of 5. McCallum (1995) was the first to discover a large border effect. Trade between a pair of Canadian provinces was 22 times larger than trade between an otherwise identical Canadian province - U.S. state pair. Other border effect research includes that by Wei (1996), Helliwell (1998), Anderson and Smith (1999), Nitsch (2000), Head and Ries (2001), Head and Mayer (2002), Hillberry (2002), Evans (2003, 2006), Anderson and van Wincoop (2003), Chen (2004), and Combes *et al.* (2005). This research tends to estimate border effects between 10 and 20. Anderson and van Wincoop (2003) is the most theoretical consistent estimation of the border effect; however, their estimate, 10.5 is fairly close to other estimates.

Of course, in these frameworks, the single stage assumption is not taken to be literally true; rather, the production function can be thought of as a reduced-form amalgam of a multiple, sequential stage process. However, the maintained assumption is that the nature of production – the mapping of production stages to regions, for example – is invariant to changes in trade costs. In other words, changes in trade costs may alter which country or region produces the (entire) good, but there is no change in the underlying nature of production. For many research questions this assumption is appropriate. However, in the context of the home bias in trade puzzle, this paper shows that the assumption leaves out forces that are quantitatively important. The reduced-form production function does change in response to changes in trade costs; this change is the source of two magnification effects.

One effect is that if different stages are produced in different countries – vertical specialization – then, as goods cross national borders while they are in-process, they incur the trade costs multiple times. Hence, the friction limiting vertical specialization is a multiple of the standard friction limiting international trade.[3] The second effect is more subtle and is best conveyed by an example. A U.S. consumer can buy cars wholly made in the United States or cars assembled in Canada from U.S.-made components. There is a tariff on imports of cars made in Canada. Note that with the second (vertically specialized) production process, the tariff is applied to the entire car, even though only assembly occurs in Canada. The relevant or effective trade cost in this example is the tariff divided by the share of assembly value-added in the total cost. The smaller the assembly value-added share, the larger is the effective tariff. The magnified effect of the tariff exerts a magnified effect on vertically specialized production, leading to a magnified reduction in international trade.

The goal of this paper is to quantitatively assess the extent to which multi-stage production can help explain the home bias in trade puzzle. I calibrate a multi-stage production model to match gross output and value-added per worker in the United States and in each of two broad regions of Canada, Ontario-Quebec and the rest of Canada.[4] I also construct measures of trade costs within and between the United States and Canada. The costs include tariffs and non-tariff barriers, transport costs, and wholesale distribution costs; also, I compute separate costs for the auto industry and the non-auto industries.[5]. I find that the trade-weighted average of the costs

[3] Yi (2003) shows that this channel can help explain the growth of world trade, because trade cost reductions have a magnified effect operating through increased vertical specialization.

[4] In the calibration, I employ an elasticity of substitution of 5. This elasticity is on the lower end of the range of estimates, which are typically between 5 and 10. However, these estimates are potentially problematic, because they do not take into account multi-stage production, which could lead to an upward bias. I show this in the next section.

[5] The auto industry accounts for only about five percent of total merchandise (agriculture, mining, and manufacturing) value-added in the the two countries, but it accounts for more than 25 percent of merchandise trade between

involved in shipping a good between Canada and the United State in 1990 was 14.8 percent. I then solve the model with these trade costs and examine its implications for trade flows within and between the United States and Canada.

The model's implications for trade can be summarized by estimating a gravity regression of the model implied trade flows on each region's income, the distance between the two regions, and a dummy variable for whether the two regions are in the same country. I also run the regression with the actual trade flows, and then compare the coefficients on the border dummy variable. The estimated border dummy coefficient from the regression with the model implied trade is more than 2/3 of the coefficient from the regression with the actual trade. The border effect is the exponential of that coefficient; the model can explain almost 3/8 of the border effect. On both dimensions, the border coefficient and the border effect, the model explains a good deal of intra-national and international trade flows from international trade costs of just 14.8 percent. To assess the "value added" of multi-stage production, I calibrate and solve a one-stage version of the model that is a special case of the Eaton and Kortum (2002) model. The one-stage model can explain less than 1/3 of the border coefficient and only 1/9 of the border effect.

A test of the model is its implications for vertical specialization trade. The model almost exactly matches the rest-of-Canada's vertical specialization trade, and captures more than two-thirds of Ontario-Quebec's. However, the model explains only about one-third of the vertical specialization in Ontario-Quebec's auto industry, and only about one-half of the auto share of Ontario-Quebec's overall vertical specialization. Hence, the implications for overall vertical specialization accord well with the data, but the model does not fully capture the important role of autos.

I also use the multi-stage model to address the following counterfactual: what level of international trade costs would be needed for the model-implied regression to generate the border dummy coefficient in the data? They would need to be 26.1 percent. With observed trade costs at 14.8 percent, this result implies that the unknown or unobserved trade costs are smaller than the observed.

Overall, my results suggest an interpretation in which multi-stage production, in conjunction with observable trade cost measures, can explain much of the pattern of intra-national and international trade. This interpretation also accounts for much of the vertical specialization trade. In addition, when observed trade costs are 14.8 percent and total trade costs are 26.1 percent, it means

the two countries. This owes in large part to the U.S.-Canada Auto Pact, which took effect in 1965. The pact was the template for the U.S.-Canada Free Trade Agreement in 1989. I find that the cost of shipping auto goods between Canada and the United States was about half the cost of non-auto goods.

that existing policy-related barriers, such as tariff rates, and technology-related barriers, such as transportation costs, are still an important component of overall trade costs. Tariff reductions and improvements in transportation infrastructure still matter.

The most closely related papers are Hillberry and Hummels (2008) and Rossi-Hansberg (2005). The former documents the fact that different stages of production tend to be located close to one another. The latter presents a model with intermediate and final goods, an agglomeration externality, and endogenous firm location. A change in tariffs leads to changes in the location of production, which affects productivity in a way that magnifies the effects of the tariff change. Neither paper conducts a quantitative analysis of the importance of multi-stage production in explaining patterns of intra-national and international trade.

Section 2 presents the model. Also, for a special case of the model, I derive an analytical expression for the border effect. The border effect is a power of the border effect from a one-stage model, where the power is increasing in the share of first stage goods used in second stage production. The special case facilitates the intuition for how multi-stage production magnifies the effects of trade costs. Section 3 presents the calibration and solution method, and the results are in Section 4. Section 5 concludes.

2 The Model

In this section, I lay out the model and describe the intuition for how multi-stage production can magnify the effects of international trade barriers. The model is a Ricardian model of trade in which trade and specialization patterns are determined by relative technology differences across countries. The model is essentially a multi-stage, multi-region version of Eaton and Kortum (2002), a model that has been successful in fitting international trade data. It also draws from Yi (2003). Both papers extend and/or generalize the celebrated Dornbusch, Fischer, and Samuelson (1977; hereafter, DFS) continuum-of-goods Ricardian model.[6]

The basic geographic unit is a region. Countries consist of more than one region. Countries have "border" barriers, but regions do not. Each region possesses technologies for producing goods along a $[0, 1]$ continuum. Each good is produced in two stages. Both stages are tradable. Consequently, there are I^2 possible production patterns, where I is the total number of regions, for each good on the continuum. Goods are produced from labor and intermediates. The model determines which

[6]In all three papers, changes in trade occur along the extensive margin. Hillberry and Hummels (2008) provide detailed micro-evidence supporting this feature.

production pattern or patterns occur in equilibrium.

2.1 Technologies and Firms

Stage 1 goods are produced from labor and intermediates:

$$y_1^i(z) = A_1^i(z) l_1^i(z)^{1-\theta_1} M^i(z)^{\theta_1} \qquad z \in [0,1] \tag{1}$$

where $A_1^i(z)$ is region i's total factor productivity associated with stage 1 good z, and $l_1^i(z)$ and $M^i(z)$ are region i's inputs of labor and aggregate intermediate M^i used to produce $y_1^i(z)$. The share of intermediates in production is θ_1.[7] This first stage is a Cobb-Douglas version of the production function in Eaton and Kortum (2002).

Stage 2 goods are also produced from labor and intermediates; however, the intermediates are the stage 1 goods:

$$y_2^i(z) = \left(A_2^i(z) l_2^i(z)\right)^{1-\theta_2} x_1^i(z)^{\theta_2} \qquad z \in [0,1] \tag{2}$$

where $x_1^i(z)$ is region i's use of $y_1(z)$ for stage 2 production, $A_2^i(z)$ is region i's total factor productivity associated with stage 2 good z, and $l_2^i(z)$ is region i's labor used in producing $y_2^i(z)$. The share of intermediates for this stage is θ_2.

Stage 2 goods are used for final consumption or to produce the aggregate intermediate, M^i.

$$M^i = \exp\left[\int_0^1 \ln(m^i(z))dz\right]$$

where $m(z)$ is the amount of the stage 2 good used to produce M.

When either stage 1 or stage 2 goods cross regional or national borders, they incur transport costs, as well as wholesale distribution costs. These are the costs associated with distance or geography. I model both costs as iceberg costs. Specifically, if 1 unit of either stage of good z is shipped from region i to region j, then $1/(1 + d^{ij}(z)) < 1$ units arrive in region j. The gross ad valorem tariff equivalent of these costs is $1 + d^{ij}(z)$. There is an additional iceberg cost, the national border cost $1+b^{ij}(z)$. This barrier is a stand-in for tariff rates, non-tariff barriers (NTBs), and other border costs associated with regulations, time, and national culture that are relevant for

[7] The first stage of production in this model differs from that in Yi (2003) in its inclusion of intermediates. This facilitates matching both gross output, trade, and value-added (GDP) in the calibration.

international trade.[8] Consequently, I assume the gross border cost exceeds one only when regions i and j are located in different countries. Total trade costs, $1+\tau^{ij}(z)$, are given by the product of the distance costs and the border costs: $1+\tau^{ij}(z) = (1+d^{ij}(z))(1+b^{ij}(z))$

In terms of the number of countries and goods, the most general Ricardian framework is that developed by Eaton and Kortum (2002, hereafter, EK). I adopt a key part of the framework, which is the use of the Frechét distribution as the probability distribution of total factor productivities:

$$F(A_k) = e^{-TA_k^{-n}} \quad k = 1,2 \tag{3}$$

The mean of A is increasing in T. n is a smoothness parameter that governs the heterogeneity of the draws from the productivity distribution. The larger n is, the lower the heterogeneity or variance of A. EK show that n plays the same role in their model as $\sigma - 1$, where σ is the elasticity of substitution between goods, in the monopolistic competition or Armington aggregator-based trade models.[9]

Firms maximize profits taking prices as given. Specifically, in each period, they hire labor and purchase inputs in order to produce their output, which they sell at market prices.

Stage 1 firms in region i maximize:

$$p_1^i(z)y_1^i(z) - w^i l_1^i(z) - P^i M^i(z) \tag{4}$$

where $p_1^i(z)$ is the factory gate price of $y_1^i(z)$, w^i is the wage rate in region i, and P^i is the price of the aggregate intermediate.

Stage 2 firms in region i maximize:

$$p_2^i(z)y_2^i(z) - w^i l_2^i(z) - (1+\tau_1^{ji}(z))p_1^j(z)x_1^i(z) \tag{5}$$

assuming the cheapest source for intermediates used to produce region i's stage 2 good z is region

[8] To the extent that the barrier includes tariffs, I assume that tariff revenue is "thrown in the ocean."

[9] The Frechét distribution facilitates a straightforward solution of the EK model in a many-country world with non-zero border barriers. Unfortunately, such a straightforward solution does not carry over in my multi-stage framework. This is because my framework requires two draws from the Frechét distribution. Neither the sum nor the product of Frechét distributions has a Frechét distribution. I thank Sam Kortum for pointing this out to me.

The EK model has an input-output production structure, which implies vertical specialization, and leads generally to more trade flows than in a model without this structure. However, this structure is invariant to changes in trade barriers, which plays a role in the result that the elasticity of trade flows with respect to trade barriers is essentially the same as in the standard trade model. This invariance in production structure to changes in trade barriers is also true for the nested CES frameworks that are commonly used in the computable general equilibrium literature.

j. $p_2^i(z)$ is the factory gate price of $y_2^i(z)$, and $1+\tau_1^{ji}(z)$ is the total trade cost incurred in shipping the stage 1 good z from region j to region i.

M firms in region i maximize:

$$P^i M^i - \int_0^1 p^i(z) m^i(z) dz \tag{6}$$

where $p^i(z)$ is the price, inclusive of trade costs, of $m^i(z)$. For example, if $m^i(z)$ is produced in region j, $p^i(z) = p_2^j(z)(1+d^{ji}(z))(1+b^{ji}(z)) = p_2^j(z)(1+\tau_2^{ji}(z))$, where $1+\tau_2^{ji}(z)$ is the total trade cost incurred in shipping the stage 2 good z from region j to region i.

2.2 Households

The representative household in region i maximizes:

$$\exp\left[\int_0^1 \ln(c^i(z))dz\right] \tag{7}$$

subject to the budget constraint:

$$\int_0^1 p^i(z)c^i(z)dz = w^i L^i \tag{8}$$

where $c^i(z)$ is consumption of good z, and $p^i(z)$ is the price, inclusive of transport and border costs, that the household pays for good z. For example, if good z is produced in region j, $p^i(z) = p_2^j(z)(1+\tau_2^{ji}(z))$. Note that the price of the aggregate consumption bundle is P^i.

2.3 Equilibrium

All factor and goods markets are characterized by perfect competition. The following market clearing conditions hold for each region:[10]

$$L^i = \int_0^1 l_1^i(z)dz + \int_0^1 l_2^i(z)dz \tag{9}$$

[10] Of course, $l_1^i(z) = 0$ whenever $y_1^i(z) = 0$, and similarly for $l_2^i(z)$.

The stage 1 goods market equilibrium condition for each z is:

$$y_1(z) \equiv \sum_{i=1}^{I} y_1^i(z) = \sum_{i=1}^{I} (1+\tau_1^i(z)) x_1^i(z) \tag{10}$$

where $1+\tau_1^i(z)$ is the total trade cost incurred by shipping the stage 1 good from its cheapest production location to region i. A similar set of conditions applies to each stage 2 good z:

$$y_2(z) \equiv \sum_{i=1}^{I} y_2^i(z) = \sum_{i=1}^{I} (1+\tau_2^i(z))(c^i(z)+m^i(z)) \tag{11}$$

Finally, each region's aggregate intermediate must be completely used:

$$M^i = \int_0^1 M^i(z)dz \tag{12}$$

If these conditions hold, then each region's exports equals its imports, i.e., balanced trade holds. I now define the equilibrium of this model:

Definition 1 *An equilibrium is a sequence of goods and factor prices,* $\{p_1^i(z), p_2^i(z), w^i\}$*, and quantities* $\{l_1^i(z), l_2^i(z), M^i(z), y_1^i(z), y_2^i(z), x_1^i(z), c^i(z), m^i(z)\}$, $z \in [0,1]$, $i = 1, ...\mathbf{I}$*, such that the first order conditions to the households' maximization problem 7, the first order conditions to the firms' maximization problems 4, 5, and 6, as well as the market clearing conditions 9,10, 11, and 12 are satisfied.*

2.4 Vertical Specialization and Border Effects

In addition to delivering implications for intra-national and international trade, the model will deliver implications for vertical specialization. In this section, I define vertical specialization and present some measures of it for Canada, as well as for two sets of provinces within Canada. I also derive for two special cases of the model the link between trade costs and border effects. These special cases will highlight the magnification effects from multi-stage production.

Under frictionless trade, there will be complete specialization. Each stage of each good will be produced by only one region. Intra-national and international trade will occur so that agents will be able to consume all goods. Under positive trade costs, complete specialization may no longer occur. A stage of a good may be produced in more than one region. If the trade costs are high enough, autarky will occur, and each region will produce every stage of every good.

I now define vertical specialization. In previous research, D. Hummels, J. Ishii, D. Rapoport, and I have documented the increasing importance of vertical specialization in OECD and other countries.[11] In order to accommodate regions as the basic geographic unit, I modify the definition from Hummels, Ishii, and Yi (2001):

1. Goods are produced in multiple, sequential stages.

2. Two or more *regions* provide value-added in the good's production sequence.

3. At least one *region* must use imported inputs in its stage of the production process, and some of the resulting output must be exported.

In this context, imports and exports refer to shipments from one region to another; in particular, these flows can occur within a country. Figure 1 illustrates an example of vertical specialization involving three regions. Region 1 produces intermediate goods and exports them to region 2. Region 2 combines the imported intermediates with other inputs and value-added to produce a final good or another intermediate good in the production chain. Finally, region 2 exports some of its output to region 3. If either the imported intermediates or exports are absent, then there is no vertical specialization.

Clearly, this definition of vertical specialization is broader than and encompasses the HIY definition. I call the HIY definition of vertical specialization, international vertical specialization, i.e., it is the vertical specialization in which all the relevant trade flows are between countries.

A necessary condition for vertically specialized production of a good to occur is for one region to be relatively more productive in the first stage of production and another region to be relatively more productive in the second stage. Under frictionless trade, if relative wages are "between" these relative productivities, then this necessary condition is also sufficient.

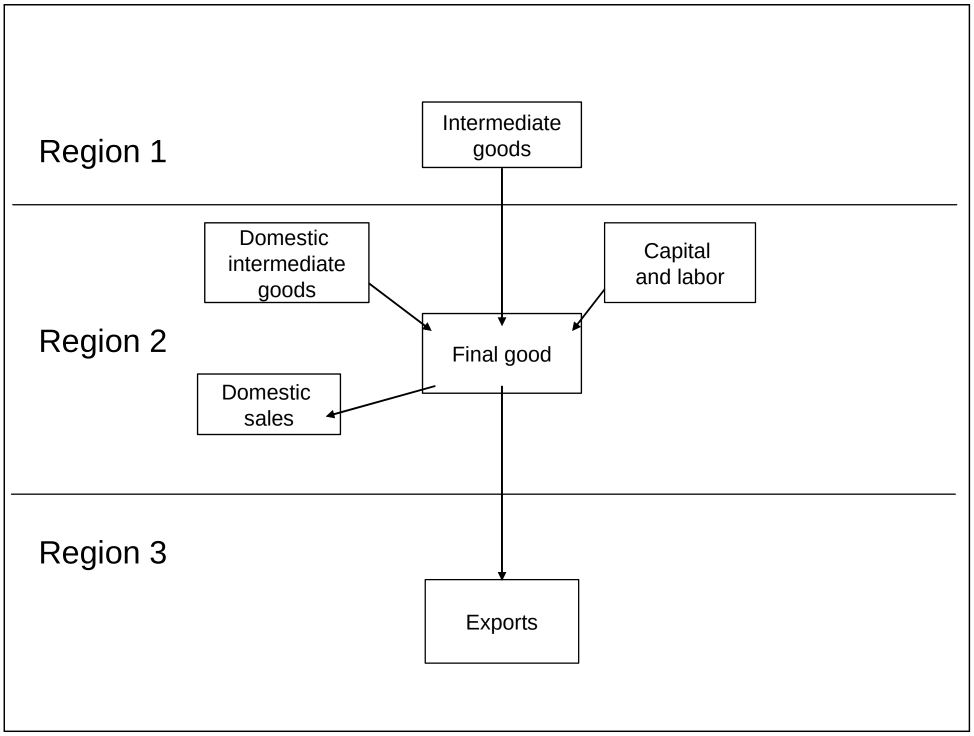
Hummels, Ishii, and Yi (HIY) develop two measures of international vertical specialization. Their primary measure is VS:

$$VS_{ki} = \left(\frac{imported\ intermediates_{ki}}{Gross\ output_{ki}} \right) Exports_{ki} \tag{13}$$

where k and i denote country and good, respectively. A regional version of VS would re-interpret k as a region.

[11] See Hummels, Rapoport, and Yi (1998), Hummels, Ishii, and Yi (2001), and Yi (2003).

Figure 1: Vertical Specialization



As discussed in HIY, ideally, VS_{ki} would be calculated at the level of individual goods, and then aggregated up. These data do not exist, at either the country or regional level. HIY relied on national input-output tables, which provide industry-level data on imported intermediates, gross output, and exports.[12] HIY find that VS_{Canada}, expressed as a share of total exports, was 0.27 in 1990. That is, for every $1 worth of exports by Canada in that year, the value of imported inputs embodied in the exports was $0.27. For motor vehicles, according to HIY, the VS share is higher, 0.51.[13]

Using the HIY industry-level VS numbers, as well as provincial-level export data by industry, I

[12] An additional advantage of using input-output tables is that they facilitate measuring the indirect import content of exports. Inputs may be imported, for example, and used to produce an intermediate good that is itself not exported, but rather, used as an input to produce a good that is. See Hummels, Ishii, and Yi (1999, 2001).

[13] How much confidence do we have in these input-output calculations? For autos, an alternative measure of $VS_{Canada,autos}$ can be calculated. I use data from Hummels, Rapoport, and Yi (2000). Owing to data limitations, $VS_{Canada,autos}$ involves imports from the U.S. only. The U.S. exported about $15.3 billion of parts *for further manufacture* to Canadian affiliates of U.S. auto manufacturers in 1989. (1989 was a benchmark year for the BEA's Direct Investment Abroad data. All dollar numbers are U.S. dollars) In addition, 80% of Canada's auto production was exported to the U.S. in that year. (Ward's Automotive Yearbook). Hence, $12.2 billion of Canada's motor vehicle exports were embodied inputs from the U.S. Total Canadian motor vehicle exports to the U.S. in 1989 was $26 billion. Consequently, according to this calculation, 47 percent of the value of Canada's exports of motor vehicles to the U.S. consisted of inputs from the U.S. This number is very close to the 51 percent value from the input-output table calculation in HIY.

impute international VS in 1990 for the sub-national regions in my calibrated model. These values, expressed as a share of merchandise GDP, are presented in Table 1 below.

Table 1: International VS (as share of region GDP)

Sector \ Region	Canada	Ontario-Quebec	Rest-of-Canada
All	0.233	0.307	0.110
Autos	0.111	0.173	0.0073
Non-autos	0.122	0.134	0.103

Sources: HIY (1999, 2001), Statistics Canada, OECD, author's calculations

Notice that autos account for almost half of Canada's VS. Also, almost all of auto's VS is from Ontario-Quebec. Almost all of Rest-of-Canada's VS is in non-autos. I will test the model by comparing these VS data against their model counterparts.

2.4.1 Border Effect in the Standard Model

To demonstrate how multi-stage production can magnify trade costs into relatively large border effects, I first develop an analytical relation linking trade costs to border effects for a special case of the model with two countries, two regions per country, and only one stage of production. This is just the EK model extended to include two regions per country. Also, there are no trade costs other than border costs, i.e., in this example trade costs and border costs are identical. To facilitate the discussion, I consider a symmetric case. I assume that the regions within a country have the same labor endowment and their (total factor) productivities are drawn from the same distribution. This implies that wages and GDPs are equalized across regions within a country.

A country's productivity for a good is defined as the maximum productivity (of that good) across the two regions: $A^h(z) = \max[A^{h1}(z), A^{h2}(z)]$, where h denotes the home country. As in DFS, without loss of generality, the goods can then be arranged in descending order of the ratio of home productivity to foreign productivity, so that $A^r(z) = \frac{A^h(z)}{A^f(z)}$ is declining in z. International imports by the home country (which equals exports) are given by:

$$\frac{(1-\underline{z}^h)w^h L^h}{1-\theta_1} \tag{14}$$

where $\underline{z}^h$ is the cutoff z that separates home and foreign production for the home market. The home country produces all goods on the interval $[0, \underline{z}^h]$ and imports all goods on the interval $[\underline{z}^h, 1]$. $w^h L^h$ is home country GDP. Imports are "blown up" by a factor of $\frac{1}{1-\theta_1}$ owing to the presence of intermediates. The foreign country produces all goods on the interval $[\underline{z}^f, 1]$ and imports all goods on the interval $[0, \underline{z}^f]$; consequently, foreign international imports are $\frac{\underline{z}^f w^f L^f}{1-\theta_1}$. Intra-national imports in the home country are:

$$\frac{\underline{z}^h w^h L^h}{2(1-\theta_1)} \tag{15}$$

This follows from the symmetry assumption about each of the two regions.

AvW define the border effect as follows:

$$\frac{Intra_b/Intra_0}{Inter_b/Inter_0} \tag{16}$$

where $Intra$ refers to intra-national trade, $Inter$ refers to international trade, the subscript b refers to actual border costs, and the subscript 0 refers to zero border costs. It is a double ratio: the ratio of intra-national trade under actual border costs to intra-national trade under zero border costs divided by the corresponding ratio for international trade. The border effect can also be thought of as the ratio of intra-national trade to international trade under actual border costs relative to what that ratio would be under zero border costs. Unlike the empirical border effect estimated directly from a gravity regression, the AvW border effect is not a model-free concept, because it relies on the counterfactual exercise of solving for trade flows when border costs are zero. While the two border effects are clearly related, I call the AvW concept the theoretical border effect, to distinguish it from the border effect estimated from the regression.[14] For the home country, the double ratio equals:

$$BorderEffect_{theoretical} = \frac{Intra_b/Intra_0}{Inter_b/Inter_0} = \frac{\underline{z}_b/\underline{z}_0}{(1-\underline{z}_b)/(1-\underline{z}_0)} \tag{17}$$

where the superscript h has been suppressed for convenience. In the one-stage model, then, the denominator of the theoretical border effect is $(1-\underline{z}_b)/(1-\underline{z}_0)$ and the numerator is given by $\underline{z}_b/\underline{z}_0$. Note that the term with intermediates, $1-\theta_1$, drops out in the above equation.

At this point, I make a further symmetry assumption, which is that the regions across countries are also identical in terms of both labor endowments and productivities. This assumption implies

[14] Table 5 of AvW (2003) gives the estimate of their border effect (10.5), as well as the estimate of the empirical border effect (16.5). As AvW show, the theoretical border effect is structural, while the empirical border effect is essentially a reduced form. I use the latter in this paper merely as a way to characterize the data; I do not give the coefficients any structural interpretation.

that wages and GDPs are equalized across countries, and relative wages and GDPs (across countries) are invariant to border costs. Assuming the productivities follow a Frechét distribution, the relative productivities will have the following functional form:

$$A^r(z) \equiv \frac{A^h(z)}{A^f(z)} = \left(\frac{1-z}{z}\right)^{\frac{1}{n}} \tag{18}$$

where $A^r(z)$ can also be interpreted as the fraction of goods z where the home productivity relative to the foreign productivity is at least A.[15] As discussed above, n is analogous to an elasticity in that a larger n implies a flatter or more "elastic" $A^r(z)$. In the appendix, I show that the solution for $\underline{z}$ is given by:

$$\underline{z} = \frac{(1+b)^n}{1+(1+b)^n} \tag{19}$$

It is easy to see that under zero border costs, $\underline{z}_0^h = \underline{z}_0^f = 0.5$. The denominator of the theoretical border effect (international trade under border costs divided by international trade under free trade) is:

$$\frac{2}{1+(1+b)^n} \tag{20}$$

This is clearly decreasing in the border cost; through international trade alone, the greater the border cost, the greater the border effect. Note that the higher the elasticity n, the greater the effect of the border cost on international trade.

The numerator of the theoretical border effect (intra-national trade under border costs divided by intra-national trade under free trade) is:

$$\frac{2(1+b)^n}{1+(1+b)^n} \tag{21}$$

This is increasing in the border cost – as the cost between countries increases, intra-national trade increases. The reason for this is essentially the idea that specialization implies that goods must be traded somewhere. If they are not traded internationally, they will typically be traded intra-nationally. This is a key insight from AvW. More specifically, consider a home country consumer in one of the regions. Under border costs, the fraction of goods purchased from home producers increases. Because the two regions within the home country are symmetric, this implies that the fraction of goods purchased from the other home region's producers, that is, intra-national trade, increases.

[15] See footnote 15 in EK (2002).

Combining the numerator and denominator yields the overall theoretical border effect, which is given by:

$$(1+b)^n \tag{22}$$

This expression is quite intuitive. Note that the border effect is independent of the intermediate input share θ_1. The presence of intermediates is necessary, but not sufficient, for magnification.

2.4.2 Border Effect in the Multi-Stage Model

With the multi-stage model, deriving analytical expressions for the theoretical border effect is considerably more difficult. To provide insight into the model, I work with a special case that has the virtue of yielding an analytical expression for the border effect. I assume that the first stage of production is produced in the country that ultimately consumes the second stage good; the second stage production location is determined by the model. Thus, if a U.S. consumer seeks to purchase an automobile, the parts and components are assumed to be produced in the United States, while final assembly can occur either in the United States or Canada. This assumption ensures that there is international vertical specialization with only one Frechét distribution of productivities compared across regions and countries for each good, which means much of the analysis from the previous sub-section can be applied here.

For goods consumed by the home country, the two possible production methods at the country level are denoted by HH and HF, where HF means that the first stage of production occurs in the Home country and the second stage of production occurs in the Foreign country. Note that production method HF involves international vertical specialization: the foreign country imports inputs and exports its resulting output. Similarly, for goods consumed by the foreign country, the two possible production methods are denoted by FF and FH, where international vertical specialization occurs with FH. I continue to assume that there are four identically sized regions; moreover, each region's productivities for both stages of production are drawn from the same distribution.[16]

If the goods are arranged in descending order of the ratio of home to foreign productivity of stage 2 production, then the analysis in the previous sub-section applies. In particular, $\underline{z}^h$ denotes the cutoff that separates home and foreign production of stage 2 goods for the home market.

[16] This latter assumption implies that the production method HH has four *ex ante* equally likely production methods distinguished by region: stage 1 can be produced in either of the two home countries' regions and likewise for stage 2 production. Two of these production methods involve intra-national vertical specialization.

International imports for the home country are now given by

$$\frac{(1+\theta_2)(1-\underline{z}^h)wL;}{1-\theta_1\theta_2} \tag{23}$$

intra-national imports are given by $\frac{(1+\theta_2)z^h wL}{2(1-\theta_1\theta_2)}$. In the appendix, I show that the solution for $\underline{z}^h$ is given by:

$$\underline{z}^h = \frac{(1+b)^{n\left(\frac{1+\theta_2}{1-\theta_2}\right)}}{1+(1+b)^{n\left(\frac{1+\theta_2}{1-\theta_2}\right)}} \tag{24}$$

Then, the numerator and denominator of the theoretical border effect are given by:

$$\frac{2(1+b)^{n\left(\frac{1+\theta_2}{1-\theta_2}\right)}}{1+(1+b)^{n\left(\frac{1+\theta_2}{1-\theta_2}\right)}} \text{and} \frac{2}{1+(1+b)^{n\left(\frac{1+\theta_2}{1-\theta_2}\right)}} \tag{25}$$

Hence, the overall theoretical border effect is:

$$(1+b)^{n\left(\frac{1+\theta_2}{1-\theta_2}\right)} \tag{26}$$

This expression differs from (22) by the presence of the $\left(\frac{1+\theta_2}{1-\theta_2}\right)$ term in the exponent. The term shows clearly that multi-stage production magnifies the effects of border costs. If $\theta_2 = 0.5$, for example, the exponent on the border cost is three times larger than in a one-stage model. Two forces underlie the $\left(\frac{1+\theta_2}{1-\theta_2}\right)$ term. The first force is the multiple border crossing or back-and-forth force. With the HF production process, the first stage encounters a border cost twice; recall that the share of stage 1 goods in stage 2 production is θ_2. Consequently, the total effect of the barrier owing to this force is $1+\theta_2$. The second force is the "effective rate of protection" force, because the concept is analogous to the concept from the literature of that name. The trade-off between HH and HF hinges on the second stage of production. The key idea is that the relevant or effective border cost is the border cost divided by the share of the second stage's value-added in the total cost. This is because the second stage is the marginal production stage, but the border cost is applied to the entire good. If the second stage value-added accounts for one-third of the total cost, for example, then the effective border cost is three times the nominal border cost. This explains the $\frac{1}{1-\theta_2}$ term.

Another way to explain the $\left(\frac{1+\theta_2}{1-\theta_2}\right)$ term is via the following decomposition. In the HF production process, the first stage encounters a border cost when it is shipped to the foreign country.

The border cost is equivalent to a cost on the second stage of production of $(1+b)^{\frac{\theta_2}{1-\theta_2}}$. A border cost is encountered again when the final good is shipped back to the home country from the foreign country. The border cost is applied to the entire good. Consequently, a cost of $1+b$ imposed on the entire HF-produced good is effectively a cost of $(1+b)^{\frac{1}{1-\theta_2}}$ on the second stage of production. The total effect is the product of these two forces. If the border cost rises, the cost of producing (internationally) vertically specialized goods rises by a multiple of the cost.

2.4.3 Discussion

The above analysis focused on a special case of the model. In the general case, in which the locations of both stage 2 and stage 1 production are endogenous, the margins described by the special case, i.e., HH vs. HF, will arise only for a fraction of the goods. For the other goods, the margins will not involve the two forces described above. One way of thinking about the model, then, is that it is a combination of the special case with its magnified border effect and other cases in which the border effect is the standard one. Consequently, in the general model, the overall effect is magnified, but the extent of the magnification is smaller than what is given by (26).

In this particular example, the share of intermediates in stage 1 production is not relevant for the border effect. What matters is the share of intermediates in stage 2 production. This is because it is these intermediates that can either stay at home or be shipped abroad for final assembly and then shipped back home. The greater the share of intermediates used in stage 2 production, the greater the fraction of the final good that crosses the border multiple times, and the larger the magnification effect. More generally, both θ_1 and θ_2 matter for the magnification effect.

The key to the magnification effect is that vertical specialization is endogenous; as border costs rise, alternative, non (internationally) vertically specialized production processes become relatively more efficient. That the model delivers changes in the *nature* of production and specialization as barriers change is what gives the model "kick" relative to the frameworks employed by AvW or EK, for example.[17]

It is common in the empirical literature to estimate the relevant elasticity of substitution from a regression of trade flows on a measure of trade costs. However, in the presence of multi-stage

[17] EK has intermediate goods, and has vertical specialization; why then, does it not have the magnification effect? As stated above, the production function in EK has just one stage. However, EK's trade cost, d_{ni}, could be re-interpreted to include all the trade costs associated with the back-and-forth flows of goods from country i to country n. In other words, EK's estimates could be interpreted as effective trade costs, not the actual trade barriers that are imposed by governments or technologies. But, EK would be silent on the mapping between these measured barriers and the effective costs. My model provides that mapping.

production, what would be estimated would be the exponent on (26). It would be the product of the elasticity (n) and the magnification term $\left(\frac{1+\theta_2}{1-\theta_2}\right)$. This suggests that estimates of substitution elasticities are upwardly biased.

Increasingly, countries apply tariffs only to the value-added that occurs abroad. These arrangements tend to arise specifically to increase opportunities for vertical specialization. When tariffs are applied only to value-added, no part of the final good is taxed more than once. However, from the above discussion, it should be clear that even under value-added tariffs, the magnification effect still exists. This is because stage 2 is still the marginal stage, and even if a tariff is levied on the final good only once, that tariff is still magnified via the effective rate of protection force. To a first approximation, the exponent that multiplies n becomes $\left(\frac{1}{1-\theta_2}\right)$. The appendix gives the derivation of the border effect in this case.

How does the magnification effect change with the number of stages of production? Obviously, this depends on how the extra stages enter. Consider a symmetric production function in which there are $2N$ stages, N is an integer, with each stage contributing $1/2N$ in value-added, and with each stage's productivity drawn from the same distribution. Let us consider two examples. In the first example, suppose the home consumer can purchase goods via only two possible production methods; one method involves involves stages $1, 3, 5, ..., 2N - 1$ produced at home while stages $2, 4, 6, ..., 2N$ are produced abroad. The other method involves all $2N$ stages made at home. Here, the back and forth force is maximized and equals $N + \frac{1}{2}$. The effective rate of protection force is $\frac{1}{1/2} = 2$, because the marginal production stages' value-added account for $\frac{1}{2}$ of the value of the good. Hence, the total magnification effect is $2N + 1$, or the number of stages plus one. In the second example, suppose again there are only two possible production methods; one method involves stages $1, 2, 3, ..., 2N - 1$ produced at home, and stage $2N$ produced abroad. The other method involves all $2N$ stages made at home. Here, the back and forth force is $2 - \frac{1}{2N}$, which is less than in the first example whenever $N > 1$. On the other hand, the effective rate of protection force is $\frac{1}{1/2N} = 2N$. The total magnification effect is $4N - 1$. The total effect in the second example exceeds the total effect in the first example whenever $2N$ exceeds 2. As with the special case presented above, both examples highlight cases that deliver the maximal magnification and border effect. These examples suggest two lessons. First, the magnification effect may be increasing in the number of stages, but note that the rate of increase linear. Second, production methods that maximize the back-and-forth force may not maximize the effective rate of protection force.

Summarizing, the discussion above suggests the following interpretation of the relation between multi-stage production and the border effect. In a world with multi-stage production, trade costs lead to a larger decrease in international trade, and a larger increase in intra-national trade, than what would be implied by a standard trade model, as indicated by (25). International trade decreases by more because of the two mechanisms discussed above: 1) the back-and-forth aspect of vertical specialization implies that at least some stages of the good are affected multiple times by trade costs and 2) the barrier is applied to the entire good, but the marginal unit of production is a single stage, whose value-added is just a fraction of the cost of the entire good. Because international trade decreases by more, intra-national trade increases by more; moreover, the ensuing increase in intra-national vertical specialization also adds to intra-national flows. Overall, the presence of multi-stage production gives rise to a larger border effect from a given trade cost than in the standard model.

3 Model Calibration and Trade Costs

I now calibrate the multi-stage model presented in sections 2.1-2.3. I focus on Canada, because it has generated the most attention in the empirical border effect literature. The United States is by far Canada's largest trading partner; consequently, the calibration will involve these two countries.[18] A central fact in U.S.-Canada trade is the importance of motor vehicles (hereafter, autos).[19] In 1990, this industry accounted for 5 percent of value-added in merchandise-producing industries in the two countries; however, it accounted for over one-fourth of U.S.-Canada trade, and almost one-half of Canada's vertical specialization (see Table 1). A key feature of the calibrated model, then, will be a two-sector framework, with autos and non-autos.[20]

The parameters and variables that are calibrated include the labor endowments of each region; the weight of the auto sector in preferences; the intermediate input shares, θ_1 and θ_2; the Frechét heterogeneity parameter n, and the Frechét mean productivity parameters T. From the discussion

[18] In 1990, over 75 percent of Canada's exports went to United States. Also, AvW (2003) estimated a two-country model (U.S. and Canada) and a multi-country model (U.S., Canada, and the rest-of-the-world). The estimates and border effect implications are very similar.

[19] The template or model for the CUSFTA in 1989 was the 1965 U.S.-Canada Auto Pact that essentially established free trade in automobiles between producers. In the five years following the Auto Pact, auto trade between the two countries soared from essentially zero to about 25-30 percent of total U.S.-Canada trade, a percentage that has remained essentially constant since then. A huge component of this trade is vertical specialization. The Auto Pact specified some domestic content restrictions; however, the available evidence suggests that these restrictions were set at the prevailing rates from that period. Also, some of these restrictions were expressed as values instead of shares; they became largely non-binding constraints after a few years.

[20] Note that in the model, wages are equalized across the two sectors in each region. The data are broadly consistent with this implication. In Canada, value-added per worker in 1990 was \$56,300 in autos and \$54,400 in non-autos. In the U.S., the numbers were \$50,700 and \$57,100, respectively.

in the previous section, the key parameters determining the magnification effect are $\theta_1$and θ_2. The intermediate shares, along with n, are key in determining the overall responsiveness of trade to the trade costs. I also construct the trade cost measures for each region and sector. The trade costs include tariffs and non-tariff barriers, transport costs, and wholesale distribution costs.

The underlying calibration strategy is straightforward. The labor endowments, the auto sector preference share, and intermediate input shares are set to match their data counterparts. The Frechét heterogeneity parameter n is taken from the existing literature. The Frechét mean productivity parameters are set so that the model matches gross output and value-added in each region-sector, as well as value-added per worker in each region. The challenge for the model is whether the trade cost data, fed into the model, will imply an allocation of stage 1 and stage 2 output to the different regions, i.e., the pattern of overall and sectoral trade and vertical specialization, in a way that matches the pattern of sectoral and overall trade and vertical specialization in the data.

At this point, it is useful to elaborate on the similarities and differences between my methodology and the usual methodology, as well as the reasons for my approach. The primary similarity with the empirical border effect literature is that it is quantitative. An additional similarity with AvW (2003) and EK (2002), and related research, is that it is structural. Two key differences are that I pursue a calibration approach, rather than an estimation approach, and that I use data on trade costs to feed into the model. One advantage of using actual data on trade costs is that I do not need to estimate an ad hoc functional form that maps distance into trade costs. Specifying and estimating a distance cost function for intra-national and international trade flows has been quite vexing, as discussed in AvW (2004).[21] A second advantage of using actual data on trade costs is that I can compare the implications of the model for trade flows against actual data on trade flows. Finally, multi-stage production renders the model sufficiently complex that there are no natural estimable equations such as equation 20 in AvW or equation 30 in EK.[22] Overall, I view my approach as complementary to existing approaches.

[21] Also, see [20] and [23].

[22] In order to estimate the model, an approach such as simulated method of moments would need to be pursued. It might be possible to run a reduced form regression of border effects on the number of production stages to see if this link is present empirically. However, data on the number of production stages by industry does not exist and appears to be difficult to construct.

3.1 Model Calibration

I divide Canada into two regions, Ontario-Quebec (OQ) and the rest-of-Canada (ROC). The latter consists of all other provinces other than the Northwest Territories, Yukon, and Nunavut. This makes sense both from a geographic perspective, as well as from a sectoral production perspective. Ontario and Quebec are the manufacturing centers of Canada, while the rest of Canada specializes more heavily in agriculture, oil, and other commodities. The United States is treated as a single region. U.S. labor is the numeraire; the U.S. wage is set to 1. I focus on 1990, which is in between the two years that McCallum and AvW focus on (1988 and 1993, respectively). As with most of the existing research, I also focus only on merchandise – agriculture, mining, and manufacturing – production and flows.

McCallum (1995) and AvW conduct their empirical analysis with data from individual provinces and states; the unit of observation is considerably more disaggregated than mine. Does the border effect puzzle disappear once the data are aggregated to just two regions in Canada and one region in the United States? Theoretically, aggregation should not matter. I run the standard gravity regression linking bilateral trade to each region's GDP, the distance between the two regions, and a dummy variable for whether the two regions are in the same country.[23] Table 2 below indicates that even with just nine observations, the regression coefficient on the border dummy is close to the McCallum coefficient; the coefficients on GDP and on distance are also similar.

Table 2: Comparison of McCallum and 3-region gravity regression

$\ln(trade_{ij}) = \beta_0 + \beta_1 \ln(gdp_i) + \beta_2 \ln(gdp_j) + \beta_3 \ln(dist_{ij}) + \beta_4 DUMMY_{ij}$

	β_1	β_2	β_3	β_4	nobs	$\exp(\beta_4)$
McCallum (1995)	1.21	1.06	-1.42	3.09	683	22.0
3-region	0.93	1.04	-0.92	3.16	9	23.6

Factor endowments are specified as follows. From the OECD's STAN database, Canada's employment in merchandise-producing industries was 2.4 million in 1990, which was 12.8 percent of U.S. employment of 18.8 million.[24] From Statistics Canada's labor force survey estimates,

[23] See the data appendix for a description of the data sources.

[24] To make the U.S. data compatible with a calibration in which the world consists of the U.S. and Canada only, I adjust the U.S. export, output, and employment data. Specifically, I subtract U.S. exports to all countries but Canada from U.S. exports and from U.S. gross output. I also calculate gross output per worker in the autos sector and the non-autos sector and use the pattern of exports to adjust U.S. employment.

employment in OQ in 1990 was 67.0 percent of total Canadian employment (in these industries).[25] Consequently, OQ's labor force is set to 0.0855 of the U.S. labor force, and ROC's labor force is set to 0.0421 of the U.S. labor force. I assume labor is not mobile across regions, but it is mobile within a region. In other words, wages or value-added per worker are equalized within a region, but not necessarily across regions.

In the model, labor is the only factor of production. With constant returns to scale production, wages and value-added per worker are the same. Of course, in the data there is a distinction between these two variables. Should the model be calibrated to match labor shares or value-added shares (in gross output)? This choice matters for calibrating two parameters, the intermediate input shares and the mean of the productivity distribution. I adopt the latter calibration.[26] Relative to the former approach, this approach results in a lower intermediate input share, which will reduce the magnification effect. Because Canada's value-added per worker relative to the U.S. is somewhat higher than its average wage relative to the U.S. (0.960 vs. 0.882) this will lead to somewhat higher mean productivity parameters, T_i,which implies a greater likelihood that Canadian consumers and firms will purchase goods produced in Canada. Hereafter, references to wages will denote value-added per worker.

Turning to the intermediate shares, θ_1 and θ_2, when $\theta_1 = \theta_2 = \theta$, it can be shown that the value-added/gross output ratio is $1 - \theta$. From the STAN database, the value-added/gross output ratio in 1990 for the U.S. and Canada, taken together, was 0.218 for autos and 0.376 for non-autos. Consequently, I set $\theta_1^{autos} = \theta_2^{autos} = 0.782$, and $\theta_1^{non-autos} = \theta_2^{non-autos} = 0.624$.

Because autos have a higher intermediate input share than non-autos, the share of autos in final expenditure is greater than the share of autos in value-added. It is 7.63 percent. Because all goods in the utility function have equal weight, I implement the two sectors by simply denoting the range $[0, 0.0763]$ as auto sector goods, and the range $[0.0763, 1]$ as non-auto sector goods. This has the same effect as a nested utility function in which the lower nest consists of a $[0, 1]$ continuum of goods combined in a Cobb-Douglas aggregator for each sector, and the upper nest consists of a Cobb-Douglas function of the auto aggregate and the non-auto aggregate, with the auto aggregate having a weight of 0.0763, and the non-auto aggregate having a weight of 0.9237.

The other key trade elasticity parameter is the heterogeneity in productivity parameter, n. As stated above, this corresponds to an elasticity of substitution in monopolistic competition or

[25] See Statistics Canada Table 282-0007.

[26] EK adopt the former calibration.

Armington aggregator models of $n+1$.[27] (Hereafter, I refer to the elasticity-equivalent of the parameter.) This elasticity is assumed identical across regions and countries. EK's estimates of n range from 3.6 to 12.86. Most of AvW's (2003) results are presented for elasticities of 5, 8, and 10. Baier and Bergstrand (2001), and Head and Ries (2001), estimate substitution elasticities of 6.43 and 7.9, respectively. In the previous section, I demonstrated that under multi-stage production the responsiveness of trade to trade costs depends on both the elasticity of substitution and the "magnification effect". Consequently, existing estimates of the substitution elasticity may be upwardly biased. Hence, I set $n = 4$.

The final parameters are the 12 productivity parameters, T, distinguished by region, sector, and stage of production ($3 \times 2 \times 2 = 12$). With no loss of generality, I normalize all U.S. T's to 1. This leaves the 8 productivity parameters for Canada. There is very little data to provide guidance on calibrating the mean productivity of a region-sector's stage 1 production relative to the mean productivity of a region-sector's stage 2 production.[28] Consequently, I set each region-sector's stage 1 and stage 2 productivity parameters equal to each other, with one exception. The exception is for OQ, which accounted for 93 percent of Canada's exports of autos in 1990. OQ specializes in auto assembly. In 1990, about 2/3 of Canada's auto exports were final vehicles, while about 2/3 of Canada's auto imports were engines and parts. Hence, for OQ, $T_2^{OQ,autos}$ is a stand-alone parameter; I assume that $T_1^{OQ,autos} = T_1^{OQ,non-autos} (= T_2^{OQ,non-autos})$. This leaves 4 parameters: $T_1^{OQ,non-autos} (= T_2^{OQ,non-autos} = T_1^{OQ,autos})$; $T_2^{OQ,autos}$; $T_1^{ROC,non-autos} (= T_2^{ROC,non-autos})$; and $T_1^{ROC,autos} (= T_2^{ROC,autos})$. I set these four parameters to match the following targets: value-added per worker in OQ, value-added per worker in ROC, auto labor in OQ, and auto labor in ROC. [29] In other words, I set these four parameters so that labor market equilibrium delivers value-added per worker and labor allocations in each region-sector that match the data. This calibration ensures that the model will match the data for total gross output and total value-added in each region-sector. As stated above, the challenge for the calibrated model is whether it will deliver the trade flows and the vertical specialization flows – the pattern of specialization – in the data. Table 3

[27] See Eaton and Kortum (2002, p. 1750, fn. 20) or Anderson and van Wincoop (2004, p. 710).

[28] A natural approach to consider would be to obtain measures of output and factor inputs for each region-sector-stage and back out a Solow residual; however, this would be a very difficult exercise because there is no data set on factor inputs and outputs by stage of production. Yi (2003) divides the input-output table into stage 1 industries and stage 2 industries, and uses Balassa revealed comparative advantage (RCA) measures to back out relative productivities according to stage. However, the RCA measures are constructed from trade data, which is what I am trying to explain. (Yi (2003) focuses on explaining the growth, not the level, of trade.)

[29] I impute labor in the auto sector for each Canadian region as follows: In 1992, 78.3 percent of auto shipments were from OQ. I assume this number holds for 1990, as well. Also, I assume value-added in autos is proportional to shipments (as is the case in the model). I then use the value-added per worker data for each region to back out the number of workers.

provides a list of all parameters, including those that are specified in advance and those that are set to hit specific targets.

The calibration of the one-stage model is similar. The parameters set in advance are identical. The four productivity parameters, $T^{OQ,autos}$, $T^{OQ,non-autos}$, $T^{ROC,autos}$, and $T^{ROC,non-autos}$, are set to match the four targets listed in Table 3.

Table 3: Calibrated Parameters

Parameter	Value	Targets
OQ labor relative to U.S. ($L^{OQ}/L^{U.S.}$)	0.0855	
ROC labor relative to U.S. ($L^{ROC}/L^{U.S.}$)	0.0421	
Size of auto sector (fraction of final expenditure)	0.0763	
Intermediate input share, $\theta_1^{autos} = \theta_2^{autos}$	0.782	
Intermediate input share, $\theta_1^{non-autos} = \theta_2^{non-autos}$	0.624	
Frechét productivity heterogeneity (n)	4	
$T_2^{OQ,autos}$		$L^{OQ,autos}/L^{U.S.}$ (0.00649)
$T_1^{OQ,autos} = T_1^{OQ,non-autos} (= T_2^{OQ,non-autos})$		$L^{ROC,autos}/L^{U.S.}$ (0.00154)
$T_1^{ROC,autos} (= T_2^{ROC,autos})$		$w^{OQ}/w^{U.S.}$ (0.910)
$T_1^{ROC,non-autos} (= T_2^{ROC,non-autos})$		$w^{ROC}/w^{U.S.}$ (1.063)

3.2 Trade Costs

I now construct the data counterparts of the trade costs between region i and region j, τ^{ij}. These include tariffs and non-tariff barriers (NTBs), transportation costs, and wholesale distribution costs for autos and for non-autos within and between the three regions.[30] The computations rely heavily on data drawn from input-output tables for the U.S., Canada, Ontario and Quebec. Some of these tables, especially the provincial input-output tables, became available only recently. I allow for asymmetry in the costs.[31] For international exports, these tables and other official data sources cover the transportation and wholesale costs only to the national border. To obtain the costs after the good crosses the border, I make two assumptions. For transport costs, I assume the cost equals the cost of an export shipment in the opposite direction. For wholesale distribution costs, I assume

[30] I do not measure retail margins. I assume that retail margins are identical across goods, regardless of production location. For example, the U.S. retail margins on a Chevrolet produced in the U.S. are assumed to be identical to the retail margins on a Chevrolet produced in Canada and exported to the U.S. Under this assumption, retail margins will not affect trade flows.

[31] Waugh (2007) finds that these asymmetries are important in explaining per capita income differences across countries.

the cost equals the cost of a domestic shipment. The data appendix describes the sources and calculations in detail and also provides a decomposition of each of the three main categories of costs.

Tables 4 and 5 below presents the total costs expressed in *ad valorem* terms for autos and non-autos in 1990. Trade costs for autos are typically about half that for non-autos. Trade costs within regions are the lowest, and international trade costs are the highest, with inter-regional costs in between. For example, the cost of shipping autos from OQ to OQ is 8.96 percent, but the cost of exporting them to the U.S. is 13.8 percent; the cost of shipping them to ROC is 12.0 percent. The international costs are higher for three reasons. First, the tariff and non-tariff barriers affect only international trade. Second, the transport costs are higher for international trade. The costs are higher because it is as costly or more costly to ship the goods to the border than to ship them to the same region or another region in the same country, and because there are additional costs to ship the good from the border to the ultimate destination. Finally, the wholesale distribution costs are also higher, for similar reasons. Overall, however, the international trade costs are not large. What matters for trade flows are international trade costs *relative to* intra-national trade costs. Computing trade-weighted averages of these two costs yields an overall relative international trade cost of 7.4 percent for autos and 15.5 percent for non-autos. Auto costs are lower because the tariffs are lower, transport costs are lower – a large fraction of auto plants in Canada are in the part of Ontario between Toronto and Windsor – and wholesale distribution costs are lower. (See appendix tables.) The lower auto trade costs will clearly deliver higher trade and vertical specialization flows, but whether they deliver the large auto share of trade and the large vertical specialization flows in the data is an open question. [32]

[32] A key assumption of the model is that as goods travel back-and-forth across countries, they incur these trade costs multiple times. This is natural for tariff barriers and non-tariff barriers, as well as for the transportation costs. I assume this is also true for wholesale distribution costs. I thank Tom Holmes and Rebecca Hellerstein for conversations with me on this subject.

Table 4: Auto Trade Costs (percent)

		From		
		Ontario-Quebec	Rest-of-Canada	U.S.
To	Ontario-Quebec	8.96	11.7	20.6
	Rest-of-Canada	12.0	8.09	19.3
	U.S.	13.8	12.2	8.28

Sources: see Appendix

Table 5: Non-Auto Trade Costs (percent)

		From		
		Ontario-Quebec	Rest-of-Canada	U.S.
To	Ontario-Quebec	14.1	20.5	39.6
	Rest-of-Canada	20.2	12.8	44.2
	U.S.	29.7	35.0	17.7

Sources: see Appendix

3.3 Solution

Given the parameterization of the model in Table 3 and the trade costs data in Tables 4 and 5, the model will deliver an equilibrium set of factor prices, goods prices, production quantities, trade flows, and vertical specialization flows. As mentioned above, I solve for the productivity parameters that yields value-added per worker in each region, as well as employment in each region and sector, that match their data counterparts.

Unlike in EK, an exact solution to the model cannot be computed. Instead, I must find an approximate solution. To do so, I approximate the $[0, 1]$ continuum with 1,500,000 equally spaced intervals; each interval corresponds to one good.[33] Further details on the solution method are in the appendix.

[33] As an accuracy check, I solve the EK model using the approximate solution method I employ. I replicate the exact EK model solution up to three significant digits.

4 Results

I now assess the quantitative importance of multi-stage production in explaining intra-national and international trade patterns for the United States and Canada. I solve both the multi-stage production model (hereafter, "benchmark" model) and the one-stage version of the model. A convenient way to characterize the model-implied trade flows is to run the McCallum gravity regression. Note that I do not give a structural interpretation to the estimated coefficients; this is a key lesson from AvW. Rather, I use the regression as a reduced form way to characterize the model-implied trade flows. Table 6 presents the main results.

Table 6: Regressions with Actual or Model-Implied Trade

$\ln(x_{ij}) = \beta_0 + \beta_1 \ln(y_i) + \beta_2 \ln(y_j) + \beta_3 \ln(dist_{ij}) + \beta_4 DUMMY_{ij}$					
	β_1	β_2	β_3	β_4	$\exp(\beta_4)$
Actual data	0.93	1.04	-0.92	3.16	23.6
Benchmark multi-stage model	0.95	0.95	-0.27	2.15	8.59
Multi-stage model with one sector	0.92	0.92	-0.27	1.94	6.96
One-stage model	0.89	0.89	-0.14	0.98	2.68

The second row of the table presents the results for the benchmark model. The regression coefficient on the border dummy variable is 2.15, which is more than 2/3 the value estimated from the actual data. The empirical border effect implied by the model is 8.59; the model can explain about 3/8 of the border effect in the data. This is a sizeable fraction resulting from our estimated international trade costs of only 14.8 percent. Nevertheless, because the implied border coefficient is less than the coefficient estimated from the actual data, the model over-predicts international trade flows and under-predicts inter-regional and intra-regional flows. For example, the model predicts that OQ exports to the U.S. are about twice their actual value, and that OQ imports from itself are about half their actual value. The model's coefficient on distance is far from the data coefficient, but the model is close to matching the coefficients on output.

What does the one-stage model imply? The fourth row of the table shows that the regression coefficient on the border dummy variable is 0.98. This is less than 1/3 the value estimated from the actual data. The coefficient implies a border effect of 2.68, less than 1/3 of the border effect implied by the benchmark model, and only about 1/9 of the data. The model predicts that OQ exports to the U.S. are about three times their actual value, and that OQ imports from itself are about one-fourth their actual value. The model's coefficients on output are also close to the data coefficients, but it does worse than the benchmark model in matching the coefficient on distance. The main result from the one-stage model is that it generates a considerably worse fit to the trade data: less than half as good as the benchmark model with respect to the border dummy coefficient, and less than one-third as good with respect to the empirical border effect.

To assess the importance of heterogeneity across the two sectors, I calibrate and solve a single sector version of the model. The parameters and trade costs are production-weighted averages across both sectors. For example, the intermediate input share is 0.637. The third row of the table shows the results. The border coefficient is considerably larger than the one-stage model's coefficient, and somewhat smaller than the benchmark model's coefficient. Hence, multi-stage production is the major reason why the border coefficient is large, with heterogeneity across sectors also playing a role.[34]

A test of the benchmark model is how well it captures international vertical specialization for Canada. The model implies that Canada's international vertical specialization flows, VS, expressed as a share of total merchandise GDP, is 22.5 percent for OQ and 10.5 percent for ROC. Table 1 above shows that the data values are 30.7 percent and 11.0 percent, respectively. So the model closely matches vertical specialization in ROC, and captures two-thirds of the vertical specialization in OQ.

Does the benchmark model capture the importance of the auto industry in VS and trade? While the model generates a disproportional importance of autos for both VS and trade, it falls short of generating the sector's actual importance, as Table 7 shows.

Table 7: Auto Sector's Importance in OQ's VS and International Trade

	VS/GDP	Share of Total VS	Share of Trade
Data	0.173	0.564	0.313
Benchmark multi-stage model	0.060	0.265	0.123
Multi-stage model with one sector	0.020	0.076	0.077

Table 7 focuses on OQ because the vast majority of auto production occurs there. The second row of the table shows that the benchmark model implies that OQ's VS for autos is 6.0 percent of

[34] Hillberry (2003) also finds that heterogeneity in trade costs matters. He estimates industry level border effects in Canada, and finds that the industry average is lower than the national border effect estimated from aggregate data.

(total merchandise) GDP; this is only about one-third of the actual value. The model also implies that autos account for 26.5 percent of OQ's total VS; this is about half of the data value. Finally, the benchmark model implies that auto trade accounts for 12.3 percent of total OQ trade, but in the data the auto share of trade is about two-and-one-half times larger. Thus, the model fails to deliver the extent of auto's role in OQ's VS and trade; however, it is useful to remember that autos represent less than 8 percent of final expenditure. In the model with just one sector, (in which there is no heterogeneity between auto goods and non-autos goods other than the weight in preferences), autos have considerably smaller shares of VS/GDP, total VS and trade, as the final row of Table 7 shows.

What happens when the key elasticity parameter, n, increases? International trade flows become more sensitive to trade costs, leading to a decline in such trade and an increase in intra-national trade. This will lead to a higher implied border effect. I find that when $n = 7.51$, the multi-stage production model can match the border dummy coefficient in the data.[35] That is, with this elasticity and the trade costs I measured, the model explains the pattern of intra-national and international trade within and between the United States and Canada. However, as elasticities increase, international vertical specialization flows are affected even more adversely than standard international trade flows. When $n = 7.51$, the model generates extremely counterfactual implications for international vertical specialization flows: 4.7 percent for OQ and 0.9 percent for ROC. This captures in a nut shell why simply raising elasticities does not resolve the border effect puzzle. It may "explain" low international trade relative to intra-national trade, but it will not explain Canada's vertical specialization flows.

All of the above analysis was with the trade costs I measured. I now subject the benchmark model to a reverse engineering exercise: What level of international trade costs is needed for the model to exactly match the border coefficient in the data? I address this by finding the number that multiplies (gross) international trade costs (only) in the two sectors so that the model implies the trade flows that yield the border dummy coefficient in the data. I find the multiplicative factor is 1.099. This translates into average international trade costs of 18.0 percent for autos, 26.8 percent for non-autos, and 26.1 percent overall. Put differently, in the presence of multi-stage production, the unobserved or unknown trade costs owing to currency differences, regulatory and

[35] With that elasticity, the coefficient implied by the one-stage model is 1.82, which translates to an empirical border effect of 6.19. Consequently, with this elasticity, the multi-stage production model could "explain" all of the border effect, while the one-stage model would explain only about 1/4 of it.

cultural factors, and other barriers, add up to only $26.1 - 14.8 = 11.3$ percentage points.[36] This result helps establish bounds on how large these other costs are.

5 Conclusion

This paper offers multi-stage production as an explanation for the home bias in trade puzzle. The presence of multi-stage production – which carries with it the possibility of vertical specialization – magnifies the effects of trade costs. I showed that there are two magnification forces. The "back and forth" trade associated with vertical specialization means that at least some stages of production bear multiple trade costs. Also, because different stages can be produced in different countries, sometimes the marginal production process is a single stage or a subset of stages. Then the "effective" trade cost is the trade cost divided by the share of the marginal stages' value-added in the total cost. Hence, under multi-stage production, a given level of trade costs can explain more of the puzzle, relative to a framework with one stage of production, because it leads to a greater reduction of international trade and a greater increase in intra-national trade.

Multi-stage production and vertical specialization break the tight link between the elasticity of trade with respect to iceberg-type trade barriers and the elasticity of substitution between goods (on either the production or consumption side) that is present in the EK model, the monopolistic competition model, and the Armington aggregator models. In these models the two elasticities are virtually identical. In a special case of the model, I demonstrate that the elasticity of trade with respect to barriers involves both the elasticity of substitution (i.e., the Frechét distribution variance parameter) and the share of stage 1 inputs used in stage 2 production. This suggests that there may be an upward bias in estimates of the substitution elasticity that do not control for multi-stage production. Chaney (2007) presents a model with firm heterogeneity and fixed costs that also breaks the link between the elasticity of trade with respect to barriers and the elasticity of substitution. Ruhl (2008) is a quantitative analysis of a similar mechanism in which there is a distinction between the short run and the long run.

The main contribution of the paper is quantitative. I develop a multi-stage, multi-region Ricardian model of trade and then pursue a calibration approach to evaluating the model's ability to address the home bias puzzle. I directly measure three types of trade costs – tariff rates and NTBs,

[36] I do the same exercise for the one-stage model. I find that when average international trade costs are 41.6 percent for autos and 52.2 percent for non-autos, or 51.5 percent overall. This implies that the unobserved or unknown trade costs are 36.7 percent, or more than twice as large as the trade costs I measured.

transport costs, and wholesale distribution costs – for the U.S. and Canada in 1990. Aided by new data from Statistics Canada, my paper also measures trade costs for autos and for non-autos; the auto sector plays a large role in U.S.-Canada trade and vertical specialization flows. I find that the international trade costs are 7.4 percent for autos, 15.5 percent for non-autos, and 14.8 percent overall. I calibrate the model to match output, value-added, and value-added per worker in each region in each sector. The challenge for the model, then, is to match the allocation of output to the different regions. In particular, the challenge is for the lower trade costs in the auto sector to generate a large volume of trade, while at the same time, the higher trade costs in the non-auto sector generate very little international trade so that a large border effect occurs.

I solve the model and estimate the usual gravity regression with the model-implied trade flows. I find that the model can explain almost 3/8 of the border effect. By contrast, a calibrated one-stage model can explain only about 1/9 of the border effect. The model generates a magnified role for the auto sector, but the extent of the magnification falls short of what is in the data. An independent test of the model is its ability to explain the extent of vertical specialization. There is a good fit, overall, but again, it fails to capture the large role played the auto sector. The fact that the most of the auto firms in Canada are affiliates of U.S. multinational parents may be a reason for this.[37]

Clearly, there are additional trade costs beyond tariffs and NTBs, transportation costs, and wholesale distribution costs. Measuring these costs – lack of a common currency or language, information barriers, regulatory burdens, security concerns, cultural differences, time, etc. – is an important task for future research.[38] However, the results from my reverse engineering exercise suggest that in the presence of multi-stage production and its magnification effect, there is an upper bound on the size of these costs, on the order of 10 percent. Put differently, the benefits of a common language or a common currency may not be as large as had been previously thought. On the other hand, my findings suggests that changes in tariffs are still important, and policies encouraging investment in transportation infrastructure may be, as well.

In the model, there are just two stages of production. Many goods are made in at least three stages. Indeed, within North America, some auto parts are already part of a three-stage vertical specialization chain: parts are made in the United States, they are exported to Mexico, where they are assembled into engines, they are exported back to the United States where they are installed in autos, and some of these autos are exported to Canada. Allowing for additional stages would

[37] See Feinberg and Keane (2006) for a study of the special role of multinationals in U.S.-Canada trade.

[38] Anderson and van Wincoop (2004) is a comprehensive survey of trade costs. Most of the costs listed above are measured indirectly.

likely enhance the explanatory power of a multi-stage production framework.

In my framework, the trade costs enter in an *ad valorem* way. Recent empirical research suggests the importance of fixed or sunk costs in international trade costs.[39] It would be useful to extend the framework to include fixed costs of exporting, as in Melitz (2003), Chaney (2007), Alessandria and Choi (2008), and Ruhl (2008). If fixed costs are increasing in the number of border crossings, it seems plausible that multi-stage production can generate magnification effects along the lines obtained with *ad valorem* costs.

This paper pursued the home bias in trade puzzle from the perspective of trade flows. There is a separate and important literature that examines the puzzle from the perspective of prices. See, for example, Engel and Rogers (1996), Gorodnichenko and Tesar (2007), and Broda and Weinstein (2008). It would be useful to merge these two literatures. For a given set of trade costs, just as multi-stage production produces magnification in terms of trade flows, multi-stage production would produce magnification in terms of prices, as well. Moreover, the magnified effects on trade flows occur precisely because of their magnified effect on prices.

A Appendix

A.1 Solution for $\underline{z}^h$ in the one-stage model

For each good consumed in the home country, there are two production methods: it can be produced at home or abroad. I follow DFS by ordering the continuum of goods according to declining home country comparative advantage. There is a cutoff $\underline{z}^h$ for which goods on the interval $[0, \underline{z}^h]$ are produced by the home country, and goods on the interval $[\underline{z}^h, 1]$ are produced by the foreign country. This cutoff is determined by the arbitrage condition that the price of purchasing this good (by a home country consumer) is the same across the two methods:

$$p^H(\underline{z}^h) \equiv \frac{(w^h)^{1-\theta_1}(P^H)^{\theta_1}}{A^h(\underline{z}^h)} = (1+b)\frac{(w^f)^{1-\theta_1}(P^F)^{\theta_1}}{A^f(\underline{z}^h)} \equiv (1+b)p^F(\underline{z}^h) \tag{27}$$

Simplifying yields:

$$\omega^{1-\theta_1}\left(\frac{P^H}{P^F}\right)^{\theta_1} = \left(\frac{A^h(\underline{z}^h)}{A^f(\underline{z}^h)}\right)(1+b) \tag{28}$$

where $\omega = w^h/w^f$. Using the result from (18), as well as the symmetry assumptions, yields:

$$1 = \left(\frac{1-\underline{z}^h}{\underline{z}^h}\right)^{\frac{1}{n}}(1+b) \tag{29}$$

Solving for $\underline{z}^h$ yields (19).

[39] See Roberts and Tybout (1997), Anderson and van Wincoop (2004), and Helpman, Melitz, and Yeaple (2004), for example.

A.2 Solution for $\underline{z}^h$ in the multi-stage production model special case

For goods ultimately consumed in the home country, there are two production methods, HH and HF. As above, ordering the continuum of goods according to declining home country comparative advantage in stage 2 production, there is a cutoff $\underline{z}^h$ for which goods on the interval $[0, \underline{z}^h]$ are produced by HH, and goods on the interval $[\underline{z}^h, 1]$ are produced by HF. This cutoff is determined by the arbitrage condition that the price of purchasing this good (by a home country consumer) is the same across the two methods:

$$p_2^{HH}(\underline{z}^h) \equiv \frac{B(w^h)^{1-\theta_1\theta_2}(P^H)^{\theta_1\theta_2}}{A_1^h(\underline{z}^h)^{\theta_2}A_2^h(\underline{z}^h)^{1-\theta_2}} = (1+b)\frac{B(1+b)^{\theta_2}(w^h)^{(1-\theta_1)\theta_2}(P^H)^{\theta_1\theta_2}(w^f)^{1-\theta_2}}{A_1^h(\underline{z}^h)^{\theta_2}A_2^f(\underline{z}^h)^{1-\theta_2}} \equiv (1+b)p_2^{HF}(\underline{z}^h) \tag{30}$$

where B is a constant. Simplifying yields:

$$\omega^{1-\theta_2} = \left(\frac{A_2^h(\underline{z}^h)}{A_2^f(\underline{z}^h)}\right)^{1-\theta_2}(1+b)^{(1+\theta_2)} \tag{31}$$

which leads to:

$$1 = \left(\frac{1-\underline{z}^h}{\underline{z}^h}\right)^{\frac{1-\theta_2}{n}}(1+b)^{(1+\theta_2)} \tag{32}$$

Solving for $\underline{z}^h$ yields (24).

A.3 Border effect with multi-stage production when barriers are value-added

When border costs such as tariffs are applied only to the value-added that occurred in the exporting country, then (30) now becomes:

$$p_2^{HH}(\underline{z}^h) \equiv (1+b(1-\theta_2))p_2^{HF}(\underline{z}^h) \tag{33}$$

Solving for $\underline{z}^h$ as before, and then plugging it into the expression for the border effect yields:

$$\Big[(1+b)^{\theta_2}(1+b(1-\theta_2))\Big]^{\frac{n}{1-\theta_2}} \tag{34}$$

When $\theta_2 \geq 1/2$, (34) is clearly greater than $(1+b)^n$. When $0 < \theta_2 < 1/2$, it can be shown that $f(\theta_2) = \big[(1+b)^{\theta_2}(1+b(1-\theta_2))\big]^{\frac{1}{1-\theta_2}} - (1+b)$, for which $f(0) = 0$ and $f(0.5) > 0$, is concave. This implies that for this range of θ_2, (34) exceeds $(1+b)^n$. Hence, the magnification effect still holds.

When $b(1-\theta_2)$ is small, $1 + b(1-\theta_2)$ can be approximated by $(1+b)^{1-\theta_2}$; hence, the border effect is:

$$(1+b)^{\frac{n}{1-\theta_2}} \tag{35}$$

With this approximation, it can be seen that the $\frac{1+\theta_2}{1-\theta_2}$ term from the “gross” barrier case is replaced by $\frac{1}{1-\theta_2}$. This is intuitive, because θ_2 is the stage 1 portion that is taxed twice in the gross case.

A.4 Data

A.4.1 Trade Costs

Tariffs and NTBs Table 2 in Lester and Morehen (1988) provides data on U.S. tariffs on Canadian goods and Canadian tariffs on U.S. goods, quantitative restrictions between the two countries, and federal procurement policies for 31 goods-producing industries for 1987. The latter two barriers are non-tariff barriers (NTBs). Lester and Morehen began with the most favored nation tariff rates on individual commodities and used import and production weights to create aggregate tariff measures for each of the 31 industries. The 31 industries correspond closely to SIC 2-digit level industries. One of these industries is autos (motor vehicles). The tariff-equivalent of the quantitative restrictions is calculated "as the difference between the observed producer price in Canada and the comparable world price (adjusted for tariffs and transportation charges)" (p. 19). The tariff-equivalent of the federal procurement policies is calculated as "the tariff rate required to reduce total imports by the same amount as the change in government imports attributable to preferential procurement practices" (p. 21).

These data are converted into tariffs plus NTBs for the U.S. vis-a-vis Ontario-Quebec (OQ) and vis-a-vis the rest of Canada (ROC) in two steps. First, the tariffs for 1987 were updated to 1990 using the "staging list" for the Canada-U.S. Free Trade Agreement (CUSFTA), obtained from Keith Head. The staging list provides the timing of the phasing out of tariffs for each industry. Most phaseouts were linear and occurred over a period of 5 to 10 years. The NTBs for 1990 are assumed to equal the NTBs for 1987. Second, Statistics Canada data on trade flows between Ontario-Quebec and the U.S. and between the rest of Canada and the U.S. were used to create import-weighted aggregate tariff plus NTB measures for each region. The tariff+NTB rate for non-autos is computed as the appropriately weighted difference between the overall aggregate tariff+NTB rate and the autos tariff+NTB rate.

Transport cost and wholesale margins for intra-national trade

1. United States. Transport cost margins and wholesale margins for intra-national U.S. trade for autos and for non-autos were calculated from the 1992 U.S. input-output tables, and include both intermediate and final demand.

2. Canada. Transport cost and wholesale margin data for intra-national Canada trade in autos and in non-autos were obtained from the Input-Output division of Statistics Canada (StatCan) and from StatCan Table 386-0001 for 1992. I assume the transport cost margins and wholesale margins (in *ad valorem* terms) are the same in 1990 as in 1992. Several adjustments were made to the Canadian trade data, because the 1992 data are incomplete; I used data from the 1997 intra-national trade tables constructed under the same methodology as the 1992 tables, as well as data from a set of 1997 tables constructed under a newer methodology, to impute the missing trade values. To calculate the transport cost margins for autos, I used the overall within-Canada transport cost margin for autos, and an additional assumption that the transport cost margins within and between OQ and ROC for autos are proportional to the transport cost margins within and between OQ and ROC for all goods, with the constant of proportionality set so that average transport cost margin matches the overall within-Canada transport cost margin for autos (obtained from the Canada input-output table). I do a similar exercise for non-autos, and for the wholesale margins. Further details are available from the author on request.

Transport cost and wholesale margins for international trade Calculating transport cost and wholesale margins for international trade involves two steps: First, calculate the transport cost and wholesale margin associated with shipping a good from the factory to the border; second, calculate the transport cost and wholesale margin associated with shipping the good from the border to its ultimate destination.

1. Transport costs, first step. I use input-output tables from the BEA and StatCan, trade data and margins data from StatCan, and inland freight data from the U.S. Census Bureau

   (a) For U.S. auto exports to OQ and ROC, I assume the transport cost margin equals the transport cost margin for U.S. auto exports to the world. These data are from the 1992 U.S. input-output tables. I assume the margins in 1990 equal the margins in 1992. For U.S. non-auto exports to OQ or to ROC, I compute the value of the margin so that the import-weighted sum of the U.S. autos and non-autos export margins equals the Census Bureau estimate of the cost of inland freight for U.S. exports to Canada. The Census estimate for 1990 was 4.5%.

   (b) For OQ and ROC exports to the U.S., I use data for 1992 from StatCan Table 386-0001 and two StatCan publications, "Interprovincial and International Trade, 1992-1998" and "Interprovincial and International Trade, 1984-1996", as well as the 1997 provincial input-output tables. I assume the export margins in 1990 are the same as in 1992; I also assume the export margin to the U.S. equals the export margin to the world. I impute the export margins for autos and for non-autos from the export margin for all goods by using the export shares of autos and non-autos, and by employing the assumption that the ratio of the auto export margin to the non-auto export margin in 1992 equals the ratio in 1997.

2. Transport costs, second step. I employ a symmetry assumption.[40] I assume the costs of shipping U.S. autos from the U.S.-Canada border to OQ, for example, are equal to the costs of shipping OQ-made autos to the U.S.-Canada border, i.e., the costs of shipping goods in the reverse direction, and similarly for non-autos. Also, I assume the costs of shipping an OQ or ROC auto from the U.S.-Canada border to its ultimate destination are the same as the costs of shipping a U.S.-made auto to the U.S.-Canada border, and similarly for non-autos.

3. Wholesale margins, first step. I use input-output tables from the BEA and Statistics Canada, as well as data from the U.S. Census Bureau. I assume the margins for 1990 equal the margins for 1992.

   (a) For U.S. auto exports, I assume the wholesale margin equals the wholesale margin for U.S. auto exports to the world. These data are from the 1992 U.S. input-output tables. For U.S. non-auto exports to OQ, I calculate the OQ-import-weighted average of the industry-level U.S. export margins for all non-auto industries (also from the 1992 U.S. input-output tables), and similarly for ROC. The import weights are based on 1990 import data from StatCan.

   (b) The first step wholesale margins for OQ exports and for ROC exports are computed from the same sources and with the same methodology as the first step transport costs for OQ and ROC exports.

[40] Rousslang and To (1993) employ similar assumptions.

4. Wholesale margins, second step. I employ a symmetry assumption. The second step wholesale margins for OQ and ROC exports are assumed to be the same as within-U.S. wholesale margins; that is, once the Canadian good crosses into the U.S., the wholesale margin to distribute the good to the retailer is assumed to be the same as the wholesale margin for a U.S. good that is shipped to a retailer. The second step wholesale margins for U.S. exports are assumed to be the same as the within-OQ or within-ROC wholesale margins.[41]

Tables Trade costs for the three categories are listed below in percent. The costs are added (I assume that each cost is applied to the factory gate price) to yield the total trade costs in Tables 4 and 5 of the text.

Table A1: Tariffs and Non-Tariff Barriers: Autos (Non-Autos)

		From		
		Ontario-Quebec	Rest-of-Canada	U.S.
To	Ontario-Quebec			1.16 (6.18)
	Rest-of-Canada			1.16 (6.22)
	U.S.	0.230 (3.56)	0.230 (2.37)	

Sources: Canada Ministry of Finance; U.N.; U.S. ITC; Head and Ries (2001), Author's calculations

Table A2: Transportation Costs: Autos (Non-Autos)

		From		
		Ontario-Quebec	Rest-of-Canada	U.S.
To	Ontario-Quebec	0.937 (2.23)	3.54 (8.40)	4.49 (8.42)
	Rest-of-Canada	2.08 (5.13)	1.06 (2.49)	4.26 (14.4)
	U.S.	4.49 (8.42)	4.26 (14.4)	2.73 (4.65)

Sources: U.S. Census Bureau, Statistics Canada, U.S. BEA, Author's calculations

Table A3: Wholesale Costs: Autos (Non-Autos)

		From		
		Ontario-Quebec	Rest-of-Canada	U.S.
To	Ontario-Quebec	8.03 (11.9)	8.19 (12.1)	14.9 (25.0)
	Rest-of-Canada	9.96 (15.0)	7.03 (10.3)	13.9 (23.6)
	U.S.	9.08 (17.7)	7.72 (18.3)	5.55 (13.1)

Sources: U.S. Census Bureau, Statistics Canada, U.S. BEA, Author's calculations

A.4.2 Other Data

All data in U.S. dollars is converted to Canadian dollars using the average exchange rate for 1990, obtained from the IMF's International Financial Statistics. The data for the gravity regression

[41] I thank Rebecca Hellerstein and Tom Holmes for useful conversations on the nature of wholesale distribution. In Goldberg and Hellerstein (2007), beer manufacturers are assumed to use one or more domestic wholesale distributors to ship beer to the wholesale distributor located near the retail outlet in the foreign country. Multinationals have manufacturing sales offices that coordinate the shipments of goods from factories to other factories and to retail outlets. These offices may also hold inventory. Goods produced in OQ, for example, may go through more than one sales office on each side of the border, before they reach their final destination.

come from several sources. All trade flow data refer to merchandise flows (agriculture, mining, and manufacturing) and are for 1990. The intra-national Canada trade flow data are from Anderson and van Wincoop (2003); they are adjusted to remove flows associated with construction and utilities. The trade flows between OQ and the U.S. and between ROC and the U.S. are obtained from Statistics Canada. The intra-national U.S. trade flow data are computed following the approach of Wei (1996), Helliwell (1998), and others, as gross output in agriculture, mining, and manufacturing minus merchandise exports. Canadian GDP for 1990 is from Statistics Canada. U.S. GDP for 1990 is from the BEA. Distances between the 3 regions are computed using great circle distance between each region's centers – Toronto for OQ, Calgary for ROC, and Chicago for the U.S. Distances within regions are computed via a slight modification of Wei (1996): the within OQ distance is 0.25 times the average of the great circle distance between Toronto and Calgary and between Toronto and Chicago, and similarly for the other two regions.

A.5 Solution Method

I compute an approximate solution to the model. I approximate the $[0, 1]$ continuum with 1,500,000 equally spaced intervals; each interval corresponds to one good. I reduce the model to nine equations in the four productivity parameters T, two aggregate intermediates M^i (for OQ and ROC), and the three aggregate regional price levels P^i. I draw a stage 1 productivity and a stage 2 productivity from the Frechét distribution for each of the 1,500,000 goods and for each region. Because there are three regions and two stages of production, there are nine possible production methods for each good. For each region's consumer, I calculate the cheapest production method (i.e., the locations of stage 1 and stage 2 production) for each good. I then calculate whether the resulting pattern of production, trade, and prices is consistent with the data targets, with intermediates goods market equilibrium, and with the candidate aggregate prices. The model uses a Gauss-Newton algorithm to adjust the candidate vector until these conditions are met. The algorithm takes about 20-30 minutes in Gauss.

References

[1] Alessandria, George, and Choi, Horag. "Establishment Heterogeneity, Exporter Dynamics, and the Effects of Trade Liberalization." Federal Reserve Bank of Philadelphia WP07-17, 2007.

[2] Anderson, James E. and van Wincoop, Eric. "Gravity with Gravitas: A Solution to the Border Puzzle." American Economic Review, March 2003, 93 (1), 170-192.

[3] Anderson, James E. and van Wincoop, Eric. "Trade Costs." Journal of Economic Literature, September 2004, 42, 691-751.

[4] Anderson, Michael and Smith, Stephen. "Do National Borders Really Matter? Canada-U.S. Regional Trade Reconsidered." Review of International Economics, 1999, 7 (2), 219-227.

[5] Baier, Scott and Bergstrand, Jeffrey. "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity." Journal of International Economics, 2001, 53, 1-27.

[6] Broda, Christian and Weinstein, David. “Understanding International Price Differences Using Barcode Data.” Manuscript, University of Chicago and Columbia University, May 2008.

[7] Chaney, Thomas. “Distorted Gravity: The Intensive and Extensive Margins of International Trade.” Manuscript, University of Chicago, May 2007.

[8] Chen, Natalie. “Intra-national versus International Trade in the European Union: Why Do National Borders Matter?” Journal of International Economics, 2004, 63, 93-118.

[9] Combes, Pierre-Philippe; Lafourcade, Miren, and Thierry Mayer. “The Trade-Creating Effects of Business and Social Networks: Evidence from France.” Journal of International Economics, 2005, 66, 1-29.

[10] Deardorff, Alan. “Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in Frankel, Jeffrey A., ed., The Regionalization of the World Economy. Chicago: University of Chicago Press, 1998.

[11] Dornbusch, Rudiger; Fischer, Stanley and Paul Samuelson. “Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods.” American Economic Review, 1977, 67, 823-839.

[12] Eaton, Jonathan and Kortum, Samuel. “Technology, Geography, and Trade.” Econometrica, September 2002, 70 (5), 1741-1779.

[13] Engel, Charles and Rogers, John H. “How Wide is the Border?” American Economic Review, December 1996, 86 (5), 1112-1125.

[14] Evans, Carolyn. “The Economic Significance of National Border Effects.” American Economic Review, September 2003, 93 (4), 1291-1312.

[15] Evans, Carolyn. “Border Effects and the Availability of Domestic Products Abroad.” Canadian Journal of Economics, February 2006, 39 (1), 211-246.

[16] Feinberg, Susan and Keane, Michael. “Accounting for the Growth of MNC-based Trade using a Structural Model of U.S. MNCs.” American Economic Review, December 2006, 96 (5), 1515-1558.

[17] Goldberg, Pinelopi K. and Hellerstein, Rebecca. “A Framework for Identifying the Sources of Local-Currency Price Stability with an Empirical Application.” Manuscript, August 2007.

[18] Goldberg, Pinelopi K. and Verboven, Frank. "The Evolution of Price Dispersion in European Car Markets." Review of Economic Studies, October 2001, 811-848.

[19] Gorodnichenko, Yuriy and Tesar, Linda. "Border Effect or Country Effect? Seattle is 110 Miles from Vancouver After All." Manuscript, University of California, Berkeley and University of Michigan, November 2007.

[20] Head, Keith and Mayer, Thierry. "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade." Manuscript, CEPII, January 2002.

[21] Head, Keith and Ries, John. "Increasing Returns Versus National Production Differentiation as an Explanation for the Pattern of U.S.-Canada Trade." American Economic Review, September 2001, 91 (4), 858-876.

[22] Helliwell, John. How Much Do National Borders Matter? Washington D.C.: Brookings Institution Press, 1998.

[23] Helliwell, John and Verdier, Genevieve. "Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada." The Canadian Journal of Economics, November 2001, 34 (4), 1024-41.

[24] Hillberry, Russell H. "Aggregation Bias, Compositional Change, and the Border Effect." Canadian Journal of Economics, August 2002, 517-530.

[25] Hillberry, Russell H. and Hummels, David. "Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data." European Economic Review, 2008, 52, 527-550.

[26] Hummels, David; Ishii, Jun and Kei-Mu Yi. "The Nature and Growth of Vertical Specialization in World Trade." Federal Reserve Bank of New York Staff Reports #72, March 1999.

[27] Hummels, David; Ishii, Jun and Kei-Mu Yi. "The Nature and Growth of Vertical Specialization in World Trade." Journal of International Economics, June 2001, 54, 75-96.

[28] Hummels, David; Rapoport, Dana and Kei-Mu Yi. "Vertical Specialization and the Changing Nature of World Trade." Federal Reserve Bank of New York Economic Policy Review, June 1998, 59-79.

[29] Lester, John and Morehen, Tony. "Trade Barriers Between Canada and the United States." Department of Finance, Canada, Working Paper No. 88-3, 1988.

[30] McCallum, John. "National Borders Matter: Canada-U.S. Regional Trade Patterns." American Economic Review, June 1995, 85, 615-623.

[31] Melitz, Marc J. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." Econometrica, November 2003, 71 (6), 1695-1725.

[32] Nitsch, Volker. "National Borders and International Trade: Evidence from the European Union." Canadian Journal of Economics, November 2000, 33 (4), 1091-1105.

[33] Obstfeld, Maurice and Rogoff, Kenneth. "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" NBER Macroeconomics Annual 2000, 2001, 339-389.

[34] Roberts, Mark and Tybout, James. "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." American Economic Review, September 1997, 87 (4), 545-564.

[35] Rossi-Hansberg, Esteban. "A Spatial Theory of Trade." American Economic Review, December 2005, 95 (5), 1464-1491.

[36] Rousslang, Donald J. and To, Theodore. "Domestic Trade and Transportation Costs as Barriers to International Trade." Canadian Journal of Economics, February 1993, 26 (1), 208-221.

[37] Ruhl, Kim. "The International Elasticity Puzzle." manuscript, University of Texas, February 2008.

[38] Waugh, Michael. "International Trade and Income Differences." manuscript, University of Iowa, November 2007.

[39] Wei, Shang-Jin. "Intra-national versus Inter-national Trade: How Stubborn Are Nations in Global Integration?" NBER Working Paper 5531, 1996.

[40] Yi, Kei-Mu. "Can Vertical Specialization Explain the Growth of World Trade?" Journal of Political Economy, February 2003, 111, 52-102.