NBER WORKING PAPER SERIES

AIRPLANES AND COMPARATIVE ADVANTAGE

James Harrigan

Working Paper 11688
http://www.nber.org/papers/w11688

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2005

This paper has benefited from audience comments at Ljubljana, Illinois, Michigan, Columbia, the World Bank, NBER, CEPR, Hitotsubashi, Tokyo, City University of Hong Kong, and Hong Kong University of Science and Technology. I thank Christina Marsh and Geoffrey Barrows for excellent research assistance. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Airplanes and Comparative Advantage
James Harrigan
NBER Working Paper No. 11688
October 2005, Revised January 2010
JEL No. F1

ABSTRACT

Airplanes are a fast but expensive means of shipping goods, a fact which has implications for comparative advantage. The paper develops a Ricardian model with a continuum of goods which vary by weight and hence transport cost. Comparative advantage depends on relative air and surface transport costs across countries and goods, as well as stochastic productivity. A key testable implication is that the U.S. should import heavier goods from nearby countries, and lighter goods from faraway counties. This implications is tested using detailed data on U.S. imports from 1990 to 2003. Looking across goods the U.S. imports, nearby exporters have lower market share in goods that the rest of the world ships by air. Looking across exporters for individual goods, distance from the US is associated with much higher import unit values. These effects are large, which establishes that the model identifies an important influence on specialization and trade.

James Harrigan
Department of Economics
University of Virginia
P.O. Box 400182
Charlottesville, VA 22904-4182
and NBER
harrigan@nber.org

1 Introduction

Countries vary in their distances from each other, and traded goods have differing physical characteristics. As a consequence, the cost of shipping goods varies dramatically by type of good and how far they are shipped. A moments reflection suggests that these facts are probably important for understanding international trade, yet they have been widely ignored by trade economists. In this paper I focus on one aspect of this set of facts, which is that airplanes are a fast but expensive means of shipping goods.

The fact that airplanes are fast and expensive means that they will be used for shipping only when timely delivery is valuable enough to outweigh the premium that must be paid for air shipment. They will also be used disproportionately for goods that are produced far from where they are sold, since the speed advantage of airplanes over surface transport is increasing in distance. In this paper I show how these considerations can be incorporated into the influential Eaton and Kortum (2002) model of comparative advantage. In this general model, differences across goods in transport costs (both air and surface) and the value that consumers place on timely delivery interact with relative distance to affect global trade patterns.

These insights are further developed in a simplified three country version of the model (three is the smallest number of countries required for distance to affect comparative advantage). In the three country model, two countries are near each other, while the third country is more distant. In equilibrium, the more remote country has lower wages, and specializes in lightweight goods which are air shipped.

The model of the paper delivers two empirical implications that I test using highly disaggregated data on all U.S. imports from 1990 to 2003. The first implication is that nearby trading partners (Canada and Mexico) should have lower market shares in goods that more distant trading partners ship primarily by air. The second implication is that goods imported from more distant locations will have higher unit values. Both of these implications are resoundingly confirmed, and the size of the effects is economically important. In short, I find that the relative distance and relative transport cost effects emphasized in the model are an important influence on U.S. trade. Finally, I show that air shipment is much more likely for goods that have a high value/weight ratio.

There is a small, recent literature that looks at some of the issues that I analyze in this paper. The most direct antecedents of my paper are Limao and Venables (2002) and Hummels and Skiba (2004). Limao and Venables (2002) is a theory paper that models the interaction between specialization and trade costs, illustrating how the equilibrium pattern of specialization involves a tradeoff between comparative production costs and comparative transport costs. The geographical structure has a central location that exports a numeraire good and imports two other goods from more remote locations. These more remote locations have a standard 2×2 production structure, and when endowments are the same at all locations and transport costs are the same for both goods the model reduces to one where greater distance from the center has simple effects on aggregate welfare: more distant countries are poorer because they face higher transport costs. When endowments and transport costs differ the analysis becomes more nuanced, with relative transport costs interacting with relative endowments to determine welfare and comparative advantage (for example, a relatively centrally located country that is abundant in the factor used intensively in the low trade cost good will have high trade volumes and high real GDP, while countries that are more distant, and/or that are abundant in the factor used intensively in the high transport cost good, will have lower trade volumes and real GDP). This rich theoretical framework is not evaluated empirically in the paper, nor to my knowledge has it been taken to the data in subsequent work.

In contrast to Limao and Venables (2002), the paper by Hummels and Skiba (2004) is mainly empirical. Like Limao-Venables the focus is on the implications of differences in transport costs across goods on trade patterns, but unlike Limao-Venables (and virtually all of trade theory) they challenge the convenient assumption that transport costs take the iceberg form. Hummels-Skiba show that actual transport costs are much closer to being per-unit than iceberg, and they use simple price theory to show the implications for trade: imports from more distant locations will have disproportionately higher f.o.b. prices. This implication is strongly confirmed using a large dataset on bilateral product-level trade. As the model of the paper is partial equilibrium, Hummels-Skiba do not address the equilibrium location of production.

A key theoretical motivation to my analysis below is Deardorff (2004). Deardorff works with a series of simple models to make a profound point about trade theory in a world of transport costs: "local comparative advantage" (defined as autarky prices in comparison to nearby countries rather than the world as a whole) is what matters in determining trade in a world with trade costs. I embed this insight into the Eaton and Kortum (2002) model of Ricardian comparative advantage in what follows.

A related literature is the "new economic geography", which is well-summarized in Fujita, Krugman, and Venables (1999) and Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2003). In new economic geography models, the interaction between increasing returns and transport costs are a force for agglomeration, and through this channel trade costs influence trade patterns. The mechanism in these models is quite different from the comparative advantage mechanism in Limao and Venables (2002) and Eaton and Kortum (2002).

David Hummels has written a series of important empirical papers that directly motivated this paper. Hummels (1999) shows that ocean freight rates have not fallen on average since World War 2, and have often risen for substantial periods. By contrast, the cost of air shipment has fallen dramatically. Chart 1 shows that these trends have continued since 1990, with the relative price of air shipping falling 40% between 1990 and 2004. Hummels (2001a) shows that shippers are willing to pay a large premium for faster delivery, a premium that has little to do with the interest cost of goods in transit[1]. Hummels (2001b) analyzes the geographical determinants of trade costs, and decomposes the negative effect of distance on trade into measured and unmeasured costs.

The following section illustrates some key features of U.S. imports by product, trading partner, and transport mode from 1990 to 2003. Section 3 presents the theory, which is then formally tested in section 4.

2 Airplanes and U.S. imports: a first look

The import data used in this paper are collected by the U.S. Customs Service and reported on CD-ROM. For each year from 1990 to 2003, the raw data include information on the value, quantity (usually number or kilograms), and weight (usually in kilograms) of U.S. imports from all sources. The data also include information on tariffs, transport mode and transport fees, including total transport charges broken down by air, vessel and (implicitly) other, plus the quantity of imports that come in by air, sea, and (implicitly) land.[2] The import data are reported at the 10-digit Harmonized System (HS) level, which is extremely detailed, with over 14,000 codes in 2003.

[1] By "the interest cost of goods in transit", I mean the financial cost of having goods in transit before they can be sold. This opportunity cost equals the value of the good × daily interest rate × days in transit.

[2] "other" transport modes include truck and rail, and are used exclusively on imports from Mexico and Canada.

I aggregate the 10-digit import data for analysis in various ways. For most of the descriptive charts and tables, I work with a broad aggregation scheme that updates Leamer's (1984) classification, which is reported in Table 1. Countries are aggregated by distance and by region, as described in Table 2. Distance from the United States is measured in kilometers from Chicago to the capital city of each country.[3]

Table 1 illustrates the great heterogeneity in the prevalence of air freight for U.S. imports, as well as some important changes over the sample. Many products come entirely or nearly entirely by surface transport (oil, iron and steel, road vehicles) while others come primarily by air (computers, telecommunications equipment, cameras, medicine). Scanning the list of products and their associated air shipment shares hints at the importance of value to weight and the demand for timely delivery in determining shipment mode. Charts 2 to 8 illustrate the variation in air freight across regions and goods (the regional aggregates are defined in Table 2, while the product aggregates correspond to the headings in Table 1). Chart 2 shows that about a quarter of US (non-oil) imports arrived by air in 2003, up from 20% in 1990 (for brevity, in what follows I'll call the proportion of imports that arrive by air "air share"). Excluding NAFTA, the non-oil air share was 35% in 2003. Chart 2 shows that this average conceals great regional variation, which is related to distance: essentially no imports come by air from Mexico and Canada, while Europe's air share is almost half by 2003, up from under 40% in 1990. East Asia's air share increased by about half from over the sample, from 20 to 30%. The airshare from the Caribbean and South America was about one-fifth over the sample. Chart 4 shows that air shipment is particularly high in labor-intensive manufactures, machinery, and chemicals (as Table 1 shows, the capital-intensive aggregate is mainly steel and other metals, which are very heavy). The biggest increase in air share came in chemicals, which (as Table 1 reveals) is accounted for by the increasing importance of pharmaceuticals, which had a 65% air share in 2003.

The remaining Charts 5 through 8 show the evolution of air share by major regions and product aggregates. The most notable fact about Chart 5 is the sharp increase in machinery's air share from East Asia, to levels similar to Europe's by 2003. Chart 6 shows a drop in the air share for labor intensive manufactures (the biggest component of which is apparel) from the Caribbean

[3] A convenient source for the distance data is http://www.macalester.edu/~robertson/index.html

and Western and South Asia. This has to do with the phenomenon documented by Evans and Harrigan (2005): as apparel production moved to Mexico during the 1990s, the shift was concentrated in goods where timely delivery is important. Essentially, U.S. apparel retailers who wanted timely delivery replaced air shipments from South Asia and the Caribbean for surface shipments from next-door Mexico.

Heavy capital-intensive goods have not shown any increase in air-share from any source since 1990 (Chart 7). Finally, the shift toward air shipment in chemicals from major suppliers was world-wide, with the exception of nearby Caribbean suppliers (Chart 8).

Tables 3 and 4 illustrate how weighted average transport costs vary by region and product category. Not surprisingly, the products that have the highest air share, Machinery and Chemicals, have the lowest air freight costs. In some of the categories, the average transport cost is lower for air than overall, which of course reflects selection: very low value/weight items, which cost a lot to move even by ship, don't get put on planes.

3 Airplanes and trade: theory

The data reviewed in the previous section clearly suggest the influence of distance and transport costs on the pattern of trade. In this section I develop a model than can be used to analyze the effects of transport costs on comparative advantage.

My basic framework comes from Eaton and Kortum (2002), simplified in some dimensions and made more complex in others. On the demand side, consumers value timely delivery, and this valuation can differ across goods. On the supply side, timely delivery can be assured in two ways: by surface transport from nearby suppliers, or by air transport from faraway suppliers. Since air transport is expensive, it will only be used by distant suppliers, and on goods which have both a high demand for timely delivery and a high value/weight ratio (and thus a relatively small cost premium for air shipment).

I derive two testable empirical implications from the model. The first implication is about the cross section of imported goods: nearby exporters will have a smaller market share in goods that faraway exporters send by air. The second implication concerns the distribution of unit values for a particular good: faraway exporters will sell goods which have on average have a higher unit value and thus lower transport costs as a share of value.

3.1 Demand

For many transactions, timely delivery is available for a substantial premium over regular delivery. Why would anybody pay such a premium? Possible answers to this question are analyzed in a few recent papers. Evans and Harrigan (2005) derive the demand for timely delivery by retailers, who benefit from ordering from their suppliers after fickle consumer demand is revealed. Evans and Harrigan show the empirical relevance of this channel using data on the variance of demand and the location of apparel suppliers: for goods where timely delivery is important, apparel suppliers to U.S. retailers are more likely to be located in Mexico, where timely delivery to the U.S. market is cheap, while goods where timeliness is less important are more likely to be located in more distant, lower-wage countries such as China. Harrigan and Venables (2006) focus on the importance of the demand for timeliness as a force for agglomeration. They analyze this question from a number of angles, including a model of the demand for "just in time" delivery. The logic is that more complex production processes are more vulnerable to disruption from faulty or delayed parts, with the result that the demand for timely delivery of intermediate goods is increasing in complexity of final production.

While the details of demand and supply differ across models, the message of Evans and Harrigan (2005) and Harrigan and Venables (2006) is that it is uncertainty that generates a willingness to pay a premium for timely delivery. For the purposes of the present paper I will model this result with a simple shortcut, and suppose that utility is higher for goods that are delivered quickly. Looking ahead, timely delivery can be assured in one of two ways: by proximity between final consumers and production, or by air shipment when producers are located far from consumers. The determination of the equilibrium location of producers is a central concern of the model.

There is a unit continuum of goods indexed by z, with consumption denoted by $x(z)$. Utility is Cobb Douglas in consumption, and the extra utility derived from timely or "fast" delivery is $f(z) > 1$.[4] Letting F denote the set of goods that are delivered in a timely matter (for brevity I will call these "fast goods"), utility is given by

$$\ln U = \int_{z \in F} \ln\left[f(z) x(z) \right] dz + \int_{z \notin F} \ln x(z) dz \tag{1}$$

[4] Eaton and Kortum (2002) assume CES preferences. Since the elasticity is substitution plays no role in the solution of their model or mine, I use Cobb-Douglas preferences for simplicity.

Order goods so that $F = [0, z'] \subseteq [0,1]$. Then the utility function can be informatively re-written as

$$\ln U = \int_0^{z'} \ln f(z)\,dz + \int_0^1 \ln x(z)\,dz$$

For nominal income *Y*, the resulting demand functions are

$$x(z) = \frac{Y}{p(z)}$$

That is, all goods have the same expenditure share, regardless of whether or not they are fast. Denoting the prices of fast goods with a superscript *f*, the indirect utility function is

$$\ln V(p,Y) = \int_0^{z'} \ln f(z)\,dz + \ln Y - \int_0^{z'} \ln p^f(z)\,dz - \int_{z'}^1 \ln p(z)\,dz$$

Changing the set of fast goods at the margin has the following effect on utility,

$$\frac{\partial \ln V(p,Y)}{\partial z'} = \ln f(z') - \ln p^f(z') + \ln p(z')$$

which is positive iff

$$f(z') > \frac{p^f(z')}{p(z')}$$

This inequality implies that consumers will prefer fast delivery of a good if and only if the marginal utility of timeliness exceeds the relative price of fast delivery.

3.2 Supply: shipping mode and geography

Atomistic producers are assumed to be perfectly competitive, which ensures that FOB price equals unit cost, but there is a choice of shipping mode (air or surface) and consequent CIF price paid.[5] Shipping costs are of the iceberg form, so that for one unit to arrive $t(z) \geq 1$ units must be shipped.

I now introduce distance into the model. Denote air and surface iceberg shipping costs from origin country *o* to destination country *d* respectively as $a_{od}(z)$ and $s_{od}(z)$, and assume that $a_{od}(z) > s_{od}(z) \geq 1\ \forall z$: air shipping is never cheaper than surface shipping. If producers are located near consumers, then (by assumption) they can achieve timely delivery using surface shipment. If producers are located far away from consumers, then they must decide if the extra expense of air shipment is worthwhile. The answer is yes if consumer preference for fast delivery $f(z)$ is higher than the relative cost of air shipment, $a_{od}(z)/s_{od}(z)$. Given the structure of costs and demand, the equilibrium shipping mode for producers located far from their customers is

$$t_{od}(z) = s_{od}(z) \quad if \quad 1 \le \frac{a_{od}(z)}{s_{od}(z) f(z)}$$

$$t_{od}(z) = a_{od}(z) \quad if \quad 1 > \frac{a_{od}(z)}{s_{od}(z) f(z)}$$

Now order z so that $\frac{a_{od}(z)}{s_{od}(z) f(z)}$ is monotonically weakly increasing in z, and define $\overline{z}_{od}$ as the implicit solution to[6]

$$\frac{a_{od}(\overline{z}_{od})}{s_{od}(\overline{z}_{od}) f(\overline{z}_{od})} = 1 \tag{2}$$

By the ordering of z, $\forall z < \overline{z}_{od}$ the optimal shipping mode is air and for all other goods the optimal mode is surface.

For every bilateral trade route from origin country *o* to destination country *d*, there will be a cutoff $\overline{z}_{od}$. This cutoff doesn't depend on wages or technology, only on bilateral transport costs. As is traditional in trade models, I will assume that preferences, including the demand for timeliness schedule $f(z)$, do not vary across countries, so that bilateral variation in transport costs determine which goods are shipped by surface and which by air.

3.3 Supply - competition

The supply side of the model is based on Eaton and Kortum (2002), henceforth EK. Labor is the only factor of production, and is paid a wage w. Labor productivity in good z in country o, $b_o(z)$, is a random variable drawn from a Fréchet distribution with parameters $T_o > 0$ and $\theta > 1$.

---

[5] FOB stands for "free on board", and refers to the price of the good before transport costs are added. CIF stands for "cost, insurance, and freight", and refers to the price after transport costs have been added.

[6] For ease of exposition, I make the innocuous assumption that $\overline{z}$ is unique.

As in EK, competition depends on the CIF price, but here the relevant price is timeliness-adjusted: a country may win the market in a good by virtue of timely delivery rather than the lowest nominal CIF price. For each good and each bilateral route we know the optimal shipping decision from the discussion above, so it will be easy to specify the probability that *o* will win the competition in *d*.

Let *C* denote the set of close country pairs, such that timely delivery is possible without air shipment. Define the timeliness-adjusted iceberg $\tilde{t}_{od}(z)$ as

$$\tilde{t}_{od}(z) = \min\left[s_{od}(z), \frac{a_{od}(z)}{f(z)} \right] \qquad \forall (o,d) \notin C \tag{3a}$$

$$\tilde{t}_{od}(z) = \frac{s_{od}(z)}{f(z)} \qquad \forall (o,d) \in C \tag{3b}$$

Perfect competition implies that the FOB price is unit cost, which is $\frac{w_o}{b_o(z)}$ Then the timeliness-adjusted CIF supply price $\tilde{p}_{od}(z)$ is

$$\tilde{p}_{od}(z) = \frac{w_o \tilde{t}_{od}(z)}{b_o(z)}$$

Country *o* will win the competition to sell good *z* in market *d* if it has the lowest timeliness-adjusted CIF price among all *N* countries, that is, if

$$\tilde{p}_{od}(z) = \min\left[\tilde{p}_{1d}(z), \ldots, \tilde{p}_{Nd}(z) \right]$$

As with EK, the probability that this happens is the probability that all the other prices on offer are greater than $\tilde{p}_{od}(z)$. The cdf of b_o is

$$F_o[b;z] = \Pr\left[B_o(z) \le b \right] = \exp\left(-T_o b^{-\theta}\right)$$

This reflects the assumption that all products *z* produced in country *o* have the same distribution of inverse unit labor requirements. Since $B_o \le b$ implies $\tilde{P}_{od}(z) \ge \tilde{p}$ where $\tilde{P}_{od}(z) = \frac{w_o \tilde{t}_{od}(z)}{B_o}$ and $\tilde{p}_{od}(z) = \frac{w_o \tilde{t}_{od}(z)}{b}$, it follows that

$$G_{od}[\tilde{p};z] = \Pr\left[P_{od}(z) \le \tilde{p} \right] = 1 - \Pr\left[B_o(z) \le b \right] = 1 - \exp\left(-T_o b^{-\theta}\right)$$

$$= 1 - \exp\left(-T_o\left(\frac{\tilde{p}}{w_o\tilde{t}_{od}(z)}\right)^{\theta}\right)$$

which is essentially the same as EK's equation (5), with the only difference that this CDF differs by goods *z*, both because of variation in the demand for timeliness and variation in shipping costs.

Following EK's logic, I next derive the CDF of the price distribution in country *d*, which is the distribution of the minimum of prices offered by all potential suppliers *o*. This is

$$G_d[\tilde{p}; z] = \Pr\left[P_d(z) \le \tilde{p}\right] = 1 - \prod_{o=1}^{N}\left[1 - G_{od}[\tilde{p}; z]\right] = 1 - \prod_{o=1}^{N}\exp\left(-T_o b^{-\theta}\right)$$

$$= 1 - \prod_{o=1}^{N}\exp\left(-T_o\left(\frac{\tilde{p}}{w_o\tilde{t}_{od}(z)}\right)^{\theta}\right) = 1 - \exp\left[\sum_{i=1}^{n} -T_o\left(\frac{\tilde{p}}{w_o\tilde{t}_{od}(z)}\right)^{\theta}\right] = 1 - \exp\left[\Phi_d(z)\tilde{p}^{\theta}\right]$$

where the price parameter for good *z* in country *d* is defined as

$$\Phi_d(z) = \sum_{o=1}^{N} T_o\left[w_o\tilde{t}_{od}(z)\right]^{-\theta} \tag{4}$$

Unlike in EK, this parameter varies by good, both because of the degree of timeliness preference and the origin-specific optimal transport mode. Note that since $\tilde{t}_{od}(z)$ depends only on technological and taste parameters, the price index $\Phi_d(z)$ has the same number of endogenous variable in it (namely the *N* wages) as in EK. The probability that *o* captures the market for *z* in *d* is

$$\pi_{od}(z) = \frac{T_o\left[w_o\tilde{t}_{od}(z)\right]^{-\theta}}{\Phi_d(z)} \tag{5}$$

which is similar to EK's equation (8).

3.4 Equilibrium

The final element of the model is market clearing. Begin by considering the demand by *d* for *o*'s labor. For good z, the probability that *d* buys from *o* is $\pi_{od}(z)$. The demand functions imply that the nominal expenditure share on good z, in CIF terms, is a constant given by

$$p(z)x(z) = Y$$

Thus the expected CIF expenditure by d on good z from o is the probability that o wins the competition in z, times aggregate expenditure in d:

$$Y_{od}(z) = \pi_{od}(z) Y_d$$

Integrating over all goods gives d's expenditure on goods from o as

$$Y_{od} = Y_d \int_0^1 \pi_{od}(z)\, dz = \pi_{od} Y_d, \qquad \pi_{od} \equiv \int_0^1 \pi_{od}(z)\, dz.$$

I now define national income for country o as the expenditure received by o from its sales to all markets:

$$Y_o = \sum_{d=1}^{N} \pi_{od} Y_d$$

or, substituting for Y,

$$w_o L_o = \sum_{d=1}^{N} \pi_{od} w_d L_d$$

In the EK case where timeliness is irrelevant and transport costs are the same across goods, this equation is identical, except that π_{od} is a simple function rather than an average across goods. As in EK, one wage can be taken as the numeraire, and solution of the model involves solving N-1 of these equations for the N-1 remaining nominal wages.

Solution of the model is conceptually straightforward. The solution algorithm is

1. Compute all the optimal bilateral transport modes and cutoffs, which depend only on model parameters.
2. Select a numeraire wage.
3. With the transport modes and cutoffs in hand, write out the *N-1* factor market clearing equations,

$$w_o L_o = \sum_{d=1}^{N} \pi_{od}(\mathbf{w}) w_d L_d \qquad o = 1,...,N\text{-}1$$

which solve for the *N-1* unknown wages. The remaining endogenous variables are found by substitution.

The welfare implications of the model are summarized by the aggregate price index. The ideal price index in country d associated with the utility function (1) is

$$P_d = \int_0^1 \tilde{p}_d(z)\,dz$$

where $\tilde{p}_d(z)$ is the timeliness-adjusted CIF price of good *z* which is a Fréchet distributed random variable with price parameter given by (4). To evaluate the price index I replace $\tilde{p}_d(z)$ by its expectation,

$$\int_0^\infty \tilde{p}\,dG_d(\tilde{p};z)\,d\tilde{p} = \gamma\left[\Phi_d(z)\right]^{\frac{-1}{\theta}}, \qquad \gamma \equiv \theta^{-1}\Gamma\left(\theta^{-1}\right)$$

where Γ is the Gamma function. The overall price index is then simply

$$P_d = \gamma\int_0^1 \left[\Phi_d(z)\right]^{\frac{-1}{\theta}}\,dz$$

Except for a different constant *γ*, this reduces to EK's price index when all goods *z* sold in *d* have the same price distribution.

3.5 Trade patterns in equilibrium

In this section, I show how relative distance affects comparative advantage. As always, comparative advantage involves the interaction of country characteristics with product characteristics. In my model, the relevant country characteristics are geographical location, and the product characteristics are timeliness-adjusted transport costs.

Consider any two origin countries 1 and 2. From equation (4), their relative probabilities of succeeding in selling product *z* in destination market *d* are

$$\frac{\pi_{1d}(z)}{\pi_{2d}(z)} = \frac{T_1}{T_2}\left(\frac{w_2}{w_1}\right)^{\theta}\left(\frac{\tilde{t}_{2d}(z)}{\tilde{t}_{1d}(z)}\right)^{\theta}$$

This expression emphasizes the three things that influence export success: overall productivity *T*, wages *w*, and timeliness-adjusted transport costs $\tilde{t}$. Only the latter varies by product for a particular pair of origin countries.

Suppose 1 and 2 are both close to *d* (one of them might even be *d*). Then using the expressions for $\tilde{t}$ from equations (3), the relative probabilities are

$$\frac{\pi_{1d}(z)}{\pi_{2d}(z)} = \frac{T_1}{T_2}\left(\frac{w_2}{w_1}\right)^{\theta}\left(\frac{s_{2d}(z)}{s_{1d}(z)}\right)^{\theta} \tag{6}$$

The expression is the same if both origin countries are faraway from *d* but the optimal shipping mode is surface. If the optimal shipping mode for both is air, then

$$\frac{\pi_{1d}(z)}{\pi_{2d}(z)} = \frac{T_1}{T_2}\left(\frac{w_2}{w_1}\right)^{\theta}\left(\frac{a_{2d}(z)}{a_{1d}(z)}\right)^{\theta} \tag{7}$$

An implication of equations (6) and (7) is that timeliness is irrelevant to export success across products when the optimal shipping mode is the same.

Now suppose that 1 is close to *d*, 2 is far, and the product *z* is sent by ship from 2. Then

$$\frac{\pi_{1d}(z)}{\pi_{2d}(z)} = \frac{T_1}{T_2}\left(\frac{w_2}{w_1}\right)^{\theta}\cdot\left(\frac{f(z)s_{2d}(z)}{s_{1d}(z)}\right)^{\theta} \tag{8}$$

Comparing this expression to (6) illustrates the mechanism in Evans and Harrigan (2005): when goods have a high value of timeliness, and are not shipped by air, then the market share is larger for the nearby country in these goods.

Lastly, suppose that 1 is close to *d*, 2 is far, and the product is shipped by air from 2. Then the degree of timeliness preference is irrelevant to export success, but the relative cost of air and surface shipping becomes important,

$$\frac{\pi_{1d}(z)}{\pi_{2d}(z)} = \frac{T_1}{T_2}\left(\frac{w_2}{w_1}\right)^{\theta}\cdot\left(\frac{a_{2d}(z)}{s_{1d}(z)}\right)^{\theta} \tag{9}$$

To recap the above discussion, the following table summarizes the competitive environment in a given destination market *d*, from the standpoint of various potential suppliers. The cells of the table indicate the optimal shipping mode:

	Type I goods	Type II goods
supplier near to *d*	surface	surface
supplier far from *d*	surface	air

For Type I goods, the nearby suppliers (including suppliers in *d*) have an advantage in timely delivery relative to faraway suppliers. For Type II goods, all suppliers make timely delivery, but nearby suppliers have a transport cost advantage because they don't have to pay air shipping charges. In equilibrium, the only goods that will not be delivered in a timely matter will be those goods sent by surface from faraway suppliers. These are likely to be goods where timely delivery is not highly valued (if timely delivery was very valuable, then the goods would probably be produced by the nearby supplier). Turning to Type II goods, air shipped goods are likely to be ones where the cost premium for air relative to surface shipment is not too large.

Continue with the case where country 1 is near and country 2 is far from destination *d*. Consider two goods z^L and z^H that are "light" and "heavy" respectively in the following sense:

$$\frac{a_{2d}\left(z^L\right)}{f\left(z^L\right)} \leq s_{2d}\left(z^L\right) \leq s_{2d}\left(z^H\right) \leq \frac{a_{2d}\left(z^H\right)}{f\left(z^H\right)} \tag{10}$$

These inequalities imply that the light good will be shipped from 2 to *d* by air, and the heavy good will be shipped by surface. Because 1 and *d* are close, both goods will be shipped from 1 to *d* by surface. Then dividing (9) by (8) gives

$$\frac{\pi_{1d}\left(z^L\right)/\pi_{2d}\left(z^L\right)}{\pi_{1d}\left(z^H\right)/\pi_{2d}\left(z^H\right)} = \left(\frac{a_{2d}\left(z^L\right)}{s_{1d}\left(z^L\right)} \times \frac{s_{1d}\left(z^H\right)}{f\left(z^H\right)s_{2d}\left(z^H\right)}\right)^{\theta} \tag{11}$$

To evaluate this ratio, I make two additional innocuous assumptions.

1. The degree of timeliness preference $f\left(z\right)$ is constant.
2. The two country's surface shipping cost schedules are proportional, or $s_{2d}\left(z\right) \propto s_{1d}\left(z\right)$.

Using these assumptions and substituting gives

$$\frac{\pi_{1d}\left(z^L\right)/\pi_{2d}\left(z^L\right)}{\pi_{1d}\left(z^H\right)/\pi_{2d}\left(z^H\right)} = \left(\frac{a_{2d}\left(z^L\right)}{f\left(z^L\right)s_{2d}\left(z^L\right)}\right)^{\theta} \leq 1$$

or

$$\frac{\pi_{1d}\left(z^L\right)}{\pi_{1d}\left(z^H\right)} \leq \frac{\pi_{2d}\left(z^L\right)}{\pi_{2d}\left(z^H\right)} \tag{12}$$

The inequality follows from the first inequality in (10), and establishes the following proposition:

Proposition 1

In a given market, distant suppliers have a comparative advantage in light weight goods, while nearby suppliers have a comparative advantage in heavy goods. That is, faraway countries have a relatively high market share in goods which are shipped by air.

This is the key empirically testable prediction of the model, and the intuition is straightforward. For heavy goods, distant producers have the double disadvantage of high shipping costs and slow delivery. In lightweight goods, distant producers can match the timely delivery of nearby suppliers by using air shipment, and their competitiveness in this range of goods depends on the cost of air shipment and the utility value of timely delivery.

Proposition 1 can be understood with the help of Figure 1. The figure incorporates the additional assumption that air shipping costs increase faster with weight than do surface shipping costs. An implication is that goods with higher value to weight ratios are more likely to be shipped by air. This commonplace observation will be confirmed in Table 11 below. In the figure, the vertical axis measures timeliness-adjusted transport costs and the horizontal index orders goods by increasing weight. The bold lower envelope $\tilde{t}_2(z) = \min\left[s_2(z), \frac{a_2(z)}{f(z)}\right]$ is the faraway country's minimized timeliness-adjusted transport cost schedule. For "light" goods $z \leq \overline{z}_2$, goods are shipped by air from country 2, while "heavy" $z > \overline{z}_2$ goods are shipped by surface regardless of which country sells them. Since $\frac{s_1(z)}{f(z)} = \tilde{t}_1(z) < \tilde{t}_2(z)$ for all *z*, the nearby supplier always has an absolute transport cost advantage, but this cost advantage is smaller for goods that are shipped by air from country 2. The Proposition follows immediately, since in the model comparative advantage is a monotonic function of differences in relative transport costs across goods.

3.6 An illustrative three-country model

Without putting more structure on the model, I can not say much about equilibrium wages and gains from trade, nor can I do comparative statics. In this section, I show that useful insights can be obtained by computing the full equilibrium for a special 3-country case. Three countries are the minimum required for relative transport costs to matter to the equilibrium of the model, and the example serves to illustrate how distance affects comparative advantage and real wages. The simplifying assumptions are

1. Country 1 and 2 are adjacent, country 3 is distant (it may be useful to keep in mind examples such as the U.S., Mexico and China, or the U.S., Canada, and Europe).
2. All countries share the same aggregate technology parameter T and the same labor supply *L*.
3. The degree of timeliness preference is constant across goods, $f(z) = f$.

4. Air shipment costs increase more rapidly with weight than surface transport costs.

The first three simplifications are for tractability only, but the fourth is more substantive. As noted above, the fourth assumption is empirically validated in the empirical analysis that follows. Since what matters for comparative advantage is the ratio of air to surface transport costs, I economize on notation by assuming that surface transport costs are the same for all goods on a given trade route.

Because countries 1 and 2 are identical in every way, they will have the same wage in equilibrium. I take this wage as the numeraire, and w_3 denotes the nominal wage in country 3.

The geography of the three country model means that the structure of transport costs is very simple. Countries 1 and 2 can ensure timely delivery to each others' markets without using expensive air shipment, so all trade between 1 and 2 takes place by surface transit at an iceberg cost of s_1. Some long distance trade involving country 3 will be by air at an iceberg cost of $a(z)$, and some by ship at a cost of s_3 where $a(z) > s_3 > s_1 > 1$.

Order goods so that $a(z)$ is increasing in z, and let $\overline{z}$ be the cutoff for air shipment between 3 and 1 or 2. By equation (2), $\overline{z}$ is determined by

$$\frac{a(\overline{z})}{s_3 f} = 1 \qquad (13)$$

For goods shipped to or from 3, the transport mode will be air for $z \in [0, \overline{z})$ ("light" goods) and surface for all other ("heavy") goods.

Expressions for the three price parameter functions $\Phi_d(z)$ and the nine market share functions $\pi_{od}(z)$ are found using equations (4) and (5). What is of interest here are the results for trade patterns. I focus on the results of competition in market 1, where the two possible foreign suppliers are nearby country 2 and faraway country 3.[7]

Consider 3's market share in light goods in 1 relative to 2's market share. These are goods where both sellers provide timely delivery, but using different transport modes. From (9),

$$\frac{\pi_{31}(z)}{\pi_{21}(z)} = \left(\frac{s_1}{w_3 a(z)}\right)^{\theta}, \qquad z < \overline{z}.$$

[7] Country 1 may of course supply its own market in any good z, but this has no bearing on the relative chances of 2 or 3 winning the competition.

Next, consider the same competition in heavy goods. These are goods where 2 provides timely delivery and 3 does not. From (8),

$$\frac{\pi_{31}}{\pi_{21}} = \left(\frac{s_1}{f\, w_3 s_3} \right)^{\theta}, \qquad \bar{z} \leq z.$$

Both of these relative market shares are decreasing in 3's wage and in the relative cost of transport used by 3. Now define the ratio of ratios as in (11), which gives comparative advantage:

$$\frac{\pi_{31}(z)/\pi_{21}(z)}{\pi_{31}/\pi_{21}} = \left(\frac{f\, s_3}{a(z)} \right)^{\theta} \geq 1$$

As expected from Proposition 1, country 3 has a comparative advantage in air-shipped goods in market 1, and 2 has a comparative advantage in heavy goods. The lighter the good (that is, the lower is $a(z)$), the greater is 3's comparative advantage.

3.7 Equilibrium in the 3 country model

Because of symmetry and low dimensionality, there are only two endogenous variables to be solved for in the three country model, with the remainder found by substitution. The first is the cutoff $\bar{z}$, which is trivial to solve for by equation (13). The second endogenous variable is the nominal wage in country 3, which can be found using the national income identity for any one of the countries.

To compute equilibrium requires a functional form for the air transport cost schedule $a(z)$. It is convenient to use

$$a(z) = \beta\, s_3 f^{1+z}$$

where the shift parameter β has a range of $\left[f^{-1}, 1 \right]$. Recall that the condition for airfreight to be profitable for trade with country 3 in good z is $a(z) \leq f s_3$. For low values of β air freight is always profitable for country 3 trade,

$$\beta = f^{-1} \rightarrow \qquad a(0) = s_3 \qquad a(1) = f s_3$$

while for high values it is never profitable:

$$\beta = 1 \rightarrow \qquad a(0) = f s_3 \qquad a(1) = f^2 s_3$$

Substituting into the cutoff condition (13) gives the solution for $\bar{z}$:

$$\bar{z} = -\frac{\log \beta}{\log f} \in [0,1]$$

Given the functional form for the air transport cost schedule and numerical values for the parameters of the model, computation of equilibrium is straightforward.[8] Figures 2 and 3 show how the equilibrium changes as the cost of air transport moves from low (such that country 3 conducts all its trade by air) to prohibitive (such that country 3 conducts all its trade by surface).

Figure 2 shows real and nominal wages. As air transport becomes more expensive, all countries suffer real wage declines, but country 3 suffers the most. The welfare effect of higher air transport costs on country 3 works both through lower nominal wages (the global demand for country 3 labor declines) and a higher consumption price index (greater timeliness-adjusted transport costs).

The fact that the distant country has lower nominal wages and a higher price level is a result familiar from economic geography models (see Redding and Venables (2004) for theory and supportive empirical evidence), and lower real income as a function of distance from the center is also a feature of the equilibrium in Limao and Venables (2002). An interesting difference from Limao-Venables is the result here that all countries gain from a drop in air transport costs, though the gain is disproportionately larger for the distant country. In Limao-Venables there is a terms-of-trade effect that can hurt countries that are near the center when more distant countries enter the world trading system, as the nearby countries now face greater competition in their export market. This terms-of-trade effect is absent in my model because the nature of competition is different: in response to greater competitiveness from remote country suppliers, the centrally located economies become more specialized, tending to move out of products where the remote country has become more competitive.

Figure 3 shows the corresponding effects on aggregate trade patterns. The overall openness of the two nearby countries doesn't fall by much when air transport becomes more expensive: the nearby countries trade more with each other (higher π_{12}), as well as consuming more of their won production (higher π_{11}). By contrast, the elimination of air transport as a viable option causes country 3 to reduce its total trade substantially.

Falling air transport costs expand the range of goods which are potentially shipped by air. The increase in $\bar{z}$ creates excess supply for country 2's labor, as some goods formerly produced in 2 are now profitable to produce in 3 and send by air. In the new equilibrium relative wages in 2 decline, and the resulting effects on market shares are illustrated in Figure 4. Country 2 increases its market share in all heavy goods, where 2's now-lower wage improves its competitiveness, and loses market share in light goods, where the lower cost of air shipping more than offsets the drop in 2's wages. Thus in equilibrium distance matters *more* to specialization rather than less when some transport costs fall, in the sense that market shares across goods are more strongly correlated with relative distance. A similar result is derived by Limao and Venables (2002), who find that countries closer to the center will specialize more strongly in the transport-intensive good when overall transport costs decline.

3.8 The model's predictions for trade data

For any given level of wages, the model delivers predictions about the cross-section of goods imported by a given country. It is these predictions which will be the focus of the empirical analysis, with the United States as the importing country. I will focus on Proposition 1, which is illustrated in Figure 4: nearby countries will have higher market shares in heavy goods and faraway countries will specialize in light goods.

As noted in Section 2, the import data are reported at the 10-digit Harmonized System (HS) level, which is extremely detailed, with over 14,000 codes in 2003. These 10-digit categories will be the empirical counterpart of the goods in the model in what follows.

The import data does not report prices, but since it reports both value and quantity I can calculate unit values, defined as the dollar value of imports per physical unit. Since shipping costs depend primarily on the physical characteristics of the good rather than on its value, low value goods will be "heavy" in the sense of having a higher shipping cost per unit of value[9]. For example, consider shoes. Quantities of shoes are reported in import data, and the units are "number" as in "number of shoes". Expensive leather shoes from Italy and cheap canvas sneakers from China weigh about the same, but the former will have a much higher unit value. In the context of the model, Italian leather shoes are "lighter" than Chinese fabric sneakers, in the economically relevant sense that the former have lower transport costs as a share of value. It is

[8] Calculations were done in *Mathematica*, and the programs are available on request.
[9] The relationship between shipping cost and shipment value is estimated by Hummels and Skiba (2004), Table 1. They find that shipping costs increase less than proportionately with price.

important to keep in mind that it is meaningless to compare unit values when the units are not comparable: dollars per number of shoes is not comparable to dollars per barrel of oil or dollars per square meter of fabric.

The model's predictions about specialization can be expressed in two ways. The first is in terms of relative quantities: nearby countries have a higher market share in heavy goods than lightweight goods. While it would be difficult if not impossible to classify goods by weight, the data does report which goods are shipped by air, so I can directly test the alternative statement of Proposition 1: faraway countries have a relatively high market share in goods which are shipped by air. Testing this formulation of the theory will be the first empirical exercise undertaken below.

A serious objection to the above strategy is that there are many reasons unrelated to weight why a country might have a high market share in a particular good. For example, Canada has a very high market share in lumber and wood products, which have relatively low value per kilo and are almost never shipped by air; since Canada is adjacent to the United States this would seem to support the model. But of course it would be absurd to explain trade in lumber while ignoring the fact that Canada is covered in trees.

Consequently, I conduct a second empirical exercise that focuses on the model's predictions about unit values of goods which are actually imported. For a particular group of goods, the model predicts a relationship between unit values and distance: imports from nearby countries should have lower unit values than imports from more distant countries. That is, the deviation of unit values from the group mean should be positively related to distance from the U.S.

To state this a bit more formally, suppose that a given HS code contains goods of varying weights, which we can order from lightest to heaviest. According to the model, each good within the HS code will be provided by one country, with the winner of the good-by-good competition being stochastic. Thus, a country that exports in this code must have won at least one competition. Conditional on exporting in this code, nearby countries are more likely to have won competitions among the heavier goods, and more distant countries are likely to have won in the lighter goods.

Note that this formulation of the model's prediction effectively controls for other, non-weight related determinants of specialization (the "Canadian trees" problem). This is because the prediction about the cross-section of unit values within an HS code is conditional on countries exporting in that category at all.

4 Airplanes and trade: empirical evidence

The trade data that was described in section 2 above will now be used to test the theory laid out in section 3. In addition to data on imports and distance, I also use data on macro variables such as real GDP per capita and aggregate price level, which come from the Penn World tables.[10]

4.1 Statistical results: market shares

The first empirical exercise is focused on the prediction that exporters that are far from the United States will have a relatively high market share of U.S. imports in goods which are shipped by air. The geography of North America suggests an obvious empirical definition of "near" and "far": Mexico and Canada are near the United States, while all other exporters are far. Thus, the prediction becomes

> Mexico and Canada will have lower shares of U.S. imports in products which the rest of the world ships by air.

As a preliminary to the statistical model, Plot 1 illustrates the relationship between the Mexico-Canada market and the share of non-NAFTA imports that arrive by air ("air share"), for 2003. There is a clear negative relationship: when the non-NAFTA air share is low, the NAFTA market share is on average higher than when the non-NAFTA air share is high. This is exactly what the theory predicts.

The prediction can be tested more formally using the following linear regression equation:

$$\pi_i^F = \beta_0 + \beta_1 a_i + \varepsilon_i \tag{14}$$

where

π_i^F = Faraway exporters' aggregate share of the U.S. import market of product *i*

a_i = share of product *i* imports that arrive by air from exporters other than Canada and Mexico

The prediction of the model is $\beta_1 > 0$. The results are reported in Tables 5 and 6.

The following issues in estimating equation (14) are important:

[10] The Penn World table data are available at http://pwt.econ.upenn.edu/php_site/pwt_index.php

1. <u>measurement of a_i</u>: In the data, a given good from a given exporter is almost invariably shipped entirely by air or entirely by surface. Aggregating across all faraway exporters (which is how a_i is constructed) introduces some heterogeneity, but about half of all goods have an air share of either zero or one. To account for this, I report two specifications. The first treats a_i as a continuous variable. The second creates two indicators for $a_i = 1$ and $a_i = 0$ respectively.
2. <u>omitted variables</u>: In addition to being near the U.S., Canada and Mexico also belong to NAFTA starting in 1994, which means they have a tariff preference compared to faraway countries which differs across products and which may be correlated with a_i. To control for this I include the average tariff for faraway countries (which is equivalent to the Canada-Mexico tariff preference) in all regressions, though to save space I do not report the coefficients. Unreported results show that excluding tariffs from the regressions has no effect on the parameter of interest.
3. <u>estimation sample</u>: Pooling across all products and exporting countries may obscure important variation in β_1. To account for this issue I estimate the model on various sub-samples in addition to the full sample. First, I break products down into manufacturing products (SITC 6 manufactured goods, SITC 7 machinery and transport equipment, and SITC 8 miscellaneous manufactures) and nonmanufacturing products. Second, I pool only high-income exporters (defined by the World Bank classification in each year), so that β_1 is identified by the market share difference between Canada and all other high-income exporters. Finally, I pool only middle-income exporters, so that β_1 is identified by the market share difference between Mexico and all other middle-income exporters.
4. <u>error structure</u>: Since market shares by construction are between zero and one, the OLS assumption that the error term has infinite range is not valid, and OLS fitted values may lie outside the unit interval. To control for this, Table 6 reports results from a double-sided Tobit specification which ensures that all fitted values lie in the unit interval. All covariance matrices are computed using the heteroskedastic-robust White estimator.

Table 5 and 6 report many numbers, but the key message of the Tables is told by the numbers highlighted in bold, which report estimates of the airshare effect in 2003. I focus my discussion here on the Tobit specification of Table 6, though it should be noted that the estimated effects are somewhat smaller in the (simpler but mis-specified) OLS specification of Table 5.

The top row of Table 6 shows that in 2003, for the full sample the coefficient on airshare was 0.09. The interpretation is that in moving from goods which were completely shipped by surface to those completely sent by air, the average market share of faraway exporters went up by 9 percentage points. The specification that looks just at the extremes of a_i implies an effect of 14 percentage points: goods shipped solely by air had an average 13.4 percentage point higher market share relative to Mexico and Canada than goods shipped entirely by surface. These are economically big effects. The rest of Table 6 shows that the effect is strongest for high-income exporters of manufactured products: compared to Canada, other rich exporters of air shipped manufactured products have a 22.9 percentage point higher market share in goods sent by air compared to goods sent by surface (central panel of table, second bolded column). By contrast, the effect is not statistically significant from zero for middle income exporters of non-manufactured products (last panel of table, last bolded column); the effect for high-income exporters of non-manufactured goods is 0.101, about the same as the overall effect.

In summary, Table 6 shows that the interaction of distance and transport mode has an important influence on the source of U.S. imports, at least for high-income exporters (The effect is weak or non-existent for middle income exporters). The effect is strongest in manufactured goods shipped by high-income exporters, which were more than 40 percent of U.S. imports in 2003. This is striking evidence in support of Proposition 1: in both the model and the data, faraway countries have a relatively high market share in goods which are shipped by air.

4.2 Statistical results: unit values

In this section I focus on what the model predicts about the price of imports across source countries: imports from faraway countries will have higher f.o.b. unit values than goods shipped from nearby countries. Statistically, I investigate this by looking at variation in unit values across exporters within 10-digit HS categories. The econometric model I use is

$$v_{ic} = \alpha_i + \beta d_c + other\ controls + \varepsilon_{ic} \tag{15}$$

where

v_{ic} = log unit value of imports of product i from country c

α_i = fixed effect for 10-digit HS code i

d_c = distance of c from United States

Note that import values are measured f.o.b, so they do not include transport charges. The model predicts $\beta > 0$ in equation (15): across exporters within a 10-digit commodity category, more distant exporters will sell products with higher unit values, controlling for other observable country-specific factors which might affect unit values. When the units are kilograms, then the prediction for unit values is a prediction about the value-weight ratio.

The fact that equation (15) uses only cross-exporter variation within each 10-digit HS differentiates it from equation (14). The advantage of using a within-product estimator is that it controls for which goods a country exports: if a country does not export product *i* to the US, then that country's distance from the US is (appropriately) irrelevant to the effect of distance on unit values within product *i*. Product fixed effects also control for differences in physical characteristics of products, making it possible to meaningfully pool information from microchips and potato chips.

The basic measurement of distance is distance in kilometers between the US and the exporting countries. The model, and common sense, give no reason to expect that any distance effect is linear, so I adopt a piecewise formulation which allows for, but does not impose, an approximately linear distance effect. Thus I measure distance by five indicator variables, based on grouping countries into similar distances from the US:

1. adjacent to the US (Mexico and Canada).
2. between 1 and 4,000 kilometers (Caribbean islands and the northern coast of South America).
3. between 4,000km and 7,800km (Europe west of Russia, most of South America, a few countries on the West Coast of Africa)
4. between 7,800km and 14,000km (most of Asia and Africa, the Middle East, and, Argentina/Chile)
5. over 14,000km (Australia/New Zealand, Thailand, Indonesia, Malaysia/Singapore)

In some of the regressions, I aggregate the distance classes into near (less than 4,000km) and far. I also include a dummy for if a country is landlocked.

There are other factors that could affect unit values within 10-digit products, and I control for some of these. *Other controls* include

1. tariffs, measured as *ad valorem* percentage, which should have a negative sign to the extent that trade costs are borne by producers.

2. macro indicator of comparative advantage (log aggregate real GDP per worker, from the Penn-World Tables). My model is silent on how these aggregate measures might affect prices, but if more advanced countries specialize in more advanced and/or higher quality goods, we would expect positive effects of these variables on log unit values. Evidence of such effects is reported by Schott (2004), Hallak (2006) and Hummels and Klenow (2005).

Before turning to estimation of equation (15), it is informative to look at a plot of the data. Plot 2 shows the distribution of log U.S. import unit values from two groups of exporters: the NAFTA countries Canada and Mexico, and exporters whose goods cross the Atlantic or Pacific Ocean. I first remove the HS10 mean log unit values, so that the plot shows deviations from product-specific averages. There are two key features visible in Plot 2. The first is that the distribution for remote exporters is clearly shifted to the right relative to the distribution for Canada and Mexico: as predicted by the model, goods have higher unit values when they travel a greater distance. The second notable feature of the data is the extremely wide scale of the horizontal axis, from -4 to +4 (a few even more extreme values are trimmed in the interests of readability). This great range of log unit values is suggestive of substantial heterogeneity even within narrowly-defined HS10 categories.

Tables 7, 8, 9, and 10 report the results of estimating equation (15)[11]. For each year, log unit value is regressed on the controls as well as fixed effects for 10-digit HS codes. In the interest of reducing the quantity of numbers presented, I report results for only four selected years (1990, 1995, 2000, and 2003), although all regressions were estimated on all 14 years from 1990 to 2003 (complete results available on request). Each column shows results for a single year's regression, with *t*-statistics in *italics*.

The specifications in Tables 7 to 10 differ in the definition of the dependent variable and the scope of the sample. Tables 7 and 9 use the broadest definition of unit value, and include all observations for which units are reported, whether those units are number, barrels, dozens, kilos, or something else. Tables 8 and 10 include only observations for which weight in kilos is reported, so the unit value in the Tables 8 and 10 regressions is precisely the value-weight ratio for all of the observations. Tables 7 and 8 include all available observations, while Tables 9 and

[11] All regressions are estimated by OLS, with product fixed effects and robust standard errors. Note that standard errors are *not* adjusted for clustering by exporting country, despite the fact

10 restrict the sample in two ways. First, very small and potentially erratic observations are eliminated by dropping all import records of less than $10,000. Second, to focus attention on manufactured goods, Tables 9 and 10 include only HS codes that belong to SITC 6, 7, and 8. The differences in the coverage of the Tables are summarized here:

	definition of dependent variable	
	log unit value	log value/weight
all available observations	Table 7	Table 8
exclude small & non-manufactured obs	Table 9	Table 10

Tables 7 and 8 show that the effect of distance on unit values is large, robust, and statistically significant. The first column for each year in Tables 7 and 8 has a single indicator for distance greater than 4000km from the United States. As the first row of Table 7 shows, for the full sample unit values are around 30 log points higher when they come from more distant locations. The effect is even larger when the sample is restricted to observations with units in kilos, with the distance effect between 40 and 50 log points (first row, Table 8).

The second four rows of Tables 7 and 8 break down distance into a larger number of categories, with Mexico/Canada as the excluded category. A striking feature of these results is the non-monotonic effect of distance. For example, in Table 7 in 2003 the closest non-adjacent distance category is associated with unit values 15.9 log points higher than Mexico/Canada, an effect which jumps to 52 log points for the next category, before falling back to 10.4 and 16.8 log points in the final two distance categories. The pattern is similar in Table 8, but the effects are substantially larger. The additional effect of being landlocked is also large, ranging between 18 and 55 log points across specifications in Tables 7 and 8.

The estimated effects of distance are invariably larger in Table 8 than in Table 7. This discrepancy is supportive of the model's predictions, since the dependent variable in Table 8 (log value/kilo), is more closely connected to the theory than the dependent variable in Table 7 (log

---

that cross-product correlation of errors within an exporter is *a priori* plausible. For an explanation, see the Appendix.

unit value). To the extent that different units within a product have different weights (which they often do), one would expect a weaker connection between unit value and distance.

The puzzling non-monotonicity of the distance effect on unit values probably reflects imperfectly measured country characteristics that affect unit values and are correlated with distance, since the 4000-7800 range includes many of the most developed countries (including all of the EU exporters). The importance of development in affecting unit values was found in Schott (2004), Hallak (2006) and Hummels and Klenow (2005), and is confirmed here: a higher aggregate productivity level raises unit values with a large and significant elasticity, between 0.4 and 0.55, in every regression in Tables 7 and 8. The very large effect of the 4000-7800km category on unit values is suggestive of a non-linear effect of log GDP on unit values, and/or some other feature of the EU countries that leads them to specialize in high unit value products within HS codes. The fact that distance is correlated with GDP per capita is a fundamental feature of the data, and in the context of a within-product data specification like equation (15) it is not possible to more precisely isolate the separate effects of distance and development on unit values.

Although the tariff effects are not the focus of the paper, it is interesting that they are consistently estimated to be small (between -0.013 and -0.002), negative and statistically significant. These negative effects are consistent with the US being a large market for most exporters, and are suggestive of a small terms of trade gain from protection.

Tables 9 and 10 repeat the specifications of Tables 7 and 8 for a narrower sample, excluding non-manufacturing imports as well as very small observations (value of less than $10,000). The results are largely consistent with Tables 7 and 8, which confirms that the overall results are not driven by a small number of observations or by non-manufacturing SITC categories.

4.3 Statistical results: the shipping mode choice

A final empirical question concerns the choice between air and surface shipment by remote exporters. According to the model, the mechanism behind the within-product specialization documented in Tables 7 through 10 is that remote exporters are more likely to ship goods by air, and that these goods are "light" in the economically relevant sense of having low air transport costs as a share of value. Since transport charges per unit are more closely related to weight than value (as common observation as well and Hummels and Skiba (2004) show), transport charges as a share of value are declining in a goods value/weight ratio. Therefore, according to the model air shipment should be the mode of choice only for high value/weight (or "light") products from distant locations. In this section I test this prediction.

The implication of the model that nearby countries will not choose air shipment is confirmed by Chart 3: virtually all U.S. imports from Canada and Mexico come by surface transport. A challenge in testing the theory that air shipment is chosen for high value/weight goods is the endogeneity between value and shipment mode: the theory says that consumers are willing to pay a premium for air shipped goods, which is why suppliers (sometimes) choose air shipment. Thus, I need an instrument for a product's value/weight: a variable that is correlated with value/weight but unrelated to shipment choice. The fact that Canada and Mexico don't use air shipment suggests a potentially powerful instrument, which is value/weight of imports from Mexico and Canada. For a given HS10 code *i* and a given non-NAFTA exporter *c*, the value/weight of Mexican-Canadian good *i* is likely to be correlated with value/weight of good *i* from *c*, but should have no independent effect on the shipping mode choice from *c*.[12]

To test the mechanism that remote exporters are more likely to ship high value/weight goods by air, I estimate a discrete choice model of the shipping mode choices of all exporters except NAFTA. Defining the indicator variable $a_{ic} = 1$ if product *i* is shipped by air from exporter *c*, I estimate the following probit model for each year:

$$\Pr\left(a_{ic} = 1 \mid data\right) = \beta_0 + \beta_1 p_{ic} + \boldsymbol{\beta}_2 \mathbf{x}_c + \varepsilon_{ic} \qquad (16)$$

where p_{ic} is the log value per kilo of imports of product *i* from country *c*, $\mathbf{x}_c$ is a vector of country characteristics including distance indicators and log aggregate productivity, and ε_{ic} is normally distributed.

The air shipment indicator a_{ic} is coded as 1 if the share of imports sent by air for that product-country is greater than 0.9, and a_{ic} is coded as 0 is the share is less than 0.1. The estimation sample is substantially smaller than in the previous section because the following observations are excluded:

1. imports from NAFTA
2. products where weight in kilos is not reported

3. products not exported by NAFTA (needed for instrument)
4. products with non-trivial share of NAFTA imports arriving by air (very few products)
5. Observations where share of goods that arrive by air is between 0.1 and 0.9.

The estimator is maximum likelihood, with p_{ic} instrumented by the average value of p_i from Mexico and Canada. This is a strong instrument, with a simple correlation of around 0.6 between the instrument and p_{ic}. The estimated covariance matrix allows for heteroskedasticity and clustering by country.[13] Distance is measured using the same categories as in the previous section, but since NAFTA observations are excluded from the estimation, the excluded dummy variable becomes the distance category of 1-4000 kilometers. Results are reported in Table 11.

Looking at the bottom two rows of the table first, the instrument appears valid: the null that there is no correlation in the first stage is rejected, while the null that the instrument can be excluded in the second stage can not be rejected. Turning to the marginal effects, there is strong evidence that higher value/weight goods are more likely to be shipped by air: a one percent increase in value/weight leads to about a 0.2 percentage point increase in the probability of air shipment. Given the huge range in value/weight[14], this is a very large effect, and it is tightly estimated. By contrast, the other explanatory variables (distance, landlocked, and aggregate productivity) do not have statistically significant effects, especially in the later years.

4.4 Summarizing the evidence

Applying Proposition 1 to the United States, the model of Section 2 predicts that more distant exporters to the U.S. will specialize in light-weight goods which are shipped by air. The empirical analysis evaluated this prediction in three ways: by looking at market shares of different goods, within-product variation in unit values across exporters, and the determinants of shipping mode.

In section 4.2, I showed that Canada and Mexico have market shares that are on average about 9 percentage points higher in goods that other countries do not ship by air. This aggregate effect is mainly driven by the difference between Canada and other high-income exporters,

---

[12] For brevity, I will sometimes use "NAFTA" as a synonym for "Mexico and/or Canada", although the NAFTA agreement was not in force in the early years of my sample.
[13] See appendix for an important caveat about estimation of the covariance matrix.
[14] In 2003, the 5th-95th percentile range of log value/weight is [-0.2,5], which is a factor of more than 120 in levels.

especially in manufactured goods, where the effect is around 20 percentage points. These results are reported in Tables 5 and 6.

In section 4.3, I focused on within-category variation in import unit values. Thus the statistical model of equation (15) asks the question: if a country exports a good to the US, is the unit value of that good related to distance? The answer is yes, as shown in Tables 7 through 10. If we focus on the final year of the sample, we find that exports that arrive from destinations more than 4000km from the U.S. (that is, from sources other than Mexico, Canada, and the Caribbean) have import unit values between 25 and 40 log points higher than those from nearby sources.

A puzzling finding is that the effect of distance on unit values is non-monotonic, with the effect seemingly peaking in the distance range of 4000-7800km. Since this distance category includes Europe, the large estimated coefficients may be conflating the effect of distance and Europe's comparative advantage in producing high-quality goods. This is an important caveat for interpreting the size of the distance effect but does not overturn the strong relationship between distance and import unit value.

The results of Tables 7 through 10 confirm the importance of distance for unit values, but they don't say anything about the role of shipment mode choice. Table 11 fills in this gap, with the unsurprising finding that air shipment is strongly related to value/weight. Thus, we can conclude that the findings of Tables 7 through 10 are driven at least in part by the mechanism studied in the model: remote exporters specialize in lightweight goods which are shipped by air.

Lastly, the results confirm the results of Schott (2004), Hallak (2006), and Hummels-Klenow (2005) that there is an important relationship between import unit values and the level of development, which probably reflects a comparative advantage that rich countries have in high-quality goods.

5 Conclusion

This paper has focused on the interaction between trade, distance, transport costs, and the choice of shipping mode. In the theory model, I showed how the existence of airplanes implies that countries that are far from their major export markets will have a comparative advantage in lightweight goods. This prediction is strongly supported by the data.

In the empirical sections, I documented the heterogeneity across regions and goods of the prevalence of air shipment in US imports. The statistical analysis finds three strong and robust empirical relationships that support the model. The first is that Canada and Mexico have much higher market shares in goods which other countries do not ship by air. The second is that U.S. imports from remote suppliers have unit values on the order of a third higher than those from nearby countries. The third (and least surprising) result is that the probability of air shipment is strongly related to distance and unit value.

Distance is not dead, and the theory and empirical results of this paper suggest that it will not be expiring any time soon. The fall in the relative cost of air shipment implies that relative distance may become even more important in determining comparative advantage, as nearby countries increasingly trade heavy goods with each other while trading lighter goods with their more distant trading partners.

References

Baldwin, Richard, Rikard Forslid, Philippe Martin, Gianmarco Ottaviano, and Frederic Robert-Nicoud, 2003, *Economic Geography and Public Policy*, Princeton: Princeton University Press.

Deardorff, Alan, 2004, "Local Comparative Advantage: Trade Costs and the Pattern of Trade", University of Michigan RSIE Working paper no. 500.

Eaton, Jonathan and Samuel Kortum, 2002, "Technology, Geography, and Trade", *Econometrica 70(5)*: 1741-1779 (September).

Evans, Carolyn E., and James Harrigan, 2005, "Distance, Time, and Specialization: Lean Retailing in General Equilibrium" *American Economic Review 95(1)*: 292-313 (March).

Fujita, Masahisa, paul Krugman, and Anthony Venables, 1999, *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge, MA: MIT Press.

Hallak, Juan Carlos, 2006, "Product Quality and the Direction of Trade", *Journal of International Economic*s 68 (1), pp. 238-265.

Harrigan, James, and A.J. Venables, 2006, "Timeliness and Agglomeration", *Journal of Urban Economics 59*: 300-316 (March).

Hummels, David, 1999, "Have International Transportation Costs Declined?", http://www.mgmt.purdue.edu/faculty/hummelsd/

Hummels, David, 2001a, "Time as a Trade Barrier", http://www.mgmt.purdue.edu/faculty/hummelsd/

Hummels, David, 2001b, "Toward a Geography of Trade Costs", http://www.mgmt.purdue.edu/faculty/hummelsd/

Hummels, David, and Peter Klenow, 2005, "The Variety and Quality of a Nation's Exports", *American Economic Review 95(3)*: 704-723 (June).

Hummels, David, and Alexandre Skiba, 2004, "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture", *Journal of Political Economy 112(6)*: 1384-1402 (December).

Leamer, Edward, 1984, *Sources of International Comparative Advantage*, Cambridge, MA: MIT Press.

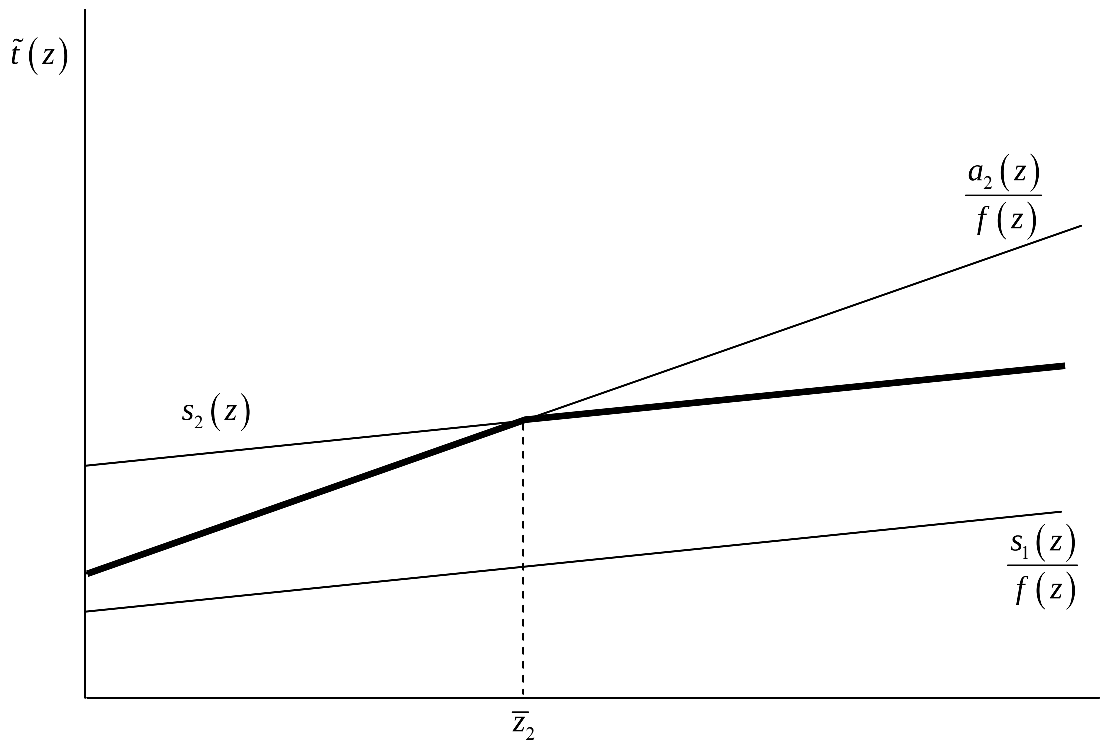
Limão, Nuno, and A.J. Venables, 2002, "Geographical Disadvantage: a Hecksher-Ohlin-Von Thunen Model of International Specialisation," *Journal of International Economics*

58(2): 239-63 (December).

Redding, Stephen, and A.J. Venables, 2004, "Economic Geography and International Inequality", *Journal of International Economics 62(1)*: 53-82.

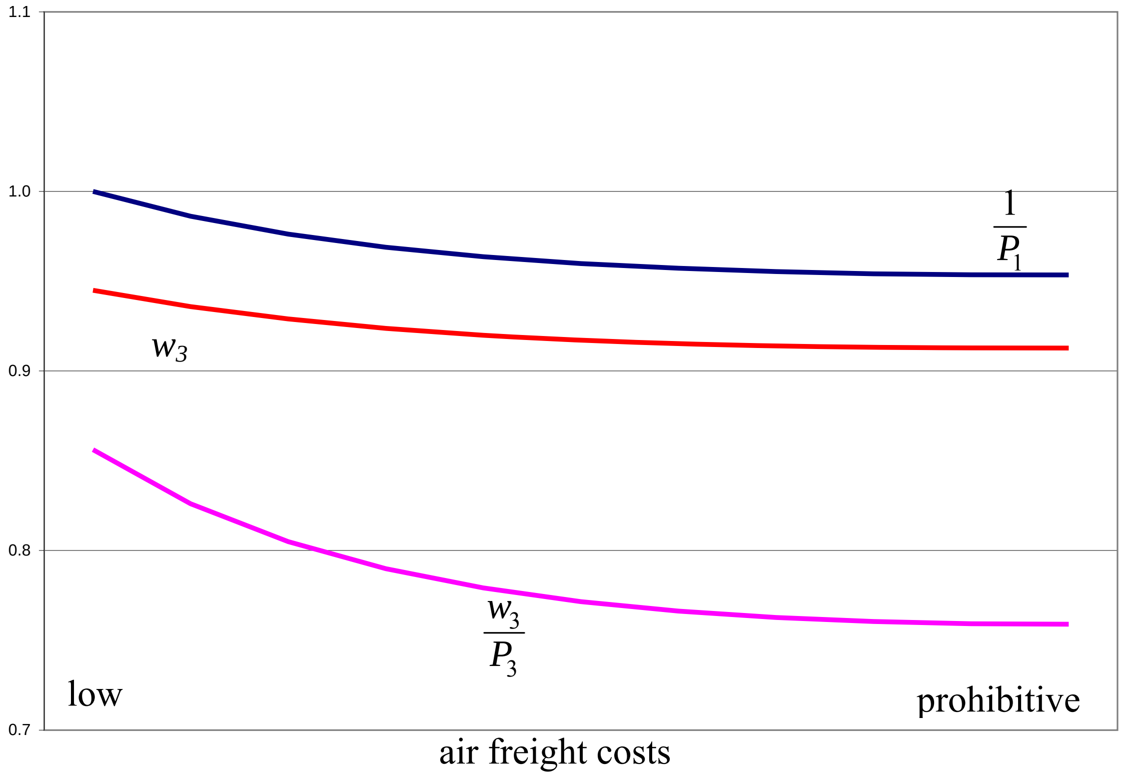
Schott, Peter K., 2004, “Across-product versus within-product specialization in international trade”, *Quarterly Journal of Economics119(2):*647-678 (May).

Figure 1 - Transport costs for nearby and faraway sellers in a given market



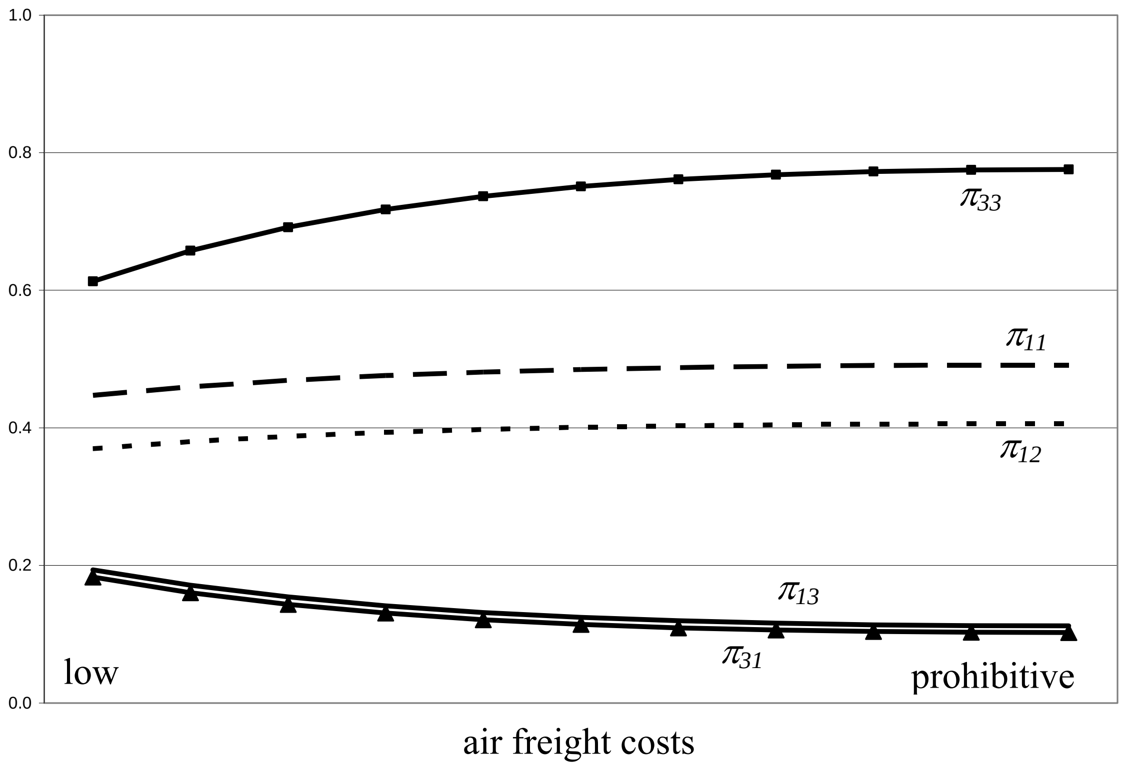
Notes to Figure 1: Goods are ordered from lightest to heaviest, so that all transport costs are increasing in *z*. The bold lower envelope $\tilde{t}_2(z) = \min\left[s_2(z), \frac{a_2(z)}{f(z)}\right]$ is the faraway country's minimized timeliness-adjusted transport cost schedule. For "light" goods $z \leq \overline{z}_2$, goods are shipped by air from country 2, while "heavy" goods are shipped by surface regardless of which country sells them.

Figure 2 - Real and nominal wages as a function of air transport costs



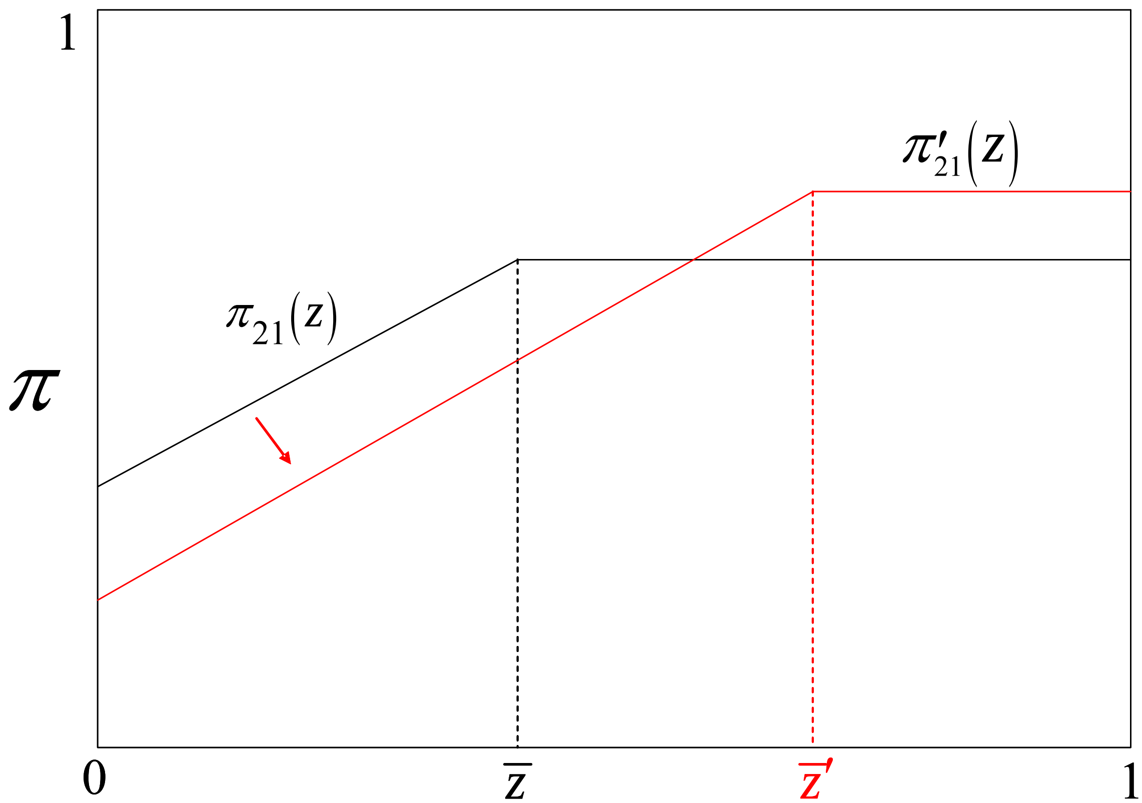
Notes to Figure 2: Illustrates equilibrium wages as a function of air freight costs for a numerical example, with air freight costs varying from prohibitive (right axis) to low enough so that country 3 always uses air freight (left axis). The nominal wage in country 1 is the numeraire and is set equal to one, and country 1 real wages are normalized to one when air freight costs are low. Equilibria are computed as the shift parameter β falls from 1 to f^{-1}. Parameter values are $f = \theta = 2$, $\tau_3 = 1.2$, $\tau_1 = 1.1$, $L_1 = L_2 = L_3 = 1$.

Figure 3 - Aggregate Market shares as a function of air transport costs



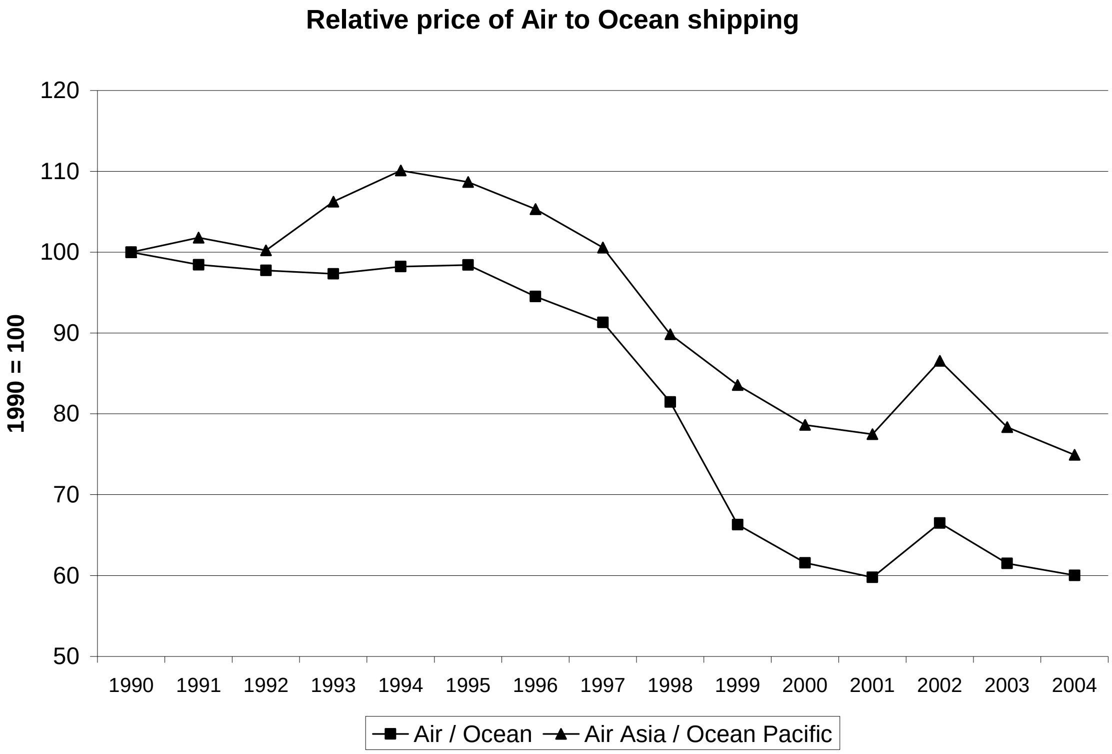
Notes to Figure 3**:** Illustrates equilibrium aggregate market shares (that is, the share of country *o* production in country *d* expenditure) as a function of air freight costs for a numerical example, with air freight costs varying from prohibitive (right axis) to low enough so that country 3 always uses air freight (left axis). Parameter values are the same as in Figure 2.

Figure 4 - Change in market shares when air freight costs fall



Notes to Figure 4: Illustrates nearby country 2's import market share in country 1. The shift in the market share schedule from $\pi_{21}(z)$ to $\pi'_{21}(z)$ is a consequence of a fall in the cost of air shipment, which causes a shift in the cutoff for air shipment by faraway country 3 from $\bar{z}$ to $\bar{z}'$. The result is lower nominal wages for country 2 relative to country 3, and a greater specialization in heavy goods by country 2.

Chart 1



Notes to Chart 1: Data are price indices for U.S. imports of air freight and ocean liner shipping services from the Bureau of Labor Statistics, www.bls.gov/mxp. The "Air/Ocean" series divides all US imports of air freight services by all imports of ocean liner services, while the "Air Asia / Ocean Pacific" series divides the index for air freight imports from Asia by the index for ocean liner imports from the Pacific region.

Chart 2

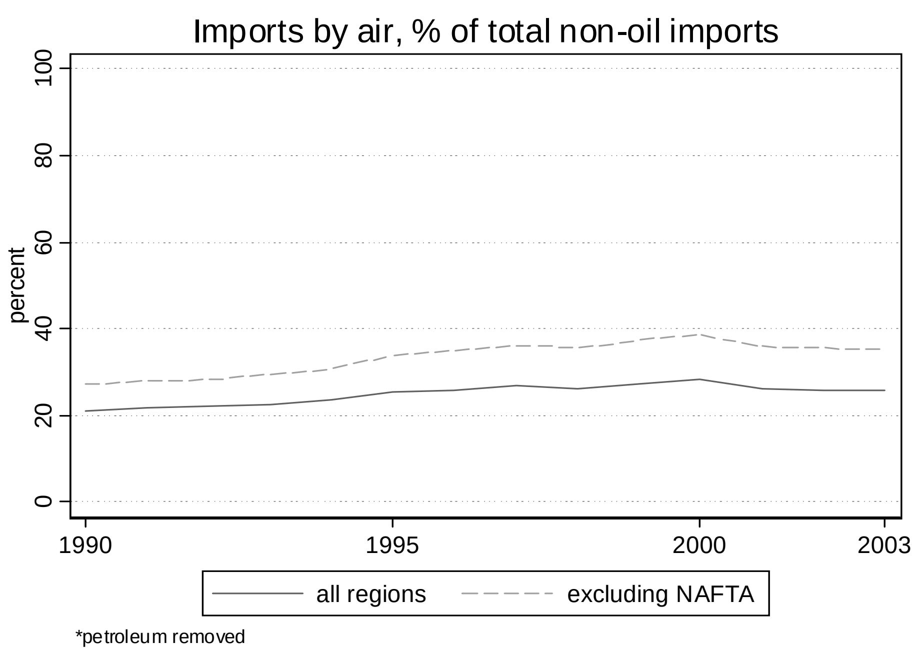


Chart 3

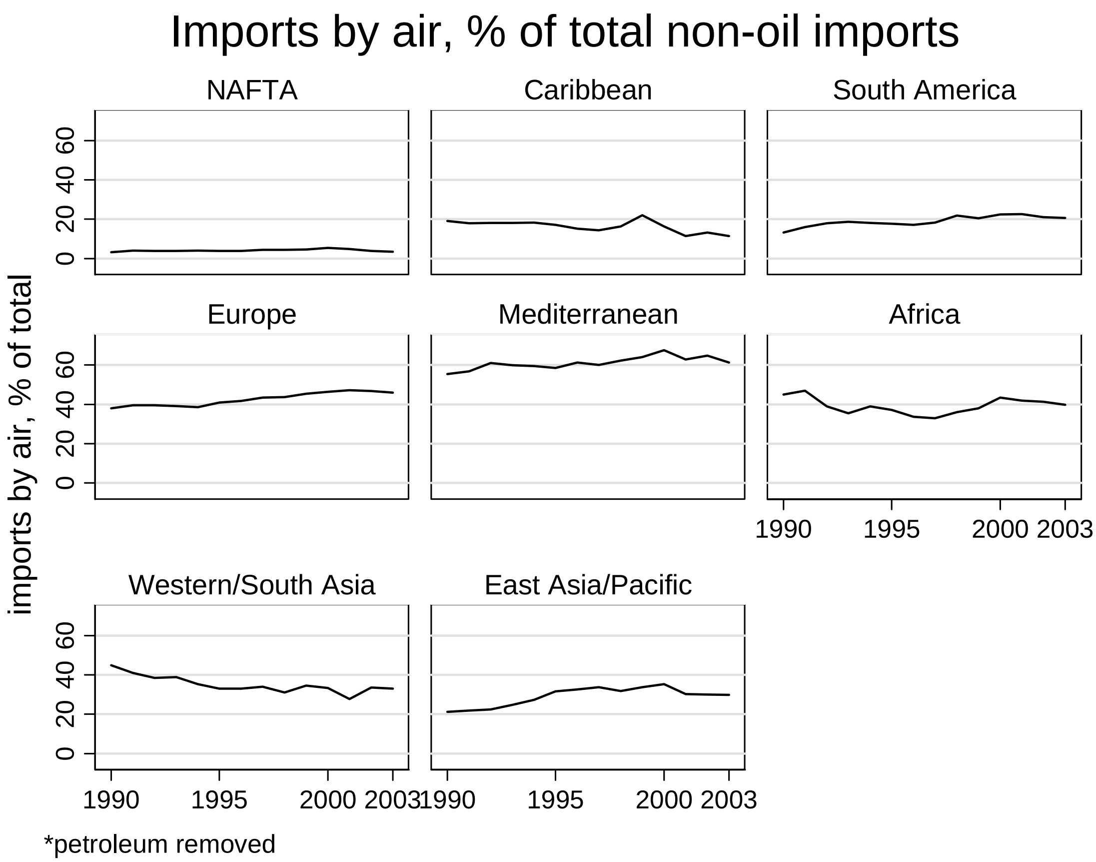


Chart 4

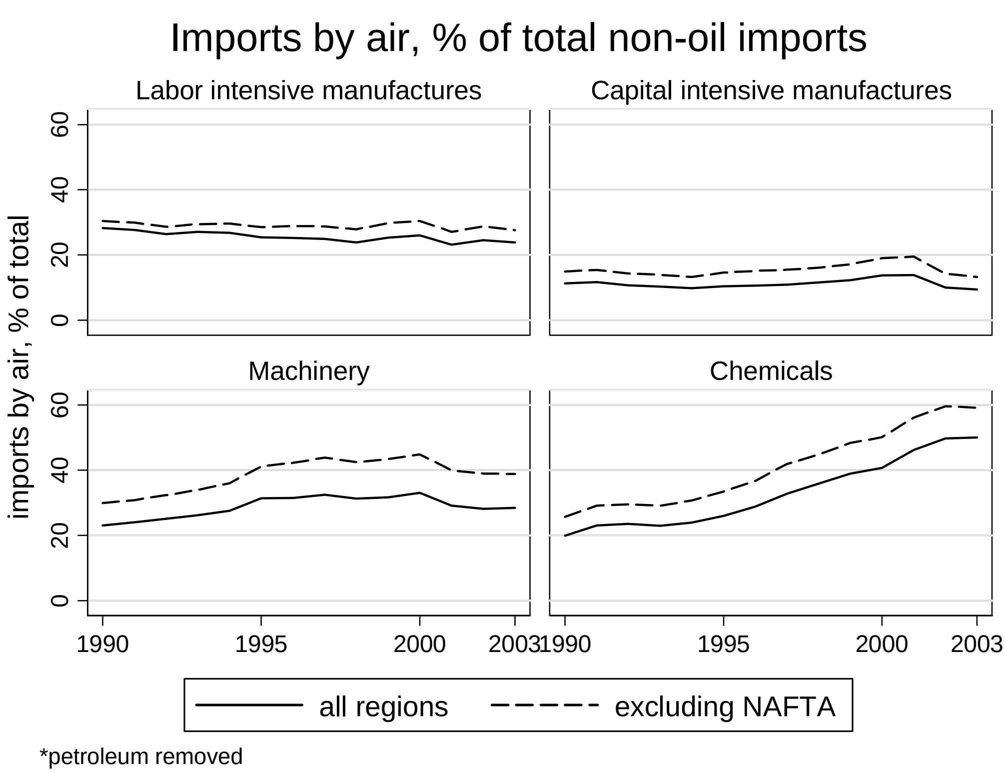


Chart 5

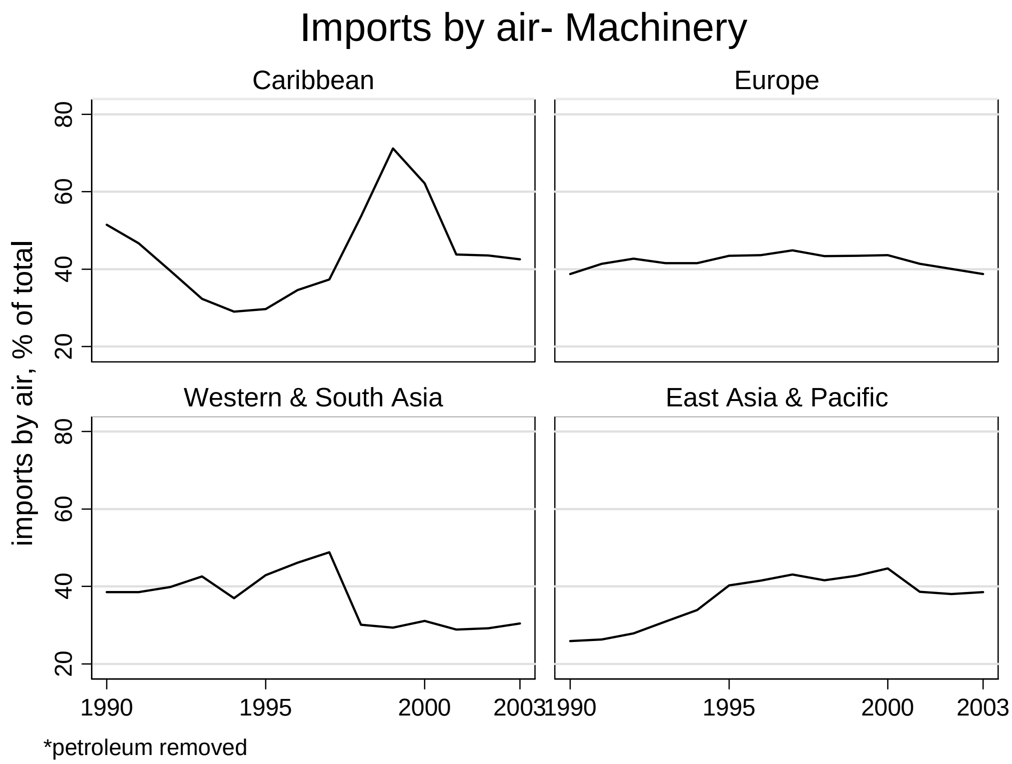


Chart 6

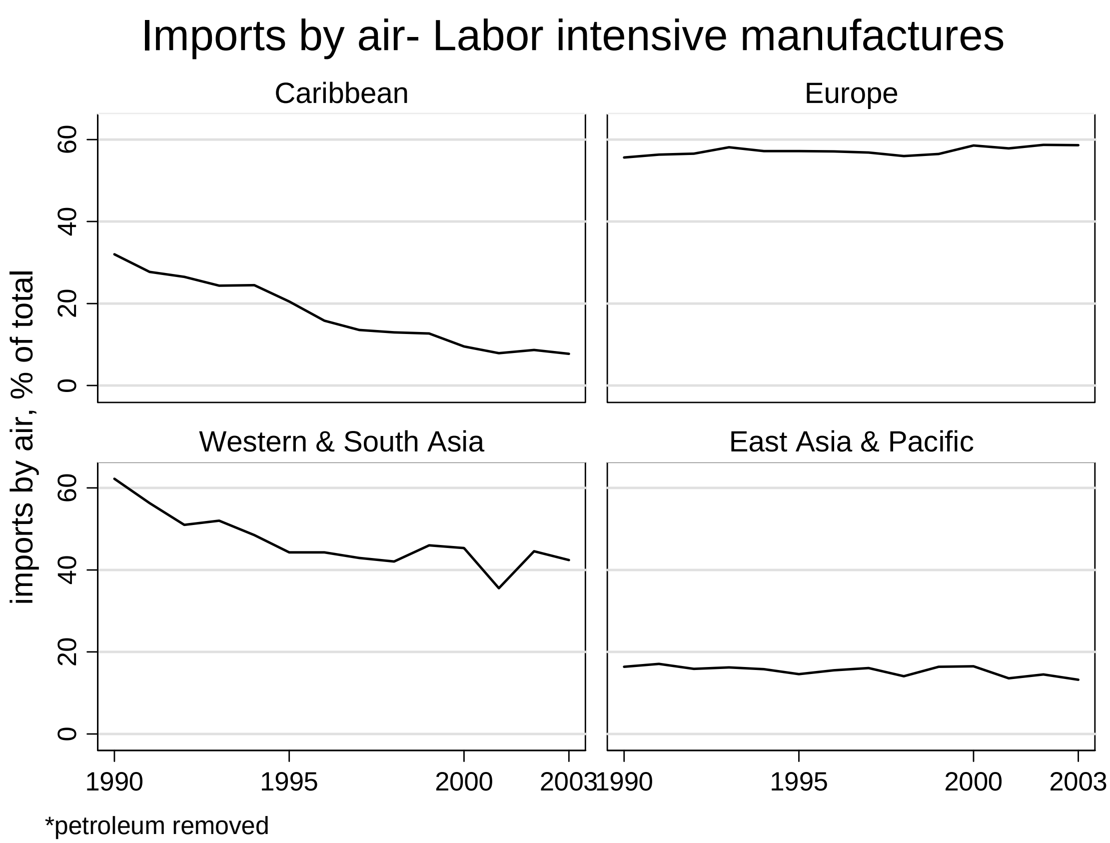


Chart 7

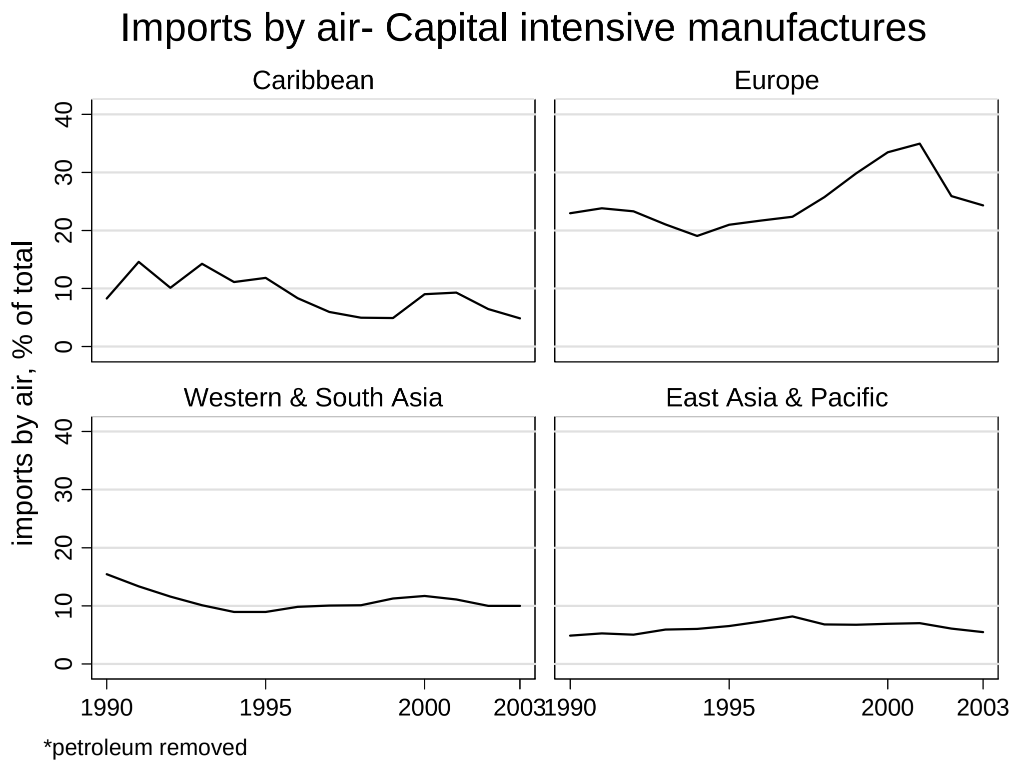


Chart 8

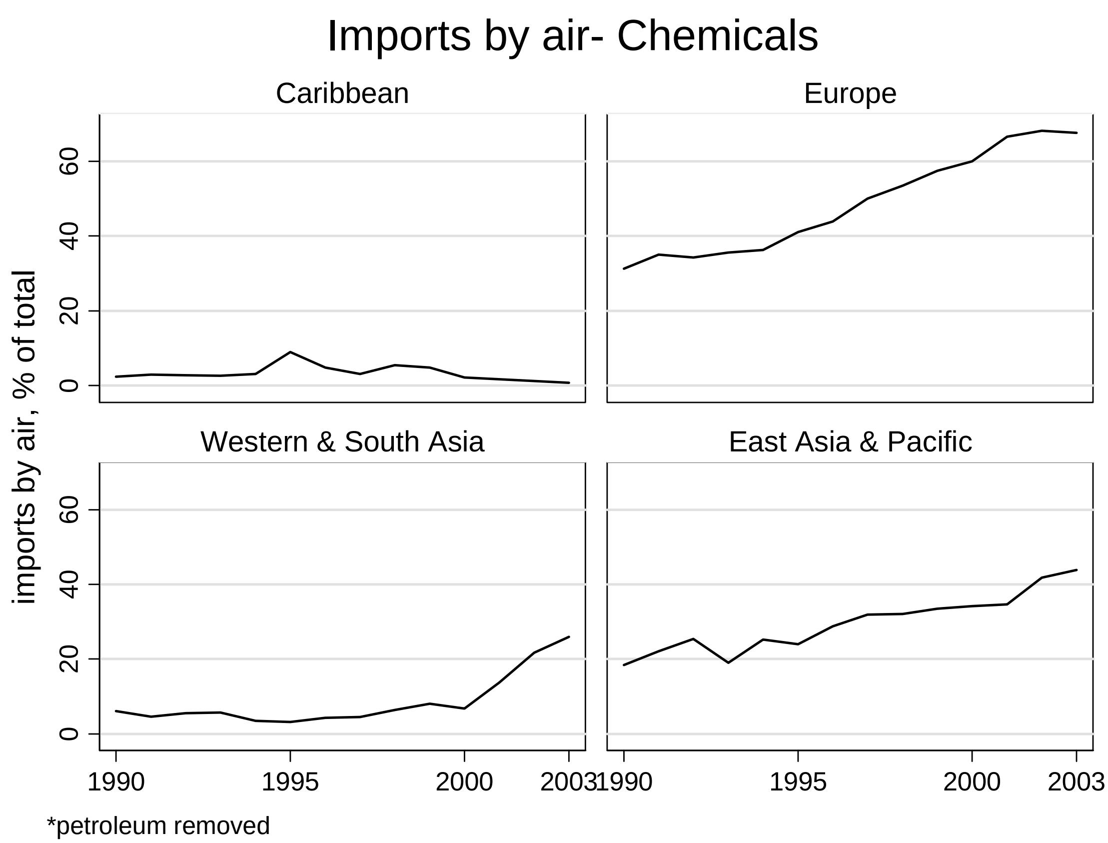


Table 1 Imports by product and percent air shipped, 1990 and 2003

Share of total 1990	Share of total 2003	SITC	% of category 1990	Imports by air, % of total 1990	% of category 2003	Imports by air, % of total 2003	SITC description
12.2	**8.9**	**Petroleum**					
		33	100.0	2.4	100.0	2.9	Petroleum, petroleum products and related materials
2.2	**3.3**	**Other fuel & raw materials**					
		28	35.0	9.1	7.5	13.6	Metalliferous ores and metal scrap
		34	30.6	1.4	74.3	7.6	Gas, natural and manufactured
		23	11.2	6.8	4.5	10.0	Crude rubber
		27	10.6	19.4	5.7	19.6	Crude fertilizers and crude minerals
		26	5.5	7.7	1.6	29.1	Textile fibres
		35	4.3	0.0	3.4	0.0	Electric current
		32	2.7	1.7	2.9	0.3	Coal, coke and briquettes
3.4	**2.8**	**Forest products**					
		64	51.2	23.8	43.5	19.5	Paper, paperboard and articles thereof
		24	18.9	4.9	21.2	4.0	Cork and wood
		25	17.3	12.3	7.6	9.5	Pulp and waste paper
		63	12.6	18.3	27.7	13.0	Cork and wood manufactures (excluding furniture)
5.7	**4.7**	**Animal and vegetable products**					
		05	19.8	7.4	19.5	5.8	Vegetables and fruit
		03	18.5	31.2	17.4	26.6	Fish, crustaceans, molluscs and aquatic invertebrates
		11	12.7	4.7	17.9	2.7	Beverages
		07	12.0	6.2	9.0	8.2	Coffee, tea, cocoa, spices, and manufactures thereof
		01	10.5	15.0	7.6	14.7	Meat and meat preparations
		29	4.4	36.8	4.8	40.6	Crude animal and vegetable materials, n.e.s.
		06	4.3	3.0	3.7	3.9	Sugars, sugar preparations and honey
		0	4.3	86.6	2.8	83.8	Live animals other than animals of division 03
		04	3.2	5.6	5.6	4.0	Cereals and cereal preparations
		12	2.3	16.2	2.2	14.1	Tobacco and tobacco manufactures

Table 1 Imports by product and percent air shipped, 1990 and 2003

		42	2.3	3.7	2.3	6.6	Fixed vegetable fats and oils, crude, refined or fractionated
		02	1.7	18.9	1.9	14.2	Dairy products and birds' eggs
		09	1.4	6.9	3.2	7.9	Miscellaneous edible products and preparations
		08	1.2	9.6	1.2	5.2	Feeding stuff for animals (not including unmilled cereals)
		22	0.7	11.8	0.5	10.8	Oil-seeds and oleaginous fruits
		21	0.6	44.2	0.2	61.1	Hides, skins and furskins, raw
		43	0.2	3.1	0.3	10.3	Animal or vegetable fats and oils
		41	0.1	13.4	0.1	18.7	Animal oils and fats
15.3	**15.9**	**Labor intensive manufactures**					
		84	33.3	56.5	31.4	47.5	Articles of apparel and clothing accessories
		89	32.9	47.0	32.6	46.4	Miscellaneous manufactured articles, n.e.s.
		85	12.8	46.3	7.7	50.1	Footwear
		66	11.6	28.9	13.7	24.5	Non-metallic mineral manufactures, n.e.s.
		82	6.6	12.8	12.2	12.3	Furniture, and parts thereof
		83	2.9	60.9	2.3	58.2	Travel goods, handbags and similar containers
8.1	**7.0**	**Capital intensive manufactures**					
		67	24.5	3.3	14.5	7.1	Iron and steel
		68	23.4	17.0	19.4	25.2	Non-ferrous metals
		69	22.2	24.2	28.5	29.4	Manufactures of metals, n.e.s.
		65	15.8	45.8	19.7	46.3	Textile yarn, fabrics, made-up articles, n.e.s
		62	8.7	14.6	9.7	27.0	Rubber manufactures, n.e.s.
		81	3.1	18.4	6.9	20.0	Prefabricated buildings, lighting & plumbing fixtures
		61	2.2	59.1	1.3	65.3	Leather, leather manufactures, n.e.s., and dressed furskins
45.2	**45.0**	**Machinery**					
		78	34.2	14.7	31.1	19.4	Road vehicles (including air-cushion vehicles)
		77	15.0	55.3	14.6	60.9	Electrical machinery, apparatus and appliances, n.e.s., and parts
		75	12.3	73.6	14.5	74.5	Office machines and automatic data-processing machines
		76	9.7	57.6	12.7	74.5	Telecommunications and sound-recording/reproducing apparatus
		74	6.4	31.5	6.8	34.7	General industrial machinery and equipment and parts, n.e.s.
		72	5.9	30.2	3.7	35.7	Machinery specialized for particular industries
		71	5.8	40.1	5.7	43.9	Power-generating machinery and equipment

Table 1 Imports by product and percent air shipped, 1990 and 2003

		79	3.3	42.0	3.6	42.8	Other transport equipment
		88	3.0	68.2	2.1	73.7	Photographic apparatus, optical goods, watches
		87	2.7	64.0	4.2	75.5	Professional, scientific and controlling instruments, n.e.s.
		73	1.7	28.9	1.0	41.4	Metalworking machinery
4.5	**8.3**	**Chemicals**					
		51	32.7	23.4	33.5	32.2	Organic chemicals
		52	14.0	11.1	7.1	16.8	Inorganic chemicals
		54	11.7	54.6	31.2	65.0	Medicinal and pharmaceutical products
		59	9.2	10.9	6.7	21.6	Chemical materials and products, n.e.s.
		57	8.9	7.2	7.1	19.5	Plastics in primary forms
		58	8.0	24.2	4.6	27.2	Plastics in non-primary forms
		53	5.9	10.4	2.4	18.4	Dyeing, tanning and colouring materials
		55	5.3	28.3	5.4	22.7	Essential oils and resinoids and perfume; cleanser
		56	4.2	8.7	2.0	6.9	Fertilizers (other than those of group 272)

Table 2 Country categories

distance	country	region	country	region
0 km from USA	Canada	NAFTA	Mexico	NAFTA
1-4000 km from USA	Bahamas	Caribbean	Barbados	Caribbean
	Belize	Caribbean	Costa.Rica	Caribbean
	Dominican.Rep.	Caribbean	El.Salvador	Caribbean
	Guatemala	Caribbean	Haiti	Caribbean
	Honduras	Caribbean	Jamaica	Caribbean
	Nicaragua	Caribbean	Panama	Caribbean
	TrinidadTobago	Caribbean	Colombia	South America
	Venezuela	South America		
4000-7800 km from USA	Bolivia	South America	Brazil	South America
	Ecuador	South America	Guyana	South America
	Paraguay	South America	Peru	South America
	Suriname	South America	Austria	Europe
	Belgium-Lux	Europe	Czechoslovakia	Europe
	Denmark	Europe	Finland	Europe
	France	Europe	Germany	Europe
	Hungary	Europe	Iceland	Europe
	Ireland	Europe	Italy	Europe
	Netherlands	Europe	Norway	Europe
	Poland	Europe	Portugal	Europe
	Spain	Europe	Sweden	Europe
	Switzerland	Europe	United.Kingdom	Europe
	Yugoslavia	Europe	Algeria	Mediterranean
	Malta	Mediterranean	Morocco	Mediterranean
	Tunisia	Mediterranean	Gambia	Africa
	Guinea	Africa	Guinea.Bissau	Africa
	Liberia	Africa	Mali	Africa
	Mauritania	Africa	Senegal	Africa
	Sierra.Leone	Africa		

Table 2 Country categories, continued

distance	country	region	country	region
7800-14000 km from USA	Argentina	South America	Chile	South America
	Uruguay	South America	Bulgaria	Europe
	Romania	Europe	Russia	Europe
	Cyprus	Mediterranean	Egypt	Mediterranean
	Greece	Mediterranean	Israel	Mediterranean
	Syria	Mediterranean	Turkey	Mediterranean
	Angola	Africa	Benin	Africa
	Burkina Faso	Africa	Burundi	Africa
	Cameroon	Africa	CentAfrRepublic	Africa
	Chad	Africa	Comoros	Africa
	Congo	Africa	Côte d'Ivoire	Africa
	Djibouti	Africa	Ethiopia	Africa
	Gabon	Africa	Ghana	Africa
	Kenya	Africa	Malawi	Africa
	Mozambique	Africa	Niger	Africa
	Nigeria	Africa	Rwanda	Africa
	Somalia	Africa	South.Africa	Africa
	Sudan	Africa	Tanzania	Africa
	Togo	Africa	Uganda	Africa
	Zaire	Africa	Zambia	Africa
	Zimbabwe	Africa	Afghanistan	W/S Asia
	Bahrain	W/S Asia	Bangladesh	W/S Asia
	Bhutan	W/S Asia	India	W/S Asia
	Iran	W/S Asia	Iraq	W/S Asia
	Jordan	W/S Asia	Kuwait	W/S Asia
	Mongolia	W/S Asia	Myanmar	W/S Asia
	Nepal	W/S Asia	Oman	W/S Asia
	Pakistan	W/S Asia	Qatar	W/S Asia
	Saudi.Arabia	W/S Asia	UAE	W/S Asia
	Yemen	W/S Asia	China	E Asia/Pacific
	Fiji	E Asia/Pacific	Hong.Kong	E Asia/Pacific
	Japan	E Asia/Pacific	Korea.RP.(S)	E Asia/Pacific
	Laos	E Asia/Pacific	Phillipines	E Asia/Pacific
	Solomon.Islands	E Asia/Pacific	Taiwan	E Asia/Pacific
over 14000 km from USA	Madagascar	Africa	Mauritius	Africa
	Seychelles	Africa	Reunion	W/S Asia
	Sri Lanka	W/S Asia	Australia	E Asia/Pacific
	Indonesia	E Asia/Pacific	Malaysia	E Asia/Pacific
	New Zealand	E Asia/Pacific	PapuaNGuinea	E Asia/Pacific
	Singapore	E Asia/Pacific	Thailand	E Asia/Pacific

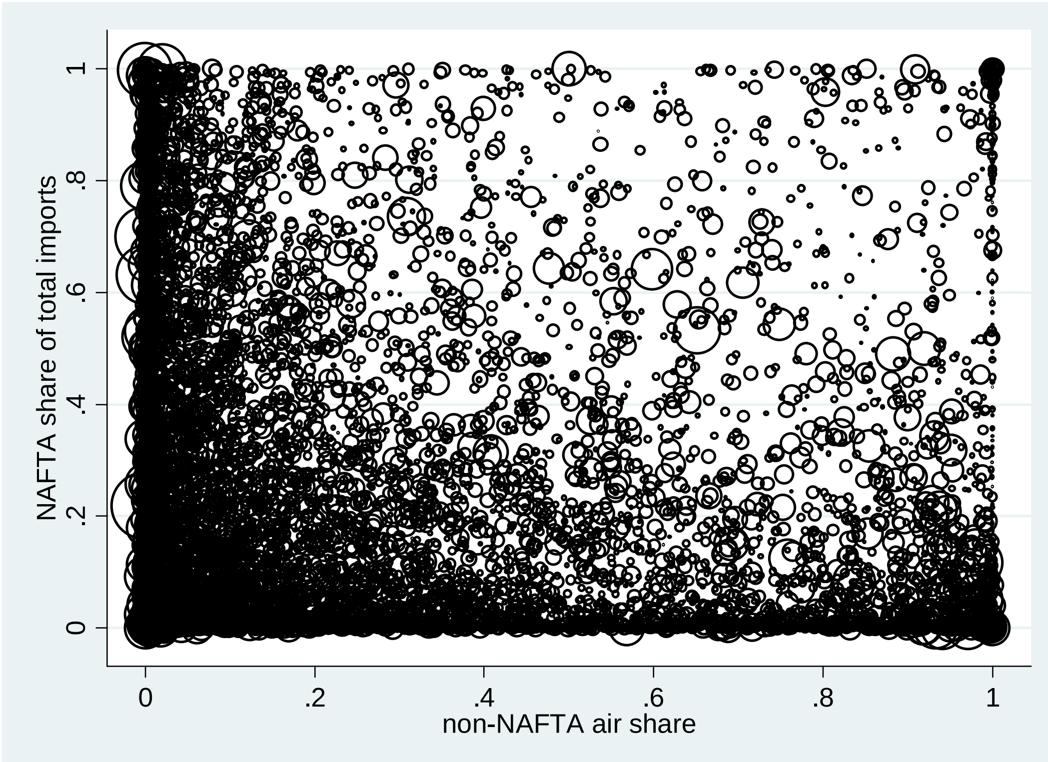
Table 3- Transport costs by region, 2003

	transport cost, % of import value	air freight cost, % of air value
NAFTA	1.50	5.17
Caribbean	2.34	6.47
South America	9.17	7.04
Europe	4.45	4.96
Mediterranean	5.09	10.18
Africa	7.02	14.57
Western/South Asia	7.12	15.38
East Asia/Pacific	6.17	12.76

Table 4- Transport costs by product, 2003

	transport cost, % of import value	air freight cost, % of air value
Petroleum	5.00	22.37
Other fuel & raw materials	4.74	3.76
Forest products	6.44	20.88
Animal and vegetable products	7.30	23.77
Labor intensive manufactures	5.71	4.43
Capital intensive manufactures	5.48	6.97
Machinery	1.97	2.37
Chemicals	2.73	1.04

Plot 1 - NAFTA share of U.S. imports vs. non-NAFTA air shipment share, 2003



Notes to Plot 1 Each point corresponds to an HS10 product, and shows the share of total imports of that product that comes from Canada and Mexico plotted against the share of non-NAFTA imports of that product that arrives by air. Circle sizes are proportional to the square root of total imports in the HS10 code.

Table 5 Linear market share regressions

	all exporters				high income exporters				middle income exporters			
	1990	1995	2000	2003	1990	1995	2000	2003	1990	1995	2000	2003
all products, continuous air share												
air share	0.066	0.057	0.054	**0.058**	0.072	0.085	0.093	**0.111**	-0.034	-0.042	-0.067	**-0.050**
	9.6	*8.0*	*7.7*	***8.2***	*10.8*	*12.2*	*13.9*	***16.6***	*-2.9*	*-4.0*	*-6.8*	***-5.4***
all products, binary air share indicator												
air share = 0	-0.048	-0.046	-0.021	-0.038	-0.060	-0.063	-0.051	-0.068	0.034	0.044	0.069	0.053
	-9.3	*-8.4*	*-4.0*	*-7.3*	*-11.1*	*-10.4*	*-8.1*	*-10.6*	*4.7*	*7.1*	*12.1*	*10.5*
air share = 1	0.012	-0.010	0.027	0.022	0.003	-0.002	0.031	0.035	0.030	0.031	0.051	0.050
	1.1	*-0.9*	*2.6*	*2.0*	*0.3*	*-0.2*	*3.4*	*4.0*	*2.0*	*2.3*	*3.6*	*3.9*
(air share = 1) -	0.060	0.036	0.049	**0.060**	0.063	0.061	0.081	**0.103**	-0.004	-0.013	-0.018	**-0.003**
(air share = 0)	*5.4*	*3.1*	*4.3*	***5.2***	*5.9*	*5.6*	*7.9*	***10.0***	*-0.3*	*-0.9*	*-1.3*	***-0.3***
sample size	12,537	13,659	14,469	14,783	12,005	12,918	13,622	13,701	7,093	8,829	10,572	11,614
manufacturing products, continuous air share												
air share	0.078	0.071	0.064	**0.063**	0.080	0.090	0.095	**0.114**	-0.029	-0.040	-0.061	**-0.046**
	11.7	*9.5*	*8.5*	***8.4***	*12.6*	*12.2*	*13.1*	***15.9***	*-2.1*	*-3.3*	*-5.3*	***-4.4***
manufacturing products, binary air share indicator												
air share = 0	-0.043	-0.046	-0.012	-0.032	-0.056	-0.066	-0.044	-0.060	0.022	0.035	0.064	0.050
	-7.1	*-6.7*	*-1.8*	*-4.9*	*-8.9*	*-8.6*	*-5.6*	*-7.3*	*2.7*	*4.9*	*9.5*	*8.4*
air share = 1	0.057	0.033	0.050	0.048	0.042	0.024	0.040	0.055	0.020	0.025	0.047	0.057
	6.6	*3.0*	*4.5*	*4.2*	*5.3*	*2.4*	*4.2*	*6.1*	*1.2*	*1.5*	*2.8*	*3.8*
(air share = 1) -	0.100	0.079	0.062	**0.080**	0.098	0.090	0.084	**0.114**	-0.003	-0.010	-0.017	**0.007**
(air share = 0)	*10.2*	*6.4*	*5.0*	***6.4***	*10.3*	*7.6*	*7.2*	***10.0***	*-0.1*	*-0.6*	*-1.0*	***0.5***
sample size	9,042	9,783	10,453	10,610	8,838	9,495	10,104	10,120	5,335	6,557	7,852	8,526
non-manufacturing products												
air share	-0.050	-0.036	-0.019	**0.003**	-0.038	0.005	0.026	**0.045**	-0.029	-0.027	-0.048	**-0.044**
	-2.4	*-1.9*	*-1.0*	***0.2***	*-1.8*	*0.3*	*1.5*	***2.5***	*-1.2*	*-1.3*	*-2.4*	***-2.3***
non-manufacturing products, binary air indicator												
air share = 0	-0.013	-0.018	-0.011	-0.020	-0.017	-0.026	-0.022	-0.044	0.067	0.061	0.042	0.050
	-1.2	*-1.7*	*-1.0*	*-2.0*	*-1.5*	*-2.3*	*-1.9*	*-3.8*	*3.4*	*3.9*	*3.4*	*4.3*
air share = 1	-0.102	-0.113	-0.030	-0.027	-0.094	-0.081	0.008	-0.011	0.077	0.050	0.035	0.025
	-3.4	*-4.2*	*-1.2*	*-1.1*	*-3.2*	*-3.0*	*0.4*	*-0.5*	*2.5*	*1.8*	*1.3*	*1.0*
(air share = 1) -	-0.089	-0.095	-0.020	**-0.006**	-0.076	-0.055	0.030	**0.033**	0.011	-0.011	-0.007	**-0.025**
(air share = 0)	*-3.0*	*-3.5*	*-0.8*	***-0.3***	*-2.6*	*-2.0*	*1.3*	***1.4***	*0.4*	*-0.5*	*-0.3*	***-1.0***
sample size	3,495	3,876	4,016	4,173	3,167	3,423	3,518	3,581	1,758	2,272	2,720	3,088

Table 6 Tobit market share regressions

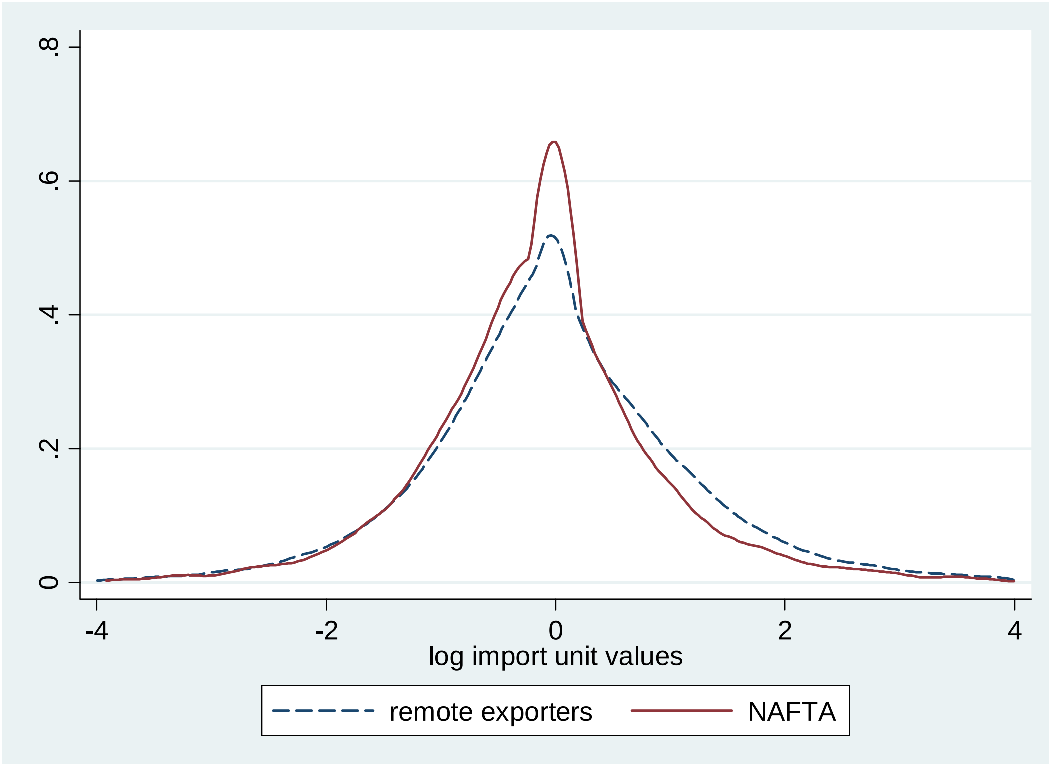
	all exporters				high income exporters				middle income exporters			
	1990	1995	2000	2003	1990	1995	2000	2003	1990	1995	2000	2003
all products, continuous air share												
air share	0.091	0.076	0.082	**0.090**	0.105	0.112	0.134	**0.166**	-0.093	-0.083	-0.123	**-0.096**
	8.3	*7.1*	*8.0*	***8.6***	*9.5*	*10.7*	*13.0*	***15.6***	*-3.7*	*-4.2*	*-6.8*	***-5.3***
all products, binary air share indicator												
air share = 0	0.002	-0.001	0.016	-0.002	-0.012	-0.017	-0.010	-0.032	0.189	0.181	0.194	0.184
	0.2	*-0.1*	*2.1*	*-0.3*	*-1.5*	*-1.9*	*-1.1*	*-3.4*	*11.7*	*14.8*	*17.5*	*17.5*
air share = 1	0.157	0.081	0.130	0.132	0.145	0.103	0.150	0.172	0.204	0.181	0.180	0.218
	7.9	*4.5*	*7.3*	*7.1*	*7.4*	*5.9*	*9.1*	*10.3*	*5.6*	*6.3*	*6.3*	*7.3*
(air share = 1) -	0.155	0.082	0.114	**0.134**	0.157	0.120	0.160	**0.204**	0.015	-0.000	-0.014	**0.034**
(air share = 0)	*7.5*	*4.4*	*6.1*	***7.0***	*7.7*	*6.3*	*8.9*	***11.0***	*0.4*	*-0.0*	*-0.5*	***1.1***
sample size	12,537	13,659	14,469	14,783	12,005	12,918	13,622	13,701	7,093	8,829	10,572	11,614
manufacturing products, continuous air share												
air share	0.129	0.111	0.110	**0.110**	0.133	0.134	0.154	**0.181**	-0.052	-0.044	-0.081	**-0.059**
	12.1	*10.3*	*10.4*	***10.2***	*12.7*	*12.2*	*14.5*	***16.7***	*-1.9*	*-2.0*	*-4.1*	***-3.1***
manufacturing products, binary air share indicator												
air share = 0	-0.016	-0.020	0.014	-0.009	-0.032	-0.037	-0.018	-0.036	0.151	0.134	0.160	0.159
	-1.9	*-2.2*	*1.7*	*-1.0*	*-3.6*	*-3.6*	*-1.7*	*-3.2*	*8.6*	*10.2*	*13.2*	*13.9*
air share = 1	0.232	0.138	0.155	0.165	0.205	0.135	0.155	0.193	0.198	0.165	0.168	0.228
	11.2	*7.2*	*8.3*	*8.2*	*10.5*	*7.4*	*9.2*	*11.2*	*4.9*	*5.2*	*5.4*	*6.8*
(air share = 1) -	0.248	0.157	0.141	**0.173**	0.237	0.172	0.173	**0.229**	0.046	0.031	0.008	**0.070**
(air share = 0)	*11.4*	*7.7*	*7.1*	***8.2***	*11.3*	*8.8*	*9.1*	***11.5***	*1.2*	*0.9*	*0.3*	***2.0***
sample size	9,042	9,783	10,453	10,610	8,838	9,495	10,104	10,120	5,335	6,557	7,852	8,526
non-manufacturing products												
air share	-0.065	-0.033	-0.022	**0.016**	-0.026	0.044	0.046	**0.101**	-0.160	-0.116	-0.133	**-0.125**
	-2.0	*-1.1*	*-0.8*	***0.6***	*-0.8*	*1.3*	*1.4*	***3.1***	*-2.6*	*-2.1*	*-2.6*	***-2.5***
non-manufacturing products, binary air indicator												
air share = 0	0.04	0.011	0.015	0.008	0.040	0.004	0.015	-0.019	0.230	0.218	0.136	0.157
	2.3	*0.6*	*0.9*	*0.5*	*1.9*	*0.2*	*0.8*	*-0.9*	*4.9*	*5.5*	*4.1*	*5.1*
air share = 1	-0.060	-0.099	0.020	0.028	-0.039	-0.037	0.100	0.085	0.184	0.164	0.116	0.102
	-1.3	*-2.4*	*0.5*	*0.7*	*-0.8*	*-0.8*	*2.2*	*2.0*	*2.2*	*2.2*	*1.6*	*1.4*
(air share = 1) -	-0.102	-0.109	0.005	**0.020**	-0.079	-0.042	0.085	**0.104**	-0.046	-0.054	-0.020	**-0.055**
(air share = 0)	*-2.3*	*-2.7*	*0.1*	***0.5***	*-1.7*	*-0.9*	*1.9*	***2.4***	*-0.6*	*-0.8*	*-0.3*	***-0.8***
sample size	3,495	3,876	4,016	4,173	3,167	3,423	3,518	3,581	1,758	2,272	2,720	3,088

Notes to Tables 5 and 6: Estimation of equation (14) in the text. Dependent variable is aggregate market share in U.S. imports of exporters other than Canada and Mexico. Unit of observation is an HS 10 code. Robust *t*-statistics are in *italics*.

For each sample, two specifications are estimated: with a continuous measure of air share, and with two indicators for air share = 0 and 1. In this second specification, the Tables report the point estimate and *t*-statistic on the difference between the coefficients on the indicators.

For Table 5, estimation is OLS. For Table 6, estimation is double sided Tobit, with upper and lower censoring of 1 and 0 respectively. The most interesting numbers in the Tables are rendered in **bold**.

Plot 2 - Densities of log import unit values, 2003



Notes to Plot 2 Kernel densities of log import unit values, expressed as deviations from HS10 means. Density labeled NAFTA is all log unit values from Canada and Mexico, density labeled "remote exporters" includes observations from exporters more than 4000 km from the United States (the plot excludes observations from exporters less than 4000 km other than Canada/Mexico).

Table 7 - Regression of U.S. import unit values on distance and other controls, all available observations

	1990		1995		2000		2003	
more than 4000km	0.319		0.377		0.317		0.323	
	26.9		*30.7*		*26.8*		*28.0*	
1-4000km		-0.020		0.078		0.231		0.159
		-1.0		*3.9*		*11.4*		*8.0*
4000-7800km		0.551		0.646		0.495		0.520
		38.6		*46.7*		*38.2*		*39.9*
7800-14,000km		0.043		0.124		0.159		0.104
		2.8		*8.1*		*10.7*		*7.0*
more than 14,000km		0.138		0.172		0.203		0.168
		7.1		*9.3*		*11.4*		*9.7*
log Y/L	0.510	0.365	0.508	0.363	0.498	0.395	0.546	0.415
	78.4	*60.2*	*87.1*	*63.5*	*90.3*	*67.2*	*103.8*	*74.4*
landlocked	0.388	0.236	0.556	0.373	0.406	0.300	0.499	0.360
	17.9	*10.7*	*26.1*	*17.4*	*21.0*	*15.3*	24.9	*17.7*
tariff	-0.013	-0.012	-0.010	-0.011	-0.002	-0.002	-0.006	-0.005
	-4.5	*-4.6*	*-7.8*	*-8.9*	*-1.8*	*-2.4*	*-6.7*	*-5.3*
R^2 within	0.113	0.142	0.107	0.137	0.102	0.114	0.120	0.136
N	88,984		108,837		121,830		127,602	
HS codes	11,815		13,131		13,788		14,103	

Notes to Table 7: Estimates of equation 15 in the text. For each year, log U.S. import unit value is regressed on indicators for exporter distance from the US, other controls, and fixed effects for 10 digit HS codes. Estimation is by OLS, *t*-statistics clustered by HS code are in italics. Y/L is aggregate real GDP per worker of the exporter, landlocked is an indicator for the exporter having no port, and tariff is ad valorem percentage. *N* is sample size, "HS codes" is the number of 10 digit HS code fixed effects, and "R^2 within" is the R^2 after removing HS10 means from the data.

Table 8 - Regression of U.S. import value/kilo on distance and other controls, all available observations

	1990		1995		2000		2003	
more than 4000km	0.408		0.520		0.474		0.454	
	35.9		*41.3*		*39.6*		*39.1*	
1-4000km		0.203		0.283		0.315		0.278
		10.7		*16.0*		*17.8*		*15.6*
4000-7800km		0.609		0.741		0.620		0.636
		44.0		*52.4*		*45.4*		*48.0*
7800-14,000km		0.347		0.465		0.465		0.358
		23.1		*31.0*		*31.1*		*23.4*
more than 14,000km		0.304		0.423		0.423		0.354
		16.1		*23.5*		*23.7*		*20.0*
log Y/L	0.432	0.367	0.450	0.381	0.440	0.397	0.505	0.422
	67.7	*58.3*	*77.1*	*63.7*	*82.1*	*67.0*	*97.0*	*73.3*
landlocked	0.436	0.352	0.543	0.444	0.464	0.412	0.554	0.461
	18.1	*14.5*	*23.9*	*19.5*	*22.7*	*20.0*	*27.9*	*73.3*
tariff	-0.011	-0.013	-0.008	-0.010	-0.004	-0.005	-0.005	-0.005
	-8.7	*-10.7*	*-6.5*	*-7.7*	*-3.5*	*-4.6*	*-6.7*	*-6.0*
R^2 within	0.164	0.183	0.167	0.186	0.156	0.162	0.181	0.225
N	52,028		66,366		74,271		78,910	
HS codes	7,422		8,518		8,910		9,139	

Notes to Table 8: Estimates of equation 15 in the text. For each year, log U.S. import value/kilo is regressed on indicators for exporter distance from the US, other controls, and fixed effects for 10 digit HS codes. Estimation is by OLS, *t*-statistics clustered by HS code are in italics. Y/L is aggregate real GDP per worker of the exporter, landlocked is an indicator for the exporter having no port, tariff is ad valorem percentage. *N* is sample size, "HS codes" is the number of 10 digit HS code fixed effects, and "R^2 within" is the R^2 after removing HS10 means from the data.

Table 9- Regression of U.S. import unit value on distance and other controls, restricted sample

	1990		1995		2000		2003	
more than 4000km	0.237		0.279		0.227		0.249	
	17.3		*18.4*		*15.8*		*17.8*	
1-4000km		-0.156		-0.062		0.091		0.023
		-6.6		*-2.77*		*3.8*		*0.9*
4000-7800km		0.499		0.582		0.420		0.462
		29.3		*34.1*		*26.8*		*29.3*
7800-14,000km		-0.123		-0.061		-0.008		-0.072
		-6.8		*-3.32*		*-0.5*		*-4.2*
more than 14,000km		0.006		0.034		0.079		0.029
		0.25		*1.53*		*3.7*		*1.4*
log Y/L	0.548	0.368	0.542	0.362	0.543	0.408	0.582	0.409
	74.0	*53.9*	*81.1*	*56.0*	*84.0*	*58.3*	*93.8*	*62.2*
landlocked	0.372	0.181	0.578	0.345	0.413	0.274	0.515	0.333
	14.2	*6.8*	*21.7*	*12.9*	*17.7*	*11.6*	*20.9*	*13.3*
tariff	-0.014	-0.014	-0.006	-0.007	0.003	0.002	-0.004	-0.002
	-6.0	*-6.4*	*-3.9*	*-4.6*	*2.4*	*1.96*	*-4.2*	*-2.0*
R^2 within	0.123	0.164	0.114	0.157	0.112	0.129	0.128	0.152
N	62,834		76,235		86,364		88,756	
HS codes	7,713		8,513		9,076		9,264	

Notes to Table 9: Estimates of equation 15 in the text. For each year, log U.S. import unit value is regressed on indicators for exporter distance from the US, other controls, and fixed effects for 10 digit HS codes. Estimation is by OLS, *t*-statistics clustered by HS code are in italics. The sample is restricted to SITC 6, 7, and 8, and small observations (imports of one unit and/or with value less than $10,000) are also dropped. Y/L is aggregate real GDP per worker of the exporter, landlocked is an indicator for the exporter having no port, tariff is ad valorem percentage. *N* is sample size, "HS codes" is the number of 10 digit HS code fixed effects, and "R^2 within" is the R^2 after removing HS10 means from the data.

Table 10- Regression of U.S. import value/kilo on distance and other controls, restricted sample

	1990		1995		2000		2003	
more than 4000km	0.340		0.434		0.401		0.412	
	27.6		*30.4*		*28.7*		*32.6*	
1-4000km		0.113		0.156		0.151		0.112
		5.4		*8.3*		*8.1*		*5.7*
4000-7800km		0.597		0.693		0.568		0.621
		37.4		*42.3*		*34.7*		*41.7*
7800-14,000km		0.208		0.300		0.303		0.187
		12.7		*17.5*		*17.3*		*11.4*
more than 14,000km		0.193		0.314		0.291		0.214
		9.1		*14.7*		*13.8*		*10.7*
log Y/L	0.459	0.357	0.475	0.372	0.468	0.385	0.528	0.391
	64.2	*51.1*	*74.4*	*56.7*	*80.1*	*58.4*	*93.9*	*61.8*
landlocked	0.470	0.357	0.557	0.417	0.514	0.426	0.598	0.451
	16.9	*12.3*	*20.7*	*15.6*	*21.6*	*17.7*	*24.1*	*17.9*
tariff	-0.008	-0.012	-0.005	-0.006	0.0003	-0.001	-0.005	-0.004
	-4.5	*-16.7*	*-3.8*	*-4.7*	*0.25*	*-0.9*	*-5.6*	*-4.1*
R^2 within	0.220	0.260	0.215	0.250	0.206	0.220	0.233	0.263
N	31,194		40,039		45,367		47,223	
HS codes	4,046		4,618		4,933		5,080	

Notes to Table 10: Estimates of equation 15 in the text. For each year, log U.S. import value/kilo is regressed on indicators for exporter distance from the US, other controls, and fixed effects for 10 digit HS codes. Estimation is by OLS, *t*-statistics clustered by HS code in italics. The sample is restricted to SITC 6, 7, and 8, and small observations (imports of one unit and/or with value less than $10,000) are also dropped. Y/L is aggregate real GDP per worker of the exporter, landlocked is an indicator for the exporter having no port, tariff is ad valorem percentage. *N* is sample size, "HS codes" is the number of 10 digit HS code fixed effects, and "R^2 within" is the R^2 after removing HS10 means from the data.

Table 11 – Probit of shipment mode choice by non-NAFTA exporters

	1990	1995	2000	2003
log value/kilo	0.182	0.202	0.234	0.221
	10.9	*13.3*	*17.5*	*16.3*
4000-7800km	-0.068	-0.001	0.151	0.078
	-1.4	*0.0*	*1.9*	*1.0*
7800-14,000km	-0.181	-0.156	-0.054	-0.074
	-5.6	*-3.5*	*-0.7*	*-1.1*
more than 14,000km	-0.112	-0.109	-0.053	-0.078
	-5.7	*-3.0*	*-0.7*	*-1.3*
log Y/L	-0.025	-0.012	-0.004	0.003
	-0.9	-0.5	-0.1	0.1
landlocked	0.093	0.113	0.091	0.098
	1.1	*1.3*	*1.2*	*1.1*
sample size	23,149	30,116	33,623	35,955
HS codes	3,541	4,479	4,917	4,854
p-values for instrument validity tests				
exogeneous?	0.75	0.89	0.68	0.84
weak?	0.00	0.00	0.00	0.00

Notes to Table 11 Maximum likelihood estimation of equation (16) in the text for all exporters except Mexico and Canada. Coefficients are marginal effects, evaluated at sample mean. Robust *t*-statistics clustered by exporting country in *italics*. Dependent variable is $a_{ic} = 1$ if product *i* is shipped by air from exporter *c*. Log value/kilo is endogenous, instrument is log of average value/kilo by HS10 from Canada and Mexico. Y/L is aggregate real GDP per worker of the exporter, landlocked is an indicator for the exporter having no port. *p*-values are the results of χ^2 tests for valid instruments. The "exogenous?" row tests the null that the instrument can be excluded from the second stage, while the "weak?" row tests the null that the instrument has no marginal explanatory power in the first stage.

Appendix - Notes on computation of the regression estimates

The data used in this paper have three dimensions: time, product, and exporting country. All regressions are run separately for each year, and a key statistical issue is how to pool the cross-product and cross-country variation within each year when estimating versions of equation (15). Let $i = 1,...,N$ index products and $c = 1,...,C$ index exporting countries.

I can write one of my regression specifications of equation (15) generically as

$$y_{ic} = \mathbf{x}_c'\boldsymbol{\beta}_1 + \mathbf{z}_{ic}'\boldsymbol{\beta}_2 + \mu_i + u_{ic} \qquad i = 1,\ldots,N \quad c = 1,\ldots,C$$

The μ_i are the product fixed effects, $\mathbf{x}_c$ is a vector of exporting country characteristics (such as distance from the US, log GDP per capita, etc) and $\mathbf{z}_{ic}$ is a vector of variables that vary over both products and exporters (such as tariffs). For each country, we can stack N_c observations as follows:

$$\mathbf{y}_c = \mathbf{X}_c\boldsymbol{\beta}_1 + \mathbf{z}_c'\boldsymbol{\beta}_2 + \boldsymbol{\mu}_c + \mathbf{u}_c \qquad c = 1,\ldots,C$$

where $N_c \leq N$ is the number of products exported by country c. The $N_c \times 1$ vector $\boldsymbol{\mu}_c$ is composed of the product fixed effects for products exported by c.

The standard OLS assumption is that $E\left(\mathbf{u}_c\mathbf{u}_c'\right) = \sigma^2 I_c, c = 1,\ldots,C$, but this is *a priori* implausible in this case: we would expect some correlation across products for each country, as well as general heteroskedasticity. A standard approach to this statistical issue is to use a robust covariance matrix which allows for arbitrary cross-commodity correlation within each country, a solution which might be called "clustering by country". The asymptotic theory behind robust covariance matrices with clustering relies on holding the number of observations per cluster fixed while increasing the number of clusters; in my notation, holding N fixed and letting C go to infinity (see, for example Wooldridge (2002), page 328-331). As a consequence, relying on this asymptotic theory in my application is not appropriate, since the number of countries (about 100) is very small relative to the number of products (as many as 14,000).

An alternative approach is to specify a two-way error components model. The generic regression model then becomes

$$y_{ic} = \mathbf{x}_c'\boldsymbol{\beta}_1 + \mathbf{z}_{ic}'\boldsymbol{\beta}_2 + \mu_i + \delta_c + u_{ic} \qquad i = 1,\ldots,N \quad c = 1,\ldots,C$$

Because $\mathbf{x}_c$ has no cross-commodity variation and $\boldsymbol{\beta}_1$ is the main parameter of interest, to identify $\boldsymbol{\beta}_1$ I need to make the random effects assumption that δ_c is a random variable which is orthogonal to $\mathbf{x}_c$ and $\mathbf{z}_{ic}$.

Computation of this mixed fixed effects-random effects model is straightforward: first remove the product means from all variables, and then run generalized least squares on the transformed data. The GLS-RE estimator runs OLS on transformed data. The transformation subtracts θ times the country-specific means from the raw data, where

$$\theta_c = 1 - \sqrt{\frac{\sigma_u^2}{N_c \sigma_\delta^2 + \sigma_u^2}}$$

and σ_δ^2 and σ_u^2 are the variance components. Notice that if σ_δ^2 is very small, then θ is close to zero, and the GLS transformation will leave the data almost unchanged. The estimator that I use for σ_δ^2 is

$$\hat{\sigma}_\delta^2 = \max\left\{0, \frac{1}{\bar{N}_c}\left(\frac{SSR_{between}}{\sum_{c=1}^{C} N_c - K} - \hat{\sigma}_u^2\right)\right\}$$

where *SSR* stands for "sum of squared residuals". Somewhat surprisingly, in my application the result of this estimator is invariably $\hat{\sigma}_\delta^2 = 0$. The intuitive reason is that $SSR_{between}$ is small relative to SSR_{within}. Despite the fact that $\hat{\sigma}_\delta^2 = 0$, the null hypothesis $\sigma_\delta^2 = 0$ can nonetheless be tested using a chi-square statistic. This null hypothesis is invariably *rejected*. Thus, the data analysis gives an odd message: there are random country effects, but they are too small to adjust for.

To summarize, as a consequence of $\hat{\sigma}_\delta^2 = 0$, the results reported in Tables 7 to 10 are computed by OLS with product fixed effects and heteroskedasticity-robust standard errors.

Turning to equation (16), the same statistical issue arises: it is necessary to allow for cross-product correlation within a country, yet clustering by country has weak statistical justification. However, in this case I face an additional complication that an estimator for random effects probit with endogeneity is not available. Since endogeneity is a more pressing concern in this context, and since this model has no product fixed effects to soak up any of the cross-country variation, I report standard errors clustered by country, but these standard errors may be biased.

Appendix reference

Wooldridge, Jeffrey M., 2002, *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.