WHAT GOODS DO COUNTRIES TRADE?
A STRUCTURAL RICARDIAN MODEL

ARNAUD COSTINOT AND IVANA KOMUNJER

ABSTRACT. The Ricardian model predicts that countries should produce and export relatively more in industries in which they are relatively more productive. Though one of the most celebrated insights in the theory of international trade, this prediction has received virtually no attention in the empirical literature since the mid-sixties. The main reason behind this lack of popularity is the absence of clear theoretical foundations to guide the empirical analysis. Building on the seminal work of Eaton and Kortum (2002), the present paper offers such foundations. Compared to the existing literature, our structural approach presents several advantages: (*i*) we do not have to rely on ad-hoc measures of export performance or bilateral comparisons inspired by a two-country model; (*ii*) we can discuss the origin of the error term, and therefore, the plausibility of our orthogonality conditions; and (*iii*) we can control for measurement errors in productivity using import penetration ratios.

1. INTRODUCTION

The Ricardian model predicts that countries should produce and export relatively more in industries in which they are relatively more productive. Though one of the most celebrated insights in the theory of international trade, this prediction has received virtually no attention in the empirical literature since the mid-sixties; see e.g. MacDougall (1951), Stern (1962), Balassa (1963).[1] The main reason behind this lack of popularity is not the existence of strong beliefs regarding the (un)importance of technological considerations. It derives instead from the obvious mismatch between the real world and the extreme assumptions of the Ricardian model. In the words of Leamer and Levinsohn (1995), "[it] is just too simple."

A seminal contribution of Eaton and Kortum (2002) is to demonstrate that random productivity shocks are sufficient to transform the Ricardian model into an empirically useful tool for the analysis of *trade volumes*. When drawn from an extreme value distribution, these shocks imply a gravity-like equation in a Ricardian framework with a continuum of goods, transport costs, and more than two countries. The objective of our paper is to generalize Eaton and Kortum's (2002) approach and develop a structural Ricardian model within which it is possible to estimate the impact of productivity differences on *trade patterns*.

Section 2 describes our economic model. We consider an economy with one factor of production, labor, and multiple goods, each available in many varieties. There are constant returns to scale in the production of each variety. The key feature of our model is that labor productivity may be separated into: a deterministic component, which is country and industry specific; and a stochastic component, randomly drawn across countries, industries, and varieties. The former, which we refer to as "fundamental productivity," captures factors such as climate, infrastructure, and institutions that affect the productivity of all producers in a given country and industry.[2] The latter, by contrast, reflects idiosyncratic differences in technological know-how across varieties. Unlike Eaton and Kortum (2002), we allow

---

[1]A notable exception is Golub and Hsieh (2000).

[2]Acemoglu, Antras, and Helpman (2006), Costinot (2005a), Cuñat and Melitz (2006), Levchenko (2004), Matsuyama (2005), Nunn (2005), and Vogel (2004) explicitly model the impact of various institutional features—e.g. labor market flexibility, the quality of contract enforcement, or credit market imperfections—on labor productivity across countries and industries.

fundamental productivity levels to vary across industries and remain agnostic about the distribution of productivity shocks.

Section 3 derives the restrictions that our economic model imposes on the pattern of trade and contrasts them with those of the standard Ricardian model.[3] Because of random productivity shocks, we can no longer predict trade flows in each variety. Yet, by assuming that each good comes in a large number of varieties, we can generate sharp predictions at the industry level. In particular, we can show that, for any pair of exporters, the ranking by industry of the ratios of their fundamental productivity levels determines the ranking of the ratios of their exports towards any importing country.

Our approach mirrors Deardorff (1980) who shows how the law of comparative advantage may remain valid, under standard assumptions, when stated in terms of correlations between vectors of trade and autarky prices. In this paper, we weaken the standard Ricardian assumptions—the "chain of comparative advantage" will only hold in terms of first-order stochastic dominance—and derive a deterministic relationship between exports and labor productivity across industries. Another perspective of this result is that the ranking of relative exports fully "reveals" comparative advantage. By observing exports across countries and industries, we can infer—according to our model—the ranking of relative productivity levels. Hence, our model also provides simple theoretical foundations to measures of revealed comparative advantage à la Balassa (1965).[4]

Section 4 uses our theoretical restrictions to develop an econometric model. According to the results of Section 3, the log of bilateral exports can be decomposed into: (*i*) an importer-exporter specific term, which captures differences in wages and bilateral trade barriers such as physical distance; (*ii*) an importer-industry specific term, which captures differences in demand and trade policy barriers; and (*iii*) an exporter-industry specific term, which isolates the only source of comparative advantage in our model, productivity differences. We augment this three-term decomposition with an error term, which aims to capture either unobserved trade barriers or measurement errors in trade flows. The error term is assumed to be independent across countries and industries and uncorrelated with fundamental productivity differences.

---

[3]We refer to a "standard" Ricardian model as a Ricardian model without random productivity shocks.

[4]We thank Kei-Mu Yi for suggesting this application.

A key benefit of our structural approach is that it allows us to control explicitly for the difference between "true" and "observed" productivity. In our model, fundamental productivity levels reflect averages across the entire universe of varieties. In practice, however, averages are computed across the varieties that are actually produced. This discrepancy between true and observed productivity may bias the estimates of export elasticities with respect to productivity. If better productivity draws are observed in less productive countries, then observed differences will be smaller than true differences in productivity, which may artificially raise the elasticity of exports. Using our theoretical foundations, we show that data on import penetration ratios are sufficient to control for this type of measurement error.

Section 5 presents our estimates of the impact of productivity differences on the pattern of trade. Using OECD data, we find that, as suggested by David Ricardo long ago, countries do tend to export relatively more—towards any importing country—in industries where they are relatively more productive. Ceteris paribus, a one standard deviation decrease in the log of the unit labor requirement raises the log of exports by 0.26 standard deviations.

We wish to be clear that the contribution of our paper is not to offer a new stylized fact on the determinants of international trade. Previous "tests" of the Ricardian model— MacDougall (1951), Stern (1962), Balassa (1963), and more recently, Golub and Hsieh (2000)—were remarkably successful. So, one should not be surprised to uncover a positive relationship between productivity and trade flows in the data. The contribution of our paper is to offer clear theoretical foundations for cross-sectional regressions that have long been perceived as hopelessly ad-hoc; see Bhagwati (1964).

Though the ad-hoc nature of Ricardian regressions has not lead to the disappearance of technological considerations from the empirical literature, it has had a strong influence on *how* the relationship between technology and trade patterns has been studied. In the Heckscher-Ohlin-Vanek literature—with or without technological differences—the factor content of trade remains the main variable of interest; see e.g. Bowen, Leamer, and Sveikauskas

(1987), Trefler (1993), Trefler (1995), and Davis and Weinstein (2001).[5] The structural Ricardian model developed in this paper puts back productivity differences at the forefront of the analysis of trade patterns across countries and industries.[6]

The tight connection between the theory and empirics that our paper offers presents several advantage over the previous literature. First, we do not have to rely on ad-hoc measures of export performance such as total exports towards the rest of the world (MacDougall, 1951; Stern, 1962); total exports to third markets (Balassa, 1963); or bilateral net exports (Golub and Hsieh, 2000). The theory tells us exactly what the dependent variable in the cross-industry regressions ought to be: log of exports, disaggregated by exporting and importing countries. Second, the careful introduction of country and industry fixed effects allows us to move away from the bilateral comparisons inspired by the two-country model, and in turn, to take advantage of a much richer data set. Third, our microtheoretical foundations make it possible to discuss the economic origins of the error term and, as a result, the plausibility of our orthogonality conditions. Finally, our economic model allows us to control explicitly for the measurement error in productivity associated with the non-random specialization of countries into different varieties.

At this point, the skeptical reader may still ask: do we really need random productivity shocks à la Eaton and Kortum (2002) to bring the Ricardian predictions to the data? After all, there are many ways to derive gravity-like equations such as the one offered in our paper; why not follow the voluminous gravity literature and impose Armington's preferences instead?[7] Our view is that if the main issue associated with Ricardian regressions is that they are ad-hoc, then invoking unjustified and arbitrary assumptions on the structure of

---

[5]Of course, technological considerations have also played an important role in the empirical literature on plant level heterogeneity and exports; see e.g. Tybout and Roberts (1997), Bernard and Jensen (1999), and Bernard, Eaton, Jensen, and Kortum (2003). However, these plant-level studies have nothing to say about trade patterns across countries and industries, which is the main focus of our analysis.

[6]Note that our model could easily accomodate multiple factors of production. The basic idea is to reinterpret differences in labor productivity as differences in total factor productivity. With multiple factors of production, the value of exports would be a function of both technological differences and differences in relative factor prices. The rest of our analysis would remain unchanged; see Appendix E.

[7]Deardorff (2004) analyzes the impact of production and trade costs on the net direction of countries' bilateral trade with a model developed along these lines. Alternative theoretical foundations for cross-industry regressions can be found in Petri (1980), Harrigan (1997), Romalis (2004), and Morrow (2006). Compared

preferences is unlikely to make them more popular.[8] An attractive feature of our approach is the weakness of the assumptions under which the gravity equation is derived. Moreover, our analysis demonstrates that the approach used to derive a gravity equation has in itself empirical content. If there are random productivity shocks within each industry, then import penetration ratios should have an impact on observed measures of productivity, and in turn, on the pattern of trade.

2. A RICARDIAN MODEL WITH RANDOM PRODUCTIVITY SHOCKS

We consider a world economy comprising $i = 1, ..., I$ countries and one factor of production, labor. There are $k = 1, ..., K$ goods and constant returns to scale in the production of each good. Labor is perfectly mobile across industries and immobile across countries. The wage of workers in country i is denoted w_i. Up to this point, this is a standard Ricardian model. We generalize this model by introducing random productivity shocks. Following Eaton and Kortum (2002), we assume that each good k may come in N^k varieties $\omega = 1, ..., N^k$, and denote $a_i^k(\omega)$ the constant unit labor requirements for the production of the ωth variety of good k in country i. Our first assumption is that:

A1. *For all countries i, goods k, and their varieties ω*

$$\ln a_i^k(\omega) = \ln a_i^k + u_i^k(\omega),$$

(1)

where $a_i^k > 0$ and $u_i^k(\omega)$ is a random variable drawn independently for each triplet (i, k, ω) from a continuous distribution $F(\cdot)$ such that: $E[u_i^k(\omega)] = 0$.

We interpret a_i^k as a measure of the fundamental productivity of country i in industry k and $u_i^k(\omega)$ as a random productivity shock. The former, which can be estimated using aggregate data, captures cross-country and cross-industry heterogeneity. It reflects factors such as climate, infrastructure, and institutions that affect the productivity of *all* producers in a given country and industry. Random productivity shocks, on the other hand, capture intra-industry heterogeneity. They reflect idiosyncratic differences in technological know-how across varieties, which are assumed to be drawn independently from a *unique* distribution

---

to our paper, Petri (1980) imposes stronger restrictions on the structure of preferences, whereas Harrigan (1997), Romalis (2004), and Morrow (2006) impose stronger restrictions on the structure of transport costs.

[8]See e.g. the discussion of Petri (1980) in Leamer and Levinsohn (1995).

$F(\cdot)$. In our setup, cross-country and cross-industry variations in the distribution of productivity levels derive from variations in a single parameter: a_i^k.

Assumption A1 generalizes Eaton and Kortum's (2002) approach along two dimensions. First, it introduces the existence of exogenous productivity differences across industries. This will allow us to shift the indeterminacy in trade in individual goods to indeterminacy in trade in varieties. Second, it does not impose any restriction on the distribution of random productivity shocks.

We assume that trade barriers take the form of "iceberg" transport costs:

A2(i). *For every unit of good k shipped from country i to country j, only $1/d_{ij}^k$ units arrive, where:*

(2)
$$\begin{cases} d_{ij}^k = d_{ij} \cdot d_j^k \geq 1, & \text{if } i \neq j, \\ d_{ij}^k = 1, & \text{otherwise.} \end{cases}$$

The indices i and j refer to the exporting and importing countries, respectively. The first parameter d_{ij} measures the trade barriers which are specific to countries i and j. It includes factors such as: physical distance, existence of colonial ties, use of a common language, or participation in a monetary union. The second parameter d_j^k measures the policy barriers imposed by country j on good k, such as import tariffs and standards. In line with "the most-favored-nation" clause of the World Trade Organization, these impediments may not vary by country of origin.

In addition, we assume that transport costs are such that there are no cross-country arbitrage opportunities:

A2(ii). *For any good k and any three countries, i, j, and l, $d_{il}^k \leq d_{ij}^k \cdot d_{jl}^k$.*

We assume that markets are perfectly competitive.[9] Together with constant returns to scale in production, perfect competition implies:

A3. *In any country j, the price $p_j^k(\omega)$ paid by buyers of variety ω of good k is*

(3)
$$p_j^k(\omega) = \min_{1 \leq i \leq I} \left[c_{ij}^k(\omega) \right],$$

---

[9]The case of Bertrand competition is discussed in detail in Appendix B.

where $c_{ij}^k(\omega) = d_{ij}^k \cdot w_i \cdot a_i^k(\omega)$ is the cost of producing and delivering one unit of this variety from country i to country j.

For each variety ω of good k, buyers in country j are "shopping around the world" for the best price available. Here, random productivity shocks lead to random costs of production $c_{ij}^k(\omega)$ and in turn, to random prices $p_j^k(\omega)$. In what follows, we let $c_{ij}^k = d_{ij}^k \cdot w_i \cdot a_i^k > 0$.

On the demand side, we assume that consumers have a two-level utility function with CES preferences across varieties. This implies:

A4(i). *In any country j, the total spending on variety ω of good k is*

$$(4) \qquad\qquad x_j^k(\omega) = \left[p_j^k(\omega)/p_j^k \right]^{1-\sigma} e_j^k,$$

where $e_j^k > 0$, $\sigma > 1$ and $p_j^k = [\sum_{\omega'=1}^{N^k} p_j^k(\omega')^{1-\sigma}]^{1/(1-\sigma)}$.

The above expenditure function is a standard feature of the "new trade" literature; see e.g. Helpman and Krugman (1985). e_j^k is an endogenous variable that represents total spending on good k in country j. It depends on the upper tier utility function in this country and the equilibrium prices. p_j^k is the CES price index, and σ is the elasticity of substitution between varieties. It is worth emphasizing that while the elasticity of substitution σ is assumed to be constant, total spending, and hence demand conditions, may vary across countries and industries: e_j^k is a function of j and k.

Finally, we assume that:

A4(ii). *In any country j, the elasticity of substitution σ between two varieties of good k is such that $E\left[p_j^k(\omega)^{1-\sigma} \right] < \infty$.*

Assumption A4(ii) is a technical assumption that guarantees the existence of a well defined price index. Whether or not A4(ii) is satisfied ultimately depends on the shape of the distribution $F(\cdot)$.[10]

---

[10]Suppose, for example, that $u_i^k(\omega)$'s are drawn from a (negative) exponential distribution with mean zero: $F(u) = \exp[\theta u - 1]$ for $-\infty < u \leq 1/\theta$ and $\theta > 0$. This corresponds to the case where labor productivity $z_i^k(\omega) \equiv 1/a_i^k(\omega)$ is drawn from a Pareto distribution: $G_i^k(z) = 1 - (b_i^k/z)^\theta$ for $0 < b_i^k \leq z$ and $b_i^k \equiv (1/a_i^k)\exp(-1/\theta)$, as assumed in various applications and extensions of Melitz's (2003) model; see e.g. Helpman, Melitz, and Yeaple (2004), Antras and Helpman (2004), Ghironi and Melitz (2005), Bernard, Redding, and Schott (2006), and Chaney (2007). Then, our assumption A4(ii) holds if the elasticity of substitution $\sigma < 1 + \theta$. Alternatively, suppose that $u_i^k(\omega)$'s are distributed as a (negative) Gumbel random

In the rest of the paper, we let $x_{ij}^k = \sum_{\omega=1}^{N^k} x_{ij}^k(\omega)$ denote the value of exports from country i to country j in industry k, where total spending on each variety $x_{ij}^k(\omega)$ is given by:

(5)
$$\begin{cases} x_{ij}^k(\omega) = x_j^k(\omega), & if\ c_{ij}^k(\omega) = \min_{1 \le i' \le I} c_{i'j}^k(\omega), \\ x_{ij}^k(\omega) = 0, & otherwise. \end{cases}$$

3. Theoretical Results

We now derive the restrictions that Assumptions A1–A4 impose on the pattern of trade and contrast them with those of the standard Ricardian model.

3.1. Pattern of Trade with Gumbel Distributions.

For expositional purposes, we first derive predictions on the pattern of trade when the distribution of a random productivity shocks, $F(\cdot)$, is a Gumbel with mean zero, as assumed in Eaton and Kortum (2002). Hence, the only difference between the present model and theirs is the existence of multiple industries. This corresponds to the case where

(6)
$$F(u) = 1 - \exp[-\exp(\theta u - \mathsf{e})]$$

with $u \in \mathbb{R}$, $\theta > 0$, and e the Euler's constant $\mathsf{e} \simeq 0.577$.

Our first result can be stated as follows:

Theorem 1. *Suppose that Assumptions A1-A4 hold. In addition, assume that the number of varieties N^k of any good k is large, and that $F(\cdot)$ satisfies Equation (6). Then, for any exporter i, any importer j, and any good k,*

(7)
$$\ln x_{ij}^k \simeq \theta_{ij} + \theta_j^k - \theta \ln a_i^k.$$

The proof of Theorem 1 mainly is a matter of algebra. First, we relate total exports x_{ij}^k to the expected value of exports coming from country i, using the law of large numbers.[11]

---

variable with mean zero: $F(u) = 1 - \exp[-\exp(\theta u - \mathsf{e})]$ for $u \in \mathbb{R}$ and $\theta > 0$, where e is Euler's constant $\mathsf{e} \simeq 0.577$. This corresponds to the case where labor productivity $z_i^k(\omega)$ is drawn from a Fréchet distribution: $G_i^k(z) = \exp(-b_i^k z^{-\theta})$ for $z \ge 0$ and $b_i^k \equiv (1/a_i^k)^\theta \exp(-\mathsf{e})$, as assumed, for example, in Eaton and Kortum (2002) and Bernard, Eaton, Jensen, and Kortum (2003). Then, like in the Pareto case, A4(ii) holds if $\sigma < 1 + \theta$.

[11]Alternatively, we could have assumed the existence a continuum of varieties and argued, like many before us, that total exports were equal to their expected value. By assuming that the number of varieties is large but finite, we avoid, however, the technical difficulties of invoking the law of large numbers with

Second, we compute the expected value explicitly using Equation (6). The proof of Theorem 1 and all subsequent theorems can be found in Appendix A.

The first term $\theta_{ij} \equiv -\theta \ln(d_{ij} \cdot w_i)$ is importer and exporter specific; it reflects wages w_i in the exporting country and trade barriers d_{ij} between countries i and j. The second term $\theta_j^k \equiv \ln e_j^k - \theta \ln d_j^k - \ln \left(\sum_{i'=1}^{I} (c_{i'j}^k)^{-\theta} \right)$ is importer and industry specific; it reflects the policy barriers d_j^k imposed by country j on good k and demand differences e_j^k across countries and industries. The main insight of Theorem 1 comes from the third term $\theta \ln a_i^k$. Since $\theta > 0$, $\ln x_{ij}^k$ should be decreasing in $\ln a_i^k$: *ceteris paribus*, countries should export less in industries where their firms are, on average, less efficient.

It is worth emphasizing that Theorem 1 *cannot* be used for comparative static analysis. If the fundamental productivity level goes up in a given country and industry, this will affect wages, demand, and, in turn, exports in other countries and industries through general equilibrium effects.[12] In other words, changes in a_i^k also lead to changes in the country and industry fixed effects, θ_{ij} and θ_j^k. By contrast, Theorem 1 *can* be used to analyze the cross-sectional variations of bilateral exports, as we shall further explore in Section 5.

Finally, note that the two fixed-effects, θ_{ij} and θ_j^k, do not depend on the elasticity of substitution σ. Thus, the predictions of Theorem 1 still hold if we relax Assumption A4(i), so that the elasticity of substitution may vary across countries and industries, $\sigma \equiv \sigma_j^k$. We shall revisit this intriguing result in the next section.

3.2. **Pattern of Trade with Unknown Distributions.** We now relax the assumption that $F(\cdot)$ is a Gumbel distribution. In this situation, we can no longer obtain a closed form solution, but we can still derive a log-linear relationship between total exports and the fundamental productivity level a_i^k, using a first-order Taylor series development around a symmetric situation where costs are identical across exporters, $\left(c_{1j}^k = ... = c_{Ij}^k \right)$.

Theorem 2. *Suppose that Assumptions A1-A4 hold. In addition, assume that the number of varieties N^k of any good k is large, and that cost differences across exporters are small:*

_____

a continuum of i.i.d. variables; see e.g. Al-Najjar (2004). Nothing substantial hinges on this particular modeling choice.

[12]To do comparative statics, one would therefore need to use labor market clearing conditions and compute equilibrium wages across countries.

$c_{1j}^k \simeq ... \simeq c_{Ij}^k$. *Then, for any exporter i, any importer $j \neq i$, and any good k,*

$$(8) \qquad\qquad\qquad \ln x_{ij}^k \simeq \gamma_{ij} + \gamma_j^k - \gamma \ln a_i^k,$$

where $\gamma > 0$.

The exact expressions for γ_{ij}, γ_j^k, and γ can be found in Appendix A. Theorem 2 predicts that, like in the Gumbel case, total exports can be decomposed into an importer-exporter specific term, γ_{ij}; an importer-industry specific term, γ_j^k; and a third term, $\gamma \ln a_i^k$, which captures the impact of productivity differences. Since $\gamma > 0$, Theorem 2 also predicts that: *ceteris paribus*, countries should export less in industries where their firms are, on average, less efficient.

The predictive power of Theorem 2 crucially relies on the fact that γ is constant across countries and industries. To understand this result, it is convenient to think about total exports in terms of their extensive and intensive margins, that is how many and how much of each variety are being exported, respectively. The unique distribution of random productivity shocks $F(\cdot)$ makes sure that marginal changes in the costs of production c_{ij}^k have the same impact on the extensive margin across countries and industries. Similarly, the constant elasticity of substitution σ guarantees that they have the same impact on the intensive margin. This is the basic idea behind Theorem 2. The other assumptions simply allow us to identify the effect of labor productivity by bundling the impact of changes in wages, demand, and transport costs into fixed effects.

Relaxing Eaton and Kortum's (2002) distributional assumption in Theorem 2 comes at the cost of two new restrictions. First, we must move from global predictions—which hold for any $\left(c_{1j}^k, ..., c_{Ij}^k\right)$—to local predictions—which only hold if costs differences across all exporters are small.[13] Second, we must assume that the elasticity of substitution is constant across countries and industries. This assumption was not necessary in Theorem 1 because of

_____

[13]Although this requirement may seem unreasonably strong, the predictions of Theorem 2 hold more generally if, for each industry and each importing country, exporters can be separated into two groups: small exporters, whose costs are very large (formally, close to infinity), and large exporters, whose costs of production are small and of similar magnitude. Then, small exporters export with probability close to zero and the results of Theorem 2 apply to the group of large exporters. In other words, Theorem 2 does not require Gambia and Japan to have similar costs of producing and delivering cars in the United States. It simply requires that Japan and Germany do.

one key property of the Gumbel distribution: the distribution of the price $p_j^k(\omega)$ of a given variety ω of good k in country j is independent of the country of origin i; see Eaton and Kortum (2002) p1748 for a detailed discussion. Formally, if $F(\cdot)$ satisfies Equation (6), then

$$(9) \qquad \Pr\left\{p_j^k(\omega) \le p\right\} = \Pr\left\{p_j^k(\omega) \le p \,\middle|\, c_{ij}^k(\omega) = \min_{1 \le i' \le I} c_{i'j}^k(\omega)\right\},$$

for any $p > 0$ and any $1 \le i \le I$. Unfortunately, this property does not generalize to other standard distributions, as we show in the following theorem.

Theorem 3. *Suppose that Assumptions A1-A4 hold and that $f(u) \equiv F'(u) > 0$ for any u in $\mathbb{R}$. Then, for any $p > 0$ and any $1 \le i \le I$, we have:*

$$\Pr\left\{p_j^k(\omega) \le p\right\} = \Pr\left\{p_j^k(\omega) \le p \,\middle|\, c_{ij}^k(\omega) = \min_{1 \le i' \le I} c_{i'j}^k(\omega)\right\} \Leftrightarrow F(\cdot) \text{ satisfies Equation (6)}.$$

Put simply, the *only* distribution with full support satisfying Property (9) is the Gumbel. For any other distribution, the intensive margin, and hence elasticities of substitution, matter.

3.3. **Relation to the Standard Ricardian Model.** In order to compare our results to those of the standard Ricardian model, we now offer a Corollary to Theorems 1 and 2. Consider an arbitrary pair of exporters, i_1 and i_2, an importer $j \ne i_1, i_2$, and an arbitrary pair of goods, k_1 and k_2. Taking the differences-in-differences in Equation (7) we get

$$\left(\ln x_{i_1 j}^{k_1} - \ln x_{i_1 j}^{k_2}\right) - \left(\ln x_{i_2 j}^{k_1} - \ln x_{i_2 j}^{k_2}\right) \simeq -\theta\left[\left(\ln a_{i_1}^{k_1} - \ln a_{i_1}^{k_2}\right) - \left(\ln a_{i_2}^{k_1} - \ln a_{i_2}^{k_2}\right)\right].$$

A similar manipulation of Equation (8) implies

$$\left(\ln x_{i_1 j}^{k_1} - \ln x_{i_1 j}^{k_2}\right) - \left(\ln x_{i_2 j}^{k_1} - \ln x_{i_2 j}^{k_2}\right) \simeq -\gamma\left[\left(\ln a_{i_1}^{k_1} - \ln a_{i_1}^{k_2}\right) - \left(\ln a_{i_2}^{k_1} - \ln a_{i_2}^{k_2}\right)\right].$$

Since $\theta > 0$ and $\gamma > 0$, we obtain

$$(10) \qquad \frac{a_{i_1}^{k_1}}{a_{i_2}^{k_1}} > \frac{a_{i_1}^{k_2}}{a_{i_2}^{k_2}} \Leftrightarrow \frac{x_{i_1 j}^{k_1}}{x_{i_2 j}^{k_1}} < \frac{x_{i_1 j}^{k_2}}{x_{i_2 j}^{k_2}},$$

under the assumptions of Theorems 1 or 2. Still considering the pair of exporters i_1 and i_2 and generalizing the above reasoning to all K goods, we derive the following Corollary:

Corollary 4. *Suppose that the assumptions of Theorem 1 or 2 hold. Then, the ranking of relative unit labor requirements determines the ranking of relative exports:*

$$\left\{ \frac{a_{i_1}^1}{a_{i_2}^1} > .... > \frac{a_{i_1}^k}{a_{i_2}^k} > ... > \frac{a_{i_1}^K}{a_{i_2}^K} \right\} \Leftrightarrow \left\{ \frac{x_{i_1 j}^1}{x_{i_2 j}^1} < .... < \frac{x_{i_1 j}^k}{x_{i_2 j}^k} < ... < \frac{x_{i_1 j}^K}{x_{i_2 j}^K} \right\}.$$

Note that, without loss of generality, we can always index the K goods so that:

$$(11) \qquad \qquad \frac{a_{i_1}^1}{a_{i_2}^1} > .... > \frac{a_{i_1}^k}{a_{i_2}^k} > ... > \frac{a_{i_1}^K}{a_{i_2}^K}.$$

Ranking (11) is at the heart of the standard Ricardian model; see e.g. Dornbusch, Fischer, and Samuelson (1977). When there are no random productivity shocks, Ranking (11) merely states that country i_1 has a comparative advantage in (all varieties of) the high k goods. If there only are two countries, the pattern of trade follows: i_1 produces and exports the high k goods, while i_2 produces and exports the low k goods. If there are more than two countries, however, the pattern of pairwise comparative advantage no longer determines the pattern of trade. In this case, the standard Ricardian model loses most of its intuitive content; see e.g. Jones (1961) and Wilson (1980).

When there are stochastic productivity differences within each industry, Assumption A1 and Ranking (11) further imply:

$$(12) \qquad \qquad \frac{a_{i_1}^1(\omega)}{a_{i_2}^1(\omega)} \succ .... \succ \frac{a_{i_1}^k(\omega)}{a_{i_2}^k(\omega)} \succ ... \succ \frac{a_{i_1}^K(\omega)}{a_{i_2}^K(\omega)},$$

where $\succ$ denotes the first-order stochastic dominance order among distributions.[14] In other words, Ranking (12) is just a stochastic—hence weaker—version of the ordering of labor requirements a_i^k, which is at the heart of the Ricardian theory. Like its deterministic counterpart in (11), Ranking (12) captures the idea that country i_1 is relatively better at producing the high k goods. But whatever k is, country i_2 may still have lower labor requirements on some of its varieties.

---

[14]To see this, note that for any $A \in \mathbb{R}^+$ we have $\Pr\left\{ a_{i_1}^k(\omega)/a_{i_2}^k(\omega) \le A \right\} = \Pr\{u_{i_1}^k(\omega) - u_{i_2}^k(\omega) \le \ln A - \ln a_{i_1}^k + \ln a_{i_2}^k\}$. Since for any $k < k'$, $u_{i_1}^k(\omega) - u_{i_2}^k(\omega)$ and $u_{i_1}^{k'}(\omega) - u_{i_2}^{k'}(\omega)$ are drawn from the same distribution by A1, Ranking (11) implies:

$$\Pr\left\{ \frac{a_{i_1}^k(\omega)}{a_{i_2}^k(\omega)} \le A \right\} < \Pr\left\{ \frac{a_{i_1}^{k'}(\omega)}{a_{i_2}^{k'}(\omega)} \le A \right\} \Leftrightarrow \frac{a_{i_1}^k(\omega)}{a_{i_2}^k(\omega)} \succ \frac{a_{i_1}^{k'}(\omega)}{a_{i_2}^{k'}(\omega)}.$$

According to Corollary 4, Ranking (12) does not imply that country i_1 should only produce and export the high k goods, but instead that it should produce and export relatively more of these goods. This is true irrespective of the number of countries in the economy. Unlike the standard Ricardian model, our stochastic theory of comparative advantage generates a clear and intuitive correspondence between labor productivity and exports. In our model, the pattern of comparative advantage for any pair of exporters fully determines their relative export performance across industries.

Another perspective on Corollary 4 is that, for any pair of exporters, the ranking of their relative exports towards any importing country fully *reveals* their comparative advantage. By observing exports across countries and industries, one can directly infer—according to our model—the ranking of relative productivity levels. Thus, our results also provide theoretical foundations to measures of revealed comparative advantage à la Balassa (1965). We explore that idea in details in Appendix C.

The previous discussion may seem paradoxical. As we have just mentioned, Ranking (12) is a weaker version of the ordering at the heart of the standard theory. If so, how does our stochastic theory lead to finer predictions? The answer is simple: it does not. While the standard Ricardian model is concerned with trade flows in each variety of each good, we only are concerned with the total trade flows in each good. Unlike the standard model, we recognize that random shocks—whose origins remain outside the scope of our model— may affect the costs of production of any variety.[15] Yet, by assuming that these shocks are identically distributed across a large number of varieties, we manage to generate sharp predictions at the industry level.

4. ECONOMETRIC MODEL

The Ricardian model with random productivity shocks presented in Section 2 has clear implications for the pattern of trade in an economy with many countries, many goods, and transport costs. According to Theorems 1 and 2, the log of bilateral exports, $\ln x_{ij}^k$, can be decomposed into three terms: (i) θ_{ij} or γ_{ij}, which is importer-exporter specific; (ii) θ_j^k or γ_j^k, which is importer-industry specific; and (iii) $\theta \ln a_i^k$ or $\gamma \ln a_i^k$, which is exporter-industry

---

[15]In this regard, our model bears a ressemblance to Davis (1995) where Ricardian differences also explains intra-industry. However, unlike in our model, factor proportions determine inter-industry trade in Davis (1995).

specific and reflects fundamental productivity differences. We now develop an econometric model based on this theoretical decomposition.

4.1. **Error Term.** Theorems 1 and 2 are a useful theoretical benchmark that captures in a simple way the logic of comparative advantage. As shown in Corollary 4, they imply that for any pair of exporters, the ranking by industry of relative exports should be constant across importers. Formally, for all exporters, i_1 and i_2, goods, k_1 and k_2, and importers, j_1 and j_2, our theoretical model predicts that

$$(13) \qquad \left(\frac{x_{i_1 j_1}^{k_1}}{x_{i_2 j_1}^{k_1}} - \frac{x_{i_1 j_1}^{k_2}}{x_{i_2 j_1}^{k_2}} \right) \cdot \left(\frac{x_{i_1 j_2}^{k_1}}{x_{i_2 j_2}^{k_1}} - \frac{x_{i_1 j_2}^{k_2}}{x_{i_2 j_2}^{k_2}} \right) > 0.$$

To fix ideas, consider two exporters, the United States and Germany, and two goods, aircrafts and cars. According to Property (13), if Germany exports relatively more cars towards France than the United States, then it should also export relatively more cars towards Mexico. The *absolute* levels of German and US exports may vary between France and Mexico due to changes in demand and transport costs, but the *relative* export performance of Germany and the United States in these two industries may not.

In practice, however, there are two important reasons why Property (13) may not hold.[16] First, there exist trade barriers that violate Assumption A2. These violations may arise because of custom unions or, more simply, because bilateral distance has a differential impact on goods of different weights; see e.g. Harrigan (2005). Second, trade data are notoriously plagued with measurement errors; see e.g. Anderson and Wincoop (2004). With this in mind, we propose a linear regression model that satisfies the three-term decomposition offered by Theorems 1 and 2, but incorporates explicitly the existence of transport costs not accounted by Assumption A2 and/or measurement error in trade flows:

$$(14) \qquad \ln x_{ij}^k = \beta_{ij} + \beta_j^k + \beta \ln a_i^k + \varepsilon_{ij}^k,$$

where β_{ij} and β_j^k are importer–exporter and importer–industry fixed effects, respectively, and ε_{ij}^k is an error term.[17] Whether we interpret ε_{ij}^k as measurement error in trade flows or

---

[16]For example, in our data, we find that 69% of the $17,955$ groups of exporters, importers, and industries satisfy Property (13) in 1997.

[17]It is worth pointing out that in our model, β is identified out of the difference-in-difference in $\ln a_i^k$. Like in a standard Ricardian model, we only use comparisons of *relative* productivities across sectors, and never *absolute* productivities.

unobserved transport costs, we shall assume that ε_{ij}^k is independent across countries i and j as well as across industries k; that ε_{ij}^k is mean zero and possibly heteroskedastic conditional on i, j and k; and that ε_{ij}^k is uncorrelated with $\ln a_i^k$.[18]

4.2. **True versus Observed Productivity.** Our econometric model has one explanatory variable, $\ln a_i^k$. By Assumption A1, this variable satisfies

$$\ln a_i^k = \ln E\left[a_i^k\left(\omega\right)\right] - \ln E\left[\exp(u)\right],$$

where the expectation is taken across all varieties of good k. In practice, however, the econometrician does not observe the entire universe of varieties. Instead, she only has access to the varieties that are actually produced in country i, and thus may only observe

$$\ln \widehat{a}_i^k = \ln E\left[a_i^k\left(\omega\right)\big|\omega \in \Omega_i^k\right],$$

where $\Omega_i^k \equiv \left\{\omega \in [\![1, N_k]\!]\,\big|\exists j \in [\![1, I]\!] : c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}$. In the rest of this paper, we refer to this measurement error in productivity, $\Delta_i^k \equiv \ln E\left[a_i^k\left(\omega\right)\right] - \ln E\left[a_i^k\left(\omega\right)\big|\omega \in \Omega_i^k\right]$, as "selection error." When this error is taken into account, our estimating equation becomes[19]

(15) $$\ln x_{ij}^k = \beta_{ij} + \beta_j^k + \beta\left(\ln \widehat{a}_i^k + \Delta_i^k\right) + \varepsilon_{ij}^k.$$

Omitting Δ_i^k from Equation (15) may bias downward the OLS estimate of β. If better productivity draws are observed when a_i^k is high, then differences in observed productivity

---

[18]Among other things, this orthogonality condition rules out situations where country j tends to discriminate more against country i in industries where i is more productive. Were these situations prevalent in practice, due to endogenous trade protection, our OLS estimates of β would be biased towards zero. To see this, suppose that trade barriers, d_{ij}^k, and exports, x_{ij}^k, are simultaneously determined according to

$$\begin{cases} \ln d_{ij}^k = \ln d_{ij} + \ln d_j^k + \mu \ln x_{ij}^k \\ \ln x_{ij}^k = \widetilde{\beta}_{ij} + \widetilde{\beta}_j^k + \beta \ln a_i^k + \beta \ln d_{ij}^k \end{cases}$$

where $\mu > 0$ captures the fact that higher levels of import penetration lead to higher levels of protection. The previous system can be rearranged as

$$\begin{cases} \ln d_{ij}^k = (1 - \mu\beta)^{-1}[\ln d_{ij} + \ln d_j^k + \mu\widetilde{\beta}_{ij} + \mu\widetilde{\beta}_j^k + \mu\eta \ln a_i^k] \\ \ln x_{ij}^k = \beta_{ij} + \beta_j^k + \beta \ln a_i^k + \varepsilon_{ij}^k \end{cases}$$

where $\beta_{ij} = (1 - \mu\beta)^{-1}[\widetilde{\beta}_{ij} + \beta \ln d_{ij}]$, $\beta_j^k = (1 - \mu\beta)^{-1}[\widetilde{\beta}_j^k + \beta \ln d_j^k]$, and $\varepsilon_{ij}^k = \mu\eta^2(1 - \mu\beta)^{-1} \ln a_i^k$. This implies $E[\ln a_i^k \varepsilon_{ij}^k] = \mu\beta^2(1 - \mu\beta)^{-1}E[\left(\ln a_i^k\right)^2] > 0$, and in turn, the upward bias in the OLS estimate of β.

[19]The fixed effects now include the constant term, $-\ln E\left[\exp(u)\right]$.

will be smaller than differences in true productivity, which may artificially raise (in absolute value) the elasticity of exports with respect to the average unit labor requirement. A key benefit of our structural approach is that it allows us to relate Δ_i^k, which is unobserved, to trade variables, which are observed, and hence control explicitly for selection error.[20]

Theorem 5. *Under the assumptions of Theorem 1, the selection error is given by*

$$(16) \qquad \Delta_i^k = -\frac{1}{\theta} \ln \left(1 - m_i^k\right)$$

where θ is the parameter of the Gumbel distribution and $m_i^k \equiv \sum_{i' \neq i} x_{i'i}^k \Big/ \sum_{i'=1}^{I} x_{i'i}^k$ is the import penetration ratio in country i and industry k.

According to Theorem 5, the difference between true and observed productivity is fully determined by the import penetration ratio. As the import penetration ratio goes down, more varieties are produced in country i, which decreases the selection error. In the extreme case where $m_i^k = 0$, country i produces the entire universe of varieties in that particular industry and the selection error disappears. In Appendix D, we derive an analogous result in the case of unknown distributions.

5. Empirical Analysis

We now use the structural Ricardian model developed in the previous sections to estimate the impact of productivity differences on the pattern of trade.

5.1. **Data Description.** Table 1 describes our data set. It includes 15 exporters, 14 European countries plus the United States; 50 importers, both OECD and large non-OECD countries; and 19 manufacturing industries from 1988 to 2001. Sample selection was entirely dictated by the availability of *both* bilateral trade data *and* productivity data comparable across countries and industries. The value of exports, x_{ij}^k, and import penetration ratios, m_i^k, are from the OECD Structural Analysis (STAN) Bilateral Trade Database. Productivity data are from the International Comparisons of Output and Productivity (ICOP) Industrial Database developed by the University of Groningen.[21] A key characteristic of the ICOP

---

[20]Of course, in practice, there may be additional sources of measurement error in productivity. Classical measurement error would bias our OLS estimates of β towards zero.

[21]See http://www.ggdc.net/index-dseries.html for details.

Table 1: Data Set Description

Sources: OECD STAN Bilateral Trade Database; International Comparisons of Output and Productivity Industrial Database

Years 1988-2001

Exporters:	Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, United Kingdom, United States
Importers:	Exporters + Argentina, Australia, Brazil, Canada, Chile, China, Cyprus, Czech Republic, Estonia, Hong Kong, Hungary, Iceland, India, Indonesia, Japan, Korea, Latvia, Lithuania, Luxembourg, Malaysia, Malta, Mexico, New Zealand, Norway, Phillipines, Poland, Russian Federation, Singapore, Slovak Republic, Slovenia, South Africa, Switzerland, Taipei, Thailand, Turkey

Product Classification System: The industrial breakdown presented for the STAN indicators database is based upon the International Standard Industrial Classification (ISIC) Revision 3.

Industry:	STAN Description:	ISIC Rev. 3
Food	Food products, beverages and tobacco	15-16
Textile	Textiles, textile products, leather and footwear	17-19
Wood	Wood and products of wood and cork	20
Paper	Pulp, paper, paper products, printing and publishing	21-22
Chemicals	Chemicals	24
Plastic	Rubber and plastics products	25
Minerals	Other non-metallic mineral products	26
Basic metals	Basic Metals	27
Metal products	Fabricated metal products, except machinery and equipment	28
Machinery	Machinery and equipment, n.e.c.	29
Office	Office, accounting and computing machinery	30
Electrical	Electrical machinery and apparatus, n.e.c.	31
Telecom	Radio, television and communication equipment	32
Medical	Medical, precision and optical instruments, watches and clocks	33
Automobile	Motor vehicles, trailers and semi-trailers	34
Shipbuilding	Building and repairing of ships and boats	351
Aircraft	Aircraft and spacecraft	353
Transport	Railroad equipment and transport equipment n.e.c.	352+359
Other	Manufacturing n.e.c.	36-37

is the use of relative producer prices, or "unit value ratios," to convert output by industry to a common currency. Throughout this section, we use "total hours worked" divided by "value added in 1997 \$US at unit value ratios" as our observed measure of productivity, $\widehat{a}_i^k$, in country i and industry k.

5.2. **Structural Estimates.** Our structural estimates of Equation (15) are presented in Table 2. For each year 1988-2001, we report the OLS estimates of β with and without controlling for selection error. In line with Theorems 1 and 2—and in spite of the potential biases towards zero associated with endogenous trade protection—we find that β is negative and significant at the 1% level for every year in the sample. The largest regression estimate

Table 2: Year-by-Year OLS Regressions
(Dependent Variable: lnx)

Variable	**2001**		**2000**		**1999**		**1998**	
ln a	-0.98	-0.94	-0.73	-0.68	-0.97	-0.84	-1.08	-0.99
	(-20.90)***	(-20.05)***	(-16.67)***	(-12.09)***	(-20.25)***	(-15.29)***	(-20.84)***	(-17.31)***
ln (1-m)	-	0.33	-	0.39	-	0.43	-	0.31
	-	(14.90)***	-	(11.62)***	-	(14.39)***	-	(11.73)***
Exp-Imp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind-Imp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11858	10067	11966	10173	11967	10226	11770	10147
R^2	0.83	0.85	0.83	0.84	0.83	0.84	0.83	0.84

Variable	**1997**		**1996**		**1995**		**1994**	
ln a	-1.42	-1.39	-1.39	-1.17	-1.27	-1.18	-1.31	-1.09
	(-25.59)***	(-22.59)***	(-24.13)***	(-19.84)***	(-22.05)***	(-19.07)***	(-20.56)***	(-15.42)***
ln (1-m)	-	0.17	-	0.30	-	0.21	-	0.14
	-	(-6.54)***	-	(9.81)***	-	(9.12)***	-	(4.85)***
Exp-Imp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind-Imp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11748	10232	11699	10357	11638	10479	9316	8768
R^2	0.83	0.84	0.83	0.84	0.82	0.83	0.83	0.84

Variable	**1993**		**1992**		**1991**		**1990**	
ln a	-1.06	-0.90	-1.20	-1.08	-1.16	-1.12	-0.99	-0.63
	(-18.79)***	(-14.71)***	(-17.87)***	(-15.23)***	(-17.25)***	(-16.20)***	(-15.01)***	(-9.52)***
ln (1-m)	-	0.41	-	0.27	-	0.09	-	0.91
	-	(7.67)***	-	(7.79)***	-	(2.93)***	-	(15.62)***
Exp-Imp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind-Imp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9223	8711	8731	8296	8003	7615	7807	7143
R^2	0.83	0.84	0.84	0.85	0.83	0.84	0.83	0.85

Variable	**1989**		**1988**	
ln a	-1.00	-0.83	-0.74	-0.60
	(-15.16)***	(-11.74)***	(-9.83)***	(-7.12)***
ln (1-m)	-	0.31	-	0.25
	-	(8.73)***	-	(6.06)***
Exp-Imp FE	Yes	Yes	Yes	Yes
Ind-Imp FE	Yes	Yes	Yes	Yes
Observations	7752	7088	6938	6270
R^2	0.83	0.84	0.83	0.83

Note: Values of t-statistics in parentheses, calculated from heteroskedasticity-consistent (White) standard errors

*   Significant at 10% confidence level

** Significant at 5% confidence level

*** Significant at 1% confidence level

of β, in absolute value, is obtained in 1997, which is the year for which the ICOP's relative producer prices were collected.

Is the impact of observed productivity on the pattern of trade economically significant? As mentioned in Section 3, we cannot use our estimate of β to predict the changes in *levels*

of exports associated with a given change in labor productivity. However, we can follow a difference-in-difference approach to predict *relative* changes in exports across countries and industries. Consider, for example, two exporters, i_1 and i_2, and two industries, k_1 and k_2, in 2001. If $\widehat{a}_{i_1}^{k_1}$ decreases by 10%, then our prediction is that:

$$\left(\Delta \ln x_{i_1 j}^{k_1} - \Delta \ln x_{i_1 j}^{k_2}\right) - \left(\Delta \ln x_{i_2 j}^{k_1} - \Delta \ln x_{i_2 j}^{k_2}\right) = \widehat{\beta}_{2001} \Delta \ln \widehat{a}_{i_1}^{k_1} \simeq 9.4\%.$$

This is consistent with a scenario where country i_1's exports of good k_1 go up by 7% and (because of the associated wage increase in country i_1) those of k_2 go down by 2.4%, while they remain unchanged in both industries in country i_2. To help understand the size of the effects reported in Table 2, we can also use the standard deviations of $\ln \widehat{a}_i^k$ and $\ln x_{ij}^k$ in 2001, 0.74 and 2.72, respectively. *Ceteris paribus*, our estimates suggest that a one standard deviation decrease in $\ln \widehat{a}_i^k$ should increase the dependent variable by 0.26 standard deviations.

In line with Theorem 5, we also find that the impact of $\ln \left(1 - m_i^k\right)$ is positive and statistically significant. In our sample, controlling for selection error always reduces the impact of productivity in absolute value. From a quantitative standpoint, the percentage difference in the OLS estimates of β due to selection error is on average equal to 15.4%, ranging from 2.3% in 1997 to 52.4% in 1990.

Finally, we wish to point out that, with or without controlling for selection error, our OLS estimates of β are smaller than those in Eaton and Kortum (2002). Under Eaton and Kortum's (2002) distributional assumptions, $\widehat{\beta}$ should be equal to the parameter of the Gumbel θ. Their estimates of θ range between 3.60 and 12.86 depending on the estimation procedure. For instance, a value of 8.27 is obtained using a simple method-of-moments estimator. Our smaller estimates may reflect, among other things, attenuation bias. As previously mentioned, if productivity is poorly measured at the industry level, then our estimates of β should be biased toward zero.

6. Concluding Remarks

The Ricardian model has long been perceived as a useful pedagogical tool with, ultimately, little empirical content. Over the last twenty years, the Heckscher-Ohlin model, which emphasizes the role of cross-country differences in factor endowments, has generated a considerable amount of empirical work; see e.g. Bowen, Leamer, and Sveikauskas (1987),

Trefler (1993), Trefler (1995), Davis and Weinstein (2001), and Schott (2004). The Ricardian model, which emphasizes productivity differences, has generated almost none.

The main reason behind this lack of popularity is not the existence of strong beliefs regarding the relative importance of factor endowments and technological considerations. Previous empirical work on the Heckscher-Ohlin model unambiguously shows that technology matters. It derives instead from the obvious mismatch between the real world and the extreme assumptions of the Ricardian model. Although the deficiencies of the Ricardian model have not lead to the disappearance of technological considerations from the empirical literature, it has had a strong influence on *how* the relationship between technology and trade has been studied. In the Heckscher-Ohlin-Vanek literature, the factor content of trade remains the main variable of interest.

In this paper, we have developed a structural Ricardian model that puts back productivity differences at the forefront of the analysis. Using this model, we have estimated the impact of productivity differences on the pattern of trade across countries and industries without having to rely on ad-hoc measures of export performance, bilateral comparisons inspired by a two-country model, or unclear orthogonality conditions. Finally, our structural approach has allowed us to control explicitly for the measurement error in productivity associated with the non-random specialization of countries into different varieties.

References

ACEMOGLU, D., P. ANTRAS, AND E. HELPMAN (2006): "Contracts and Technology Adoption," *mimeo Harvard University*.

AL-NAJJAR, N. (2004): "Aggregation and the Law of Large Numbers in Large Economies," *Games and Economic Behavior*, pp. 1–35.

ANDERSON, J. A., AND E. V. WINCOOP (2004): "Trade Costs," *Journal of Economic Literature*, pp. 691–751.

ANTRAS, P., AND E. HELPMAN (2004): "Global Sourcing," *Journal of Political Economy*, pp. 552–580.

BALASSA, B. (1963): "An Empirical Demonstration of Comparative Cost Theory," *Review of Economics and Statistics*, pp. 231–238.

——— (1965): "Trade Liberalization and Revealed Comparative Advantage," *The Manchester School of Economic and Social Studies*, 33, 99–123.

BERNARD, A. B., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *The American Economic Review*, pp. 1268–1290.

BERNARD, A. B., AND J. B. JENSEN (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, pp. 1–25.

BERNARD, A. B., S. J. REDDING, AND P. K. SCHOTT (2006): "Comparative Advantage and Heterogeneous Firms," *(forthcoming) The Review of Economic Studies*.

BHAGWATI, J. (1964): "The Pure Theory of International Trade: A Survey," *The Economic Journal*, pp. 1–84.

BILLINGSLEY, P. (1995): *Probability and Measure*. John Wiley & Sons, Inc.

BOWEN, H. P., E. E. LEAMER, AND L. SVEIKAUSKAS (1987): "Multicountry, Multifactor Tests of the Factor Abundance Theory," *The American Economic Review*, pp. 791–809.

CHANEY, T. (2007): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *mimeo University of Chicago*.

CHOR, D. (2006): "Unpacking Sources of Comparative Advantage: A Quantitative Approach," *mimeo Harvard University*.

COSTINOT, A. (2005a): "Contract Enforcement, Division of Labor, and the Pattern of Trade," *mimeo University of California, San Diego*.

——— (2005b): "Three Essays on Institutions and Trade," *Ph.D. Dissertation, Princeton University*.

CUÑAT, A., AND M. MELITZ (2006): "Labor Market Flexibility and Comparative Advantage," *mimeo Harvard University*.

DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford.

DAVIS, D. R. (1995): "Intra-Industry Trade: A Heckscher-Ohlin-Ricardo Approach," *Journal of International Economics*.

DAVIS, D. R., AND D. E. WEINSTEIN (2001): "An Account of Global Factor Trade," *The American Economic Review*, pp. 1423–1453.

DEARDORFF, A. V. (1980): "The General Validity of the Law of Comparative Advantage," *Journal of Political Economy*, pp. 941–957.

——— (2004): "Local Comparative Advantage: Trade Costs and the Pattern of Trade," *mimeo University of Michigan.*

DORNBUSCH, R., S. FISCHER, AND P. SAMUELSON (1977): "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods," *American Economic Review*, 67, 823–839.

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

GHIRONI, F., AND M. MELITZ (2005): "International Trade and Macroeconomic Dynamics with Heterogeneous Firms," *Quarterly Journal of Economics*, pp. 865–915.

GOLUB, S. S., AND C.-T. HSIEH (2000): "Classical Ricardian Theory of Comparative Advantage Revisited," *Review of International Economics*, pp. 221–234.

HARRIGAN, J. (1997): "Technology, Factor Supplies, and International Specialization: Estimating the Neo-Classical Model," *The American Economic Review*, pp. 475–494.

——— (2005): "Airplanes and Comparative Advantage," *NBER working paper.*

HELPMAN, E., AND P. R. KRUGMAN (1985): *Market Structure and Foreign Trade.* MIT Press.

HELPMAN, E., M. MELITZ, AND S. YEAPLE (2004): "Export versus FDI with Heterogeneous Firms," *The American Economic Review*, pp. 300–316.

JONES, R. W. (1961): "Comparative Advantage and the Theory of Tariffs: A Multi-Country, Multi-Commodity Model," *The Review of Economic Studies*, 28(3), 161–175.

LEAMER, E., AND J. LEVINSOHN (1995): "International Trade Theory: The Evidence," in *Handbook of International Economics*, ed. by G. Grossman, and K. Rogoff, vol. 3, pp. 1139–1194. Elsevier Science, New York.

LEVCHENKO, A. (2004): "Institutional Quality and International Trade," *IMF working paper.*

MACDOUGALL, G. (1951): "British and American Exports: A Study Suggested by the Theory of Comparative Costs, Part," *Economic Journal*, pp. 697–724.

MATSUYAMA, K. (2005): "Credit Market Imperfections and Patterns of International Trade and Capital Flows," *Journal of the European Economic Association*, pp. 714–723.

MELITZ, M. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Inudstry Productivity," *Econometrica*, pp. 1695–1725.

MORROW, P. (2006): "East is East and West is West: A Ricardian-Heckscher-Ohlin Model of Comparative Advantage," *mimeo Michigan University*.

NUNN, N. (2005): "Relationship Specificity, Incomplete Contracts and the Pattern of Trade," *mimeo University of Toronto*.

PETRI, P. A. (1980): "A Ricardian Model of Market Sharing," *Journal of International Economics*, pp. 201–211.

ROMALIS, J. (2004): "Factor Proportions and the Structure of Commodity Trade," *The American Economic Review*, pp. 67–97.

SCHOTT, P. (2004): "Across-Product versus Within-Product Specialization in International Trade," *Quarterly Journal of Economics*, pp. 647–678.

STERN, R. M. (1962): "British and American Productivity and Comparative Costs in International Trade," *Oxford Economic Papers*, pp. 275–303.

TREFLER, D. (1993): "International Factor Price Differences: Leontief was Right!," *Journal of Political Economy*, pp. 961–987.

——— (1995): "The Case of the Missing Trade and Other Mysteries," *The American Economic Review*, pp. 1029–1046.

TYBOUT, J. R., AND M. ROBERTS (1997): "The Decision to Export in Columbia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, pp. 545–563.

VOGEL, J. (2004): "Institutions and Moral Hazard in Open Economies," *mimeo Princeton University*.

WILSON, C. A. (1980): "On the General Structure of Ricardian Models with a Continuum of Goods: Applications to Growth, Tariff Theory, and Technical Change," *Econometrica*, pp. 1675–1702.

APPENDIX A: PROOFS OF THEOREMS 1-5

Proof of Theorem 1. Fix $i \neq j$; by the definition of total exports x_{ij}^k and Assumption A4(i), we have

$$
\begin{aligned}
x_{ij}^k &= \sum_{\omega=1}^{N^k} x_j^k(\omega) \cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\} \\
&= \frac{e_j^k}{\left(p_j^k\right)^{1-\sigma}} \sum_{\omega=1}^{N^k} p_j^k(\omega)^{1-\sigma} \cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\} \\
&= e_j^k \left[\frac{1}{N^k}\sum_{\omega'=1}^{N^k} p_j^k(\omega')^{1-\sigma}\right]^{-1} \left[\frac{1}{N^k}\sum_{\omega=1}^{N^k} p_j^k(\omega)^{1-\sigma} \cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}\right],
\end{aligned}
$$

where the function $\mathbb{I}\{\cdot\}$ is the standard indicator function, i.e. for any event A, we have $\mathbb{I}\{A\} = 1$ if A true, and $\mathbb{I}\{A\} = 0$ otherwise. By Assumption A1, $u_i^k(\omega)$ is independent and identically distributed (i.i.d.) across varieties so same holds for $c_{ij}^k(\omega)$. In addition, $u_i^k(\omega)$ is i.i.d. across countries so $\mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}$ is i.i.d. across varieties as well. This implies that $p_j^k(\omega)^{1-\sigma}$ and $p_j^k(\omega)^{1-\sigma} \cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}$ are i.i.d. across varieties. Moreover, by Assumption A4(ii), $E\left[p_j^k(\omega)^{1-\sigma}\right] < \infty$ so we can use the strong law of large numbers for i.i.d. random variables (e.g. Theorem 22.1 in Billingsley (1995)) to show that

$$
(17) \qquad \frac{1}{N^k}\sum_{\omega'=1}^{N^k}\left[p_j^k(\omega')\right]^{1-\sigma} \overset{a.s.}{\to} E\left[p_j^k(\omega)^{1-\sigma}\right],
$$

as $N^k \to \infty$. Note that $a_i^k > 0$, $d_{ij}^k \geq 1$ ensure that $c_{ij}^k > 0$ whenever $w_i > 0$; hence $E\left[p_j^k(\omega)^{1-\sigma}\right] > 0$. Similarly, Assumption A4(ii) implies that

$$
E\left[p_j^k(\omega)^{1-\sigma} \cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}\right] < \infty,
$$

so we can again use the strong law of large numbers for i.i.d. random variables (e.g. Theorem 22.1 in Billingsley (1995)) to show that

$$
\begin{aligned}
(18) \qquad &\frac{1}{N^k}\sum_{\omega=1}^{N^k} p_j^k(\omega)^{1-\sigma} \cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\} \\
&\qquad \overset{a.s.}{\to} E\left[p_j^k(\omega)^{1-\sigma} \cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}\right],
\end{aligned}
$$

as $N^k \to \infty$. Combining Equations (18) and (17) together with the continuity of the inverse function $x \mapsto x^{-1}$ away from 0, yields by continuous mapping theorem (e.g. Theorem 18.10

(i) in Davidson (1994))

$$(19) \quad \left[\frac{1}{N^k}\sum_{\omega'=1}^{N^k} p_j^k(\omega')^{1-\sigma}\right]^{-1} \left[\frac{1}{N^k}\sum_{\omega=1}^{N^k} p_j^k(\omega)^{1-\sigma}\cdot \mathbb{I}\left\{c_{ij}^k(\omega)=\min_{1\leq i'\leq I} c_{i'j}^k(\omega)\right\}\right]$$

$$\overset{a.s.}{\to}\left\{E\left[p_j^k(\omega)^{1-\sigma}\right]\right\}^{-1}\left\{E\left[p_j^k(\omega)^{1-\sigma}\cdot\mathbb{I}\left\{c_{ij}^k(\omega)=\min_{1\leq i'\leq I} c_{i'j}^k(\omega)\right\}\right]\right\},$$

as $N^k\to\infty$. Note that the quantities in Equation (19) are positive; hence, applying again the continuous mapping theorem (e.g. Theorem 18.10 (i) in Davidson (1994)) to their logarithm we get, with probability one,

$$(20)\quad \ln x_{ij}^k\to\ln e_j^k+\ln E\left[p_j^k(\omega)^{1-\sigma}\cdot\mathbb{I}\left\{c_{ij}^k(\omega)=\min_{1\leq i'\leq I} c_{i'j}^k(\omega)\right\}\right]-\ln E\left[p_j^k(\omega)^{1-\sigma}\right],$$

as $N^k\to\infty$.

Consider $H_i(c_{1j}^k,...,c_{Ij}^k)\equiv E\left[p_j^k(\omega)^{1-\sigma}\cdot\mathbb{I}\left\{c_{ij}^k(\omega)=\min_{1\leq i'\leq I} c_{i'j}^k(\omega)\right\}\right]$. Assumptions A1, A3 and straightforward computations yield

$$(21)\quad H_i(c_{1j}^k,...,c_{Ij}^k)=\left(c_{ij}^k\right)^{1-\sigma}\int_{-\infty}^{+\infty}\exp\left[(1-\sigma)u\right]f(u)\prod_{n\neq i}\left[1-F(\ln c_{ij}^k-\ln c_{nj}^k+u)\right]du.$$

where we let $f(u)\equiv F'(u)$.

Using Equation (21) together with the expressions for the (negative) Gumbel distribution and density, we then have

$$E\left[p_j^k(\omega)^{1-\sigma}\cdot\mathbb{I}\left\{c_{ij}^k(\omega)=\min_{1\leq i'\leq I} c_{i'j}^k(\omega)\right\}\right]$$

$$=(c_{ij}^k)^{1-\sigma}\int_{-\infty}^{+\infty}\theta\exp\left\{(\theta+1-\sigma)u-\mathsf{e}-\left[1+\sum_{i'\neq i}(c_{ij}^k/c_{i'j}^k)^\theta\right]\exp\left(\theta u-\mathsf{e}\right)\right\}du$$

$$=(c_{ij}^k)^{1-\sigma}\exp\left(-\mathsf{e}\frac{\sigma-1}{\theta}\right)\Gamma(\frac{\theta+1-\sigma}{\theta})\left[1+\sum_{i'\neq i}(c_{ij}^k/c_{i'j}^k)^\theta\right]^{-(\theta+1-\sigma)/\theta}$$

$$(22)\quad =\exp\left(-\mathsf{e}\frac{\sigma-1}{\theta}\right)\Gamma(\frac{\theta+1-\sigma}{\theta})\frac{(c_{ij}^k)^{-\theta}}{\left[\sum_{i'=1}^{I}(c_{i'j}^k)^{-\theta}\right]^{(\theta+1-\sigma)/\theta}},$$

where the second equality uses the change of variable $v\equiv\left(1+\sum_{i'\neq i}(c_{ij}^k/c_{i'j}^k)^\theta\right)\exp(\theta u-\mathsf{e})$, and where $\Gamma(\cdot)$ denotes the Gamma function, $\Gamma(t)=\int_0^{+\infty} v^{t-1}\exp(-v)dv$ for any $t>0$. Note

that

$$E\left[p_j^k(\omega)^{1-\sigma}\right] = \sum_{i=1}^{I} E\left[p_j^k(\omega)^{1-\sigma} \cdot \mathbb{1}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}\right],$$

so that by using Equation (22) we get

$$E\left[p_j^k(\omega)^{1-\sigma}\right] = \exp\left(-\mathrm{e}\frac{\sigma-1}{\theta}\right) \Gamma(\frac{\theta+1-\sigma}{\theta}) \frac{1}{\left[\sum_{i'=1}^{I}(c_{i'j}^k)^{-\theta}\right]^{(1-\sigma)/\theta}},$$

and hence

(23)
$$\ln x_{ij}^k \simeq \ln e_j^k - \theta \ln c_{ij}^k - \ln\left(\sum_{i'=1}^{I}(c_{i'j}^k)^{-\theta}\right).$$

for N^k large. Combining the above with the definition of $c_{ij}^k = d_{ij}^k \cdot \mathrm{w}_i \cdot a_i^k$ and Assumption A2, then gives

$$\ln x_{ij}^k \simeq \theta_{ij} + \theta_j^k - \theta \ln a_i^k,$$

for N^k large, where we have let $\theta_{ij} \equiv -\theta \ln(d_{ij} \cdot \mathrm{w}_i)$ and $\theta_j^k \equiv \ln e_j^k - \theta \ln d_j^k - \ln\left(\sum_{i'=1}^{I}(c_{i'j}^k)^{-\theta}\right)$. This completes the proof of Theorem 1. $\qquad\square$

Proof of Theorem 2. Since Assumptions A1-A4 hold, the results of Theorem 1 apply. In particular, we know that, with probability one

$$\ln x_{ij}^k \to \ln e_j^k + \ln E\left[p_j^k(\omega)^{1-\sigma} \cdot \mathbb{1}\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}\right] - \ln E\left[p_j^k(\omega)^{1-\sigma}\right],$$

as $N^k \to \infty$. We now approximate $\ln \tilde{H}_i(c_{1j}^k, ..., c_{Ij}^k) \equiv \ln H_i(c_{1j}^k, ..., c_{Ij}^k) - (1-\sigma)\ln c_{ij}^k$ obtained from Equation (21) by its first order Taylor series around the symmetric case $\ln c_{1j}^k = ... = \ln c_{Ij}^k = \ln c$. Without loss of generality, we choose units of account in each industry k such that $\ln c = 0$. We have

(24)
$$\tilde{H}_i(c_{1j}^k, ..., c_{Ij}^k)\Big|_{(0,...,0)} = \int_{-\infty}^{+\infty} \exp\left[(1-\sigma)u\right] f(u)\left[1-F(u)\right]^{I-1} du,$$

(25)
$$\frac{\partial \tilde{H}_i(c_{1j}^k, ..., c_{Ij}^k)}{\partial \ln c_{ij}^k}\Bigg|_{(0,...,0)} = -(I-1)\int_{-\infty}^{+\infty} \exp\left[(1-\sigma)u\right] f^2(u)\left[1-F(u)\right]^{I-2} du,$$

and, for $i' \neq i$,

(26)
$$\frac{\partial \tilde{H}_i(c_{1j}^k, ..., c_{Ij}^k)}{\partial \ln c_{i'j}^k}\Bigg|_{(0,...,0)} = \int_{-\infty}^{+\infty} \exp\left[(1-\sigma)u\right] f^2(u)\left[1-F(u)\right]^{I-2} du.$$

Let

$$\kappa \equiv \int\limits_{-\infty}^{+\infty} \exp\left[(1-\sigma)u\right] f(u) \left[1 - F(u)\right]^{I-1} du,$$

and

$$\delta \equiv \kappa^{-1} \left[\int\limits_{-\infty}^{+\infty} \exp\left[(1-\sigma)u\right] f^2(u) \left[1 - F(u)\right]^{I-2} du\right].$$

Combining Equations (24), (25), and (26), we then get

$$\ln H_i(c_{1j}^k, ..., c_{Ij}^k) = \ln \kappa + (1-\sigma)\ln c_{ij}^k - (I-1)\,\delta \ln c_{ij}^k + \delta \sum_{i' \neq i} \ln c_{i'j}^k + o\left(\left\|\ln c_j^k\right\|\right)$$

$$(27) \qquad\qquad = \ln \kappa - (\delta I + \sigma - 1)\ln c_{ij}^k + \delta \sum_{i'=1}^{I} \ln c_{i'j}^k + o\left(\left\|\ln c_j^k\right\|\right),$$

where $\left\|\ln c_j^k\right\|^2 = \sum_{i'=1}^{I} \left[\ln c_{i'j}^k\right]^2$ denotes the usual L_2-norm, and $\delta > 0$ only depends on $f(\cdot)$, $F(\cdot)$, σ and I. Combining Equation (27) with the definition of $c_{ij}^k = d_{ij}^k \cdot \mathrm{w}_i \cdot a_i^k$ and Assumption A2, then gives

$$(28) \qquad\qquad \ln H_i(c_{1j}^k, ..., c_{Ij}^k) \simeq \gamma_{ij} + g_j^k - \gamma \ln a_i^k,$$

where

$$\gamma_{ij} \equiv \ln \kappa - (\delta I + \sigma - 1)\ln(d_{ij} \cdot \mathrm{w}_i)$$

$$g_j^k \equiv -(\delta I + \sigma - 1)\ln d_j^k + \delta \sum_{i'=1}^{I} \ln c_{i'j}^k$$

$$\gamma \equiv \delta I + \sigma - 1$$

Note that γ_{ij} does not depend on the good index k, g_j^k does not depend on the country index i, and $\gamma > 0$ is a positive constant which only depends on $f(\cdot)$, $F(\cdot)$, σ and I. Combining Equations (20) and (28) then yields

$$\ln x_{ij}^k \simeq \gamma_{ij} + \gamma_j^k - \gamma \ln a_i^k,$$

for N^k large, where we have let $\gamma_j^k \equiv \ln e_j^k + g_j^k - \ln E\left[p_j^k(\omega)^{1-\sigma}\right]$. This completes the proof of Theorem 2.   □

Proof of Theorem 3. That Equation (6) is sufficient for Property (9) to hold is a matter of simple algebra. We now show that it is also necessary: if Equation (9) is satisfied, then $F(\cdot)$ is Gumbel. First, note that Property (9) is equivalent to

$$\frac{\Pr\left\{p_j^k(\omega) \leq p, c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}}{\Pr\left\{c_{ij}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}} = \Pr\left\{p_j^k(\omega) \leq p\right\},$$

for all $p > 0$ and any $1 \leq i \leq I$, which in turn is equivalent to having

(29)
$$\frac{\Pr\left\{p_j^k(\omega) \leq p, c_{i_1j}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}}{\Pr\left\{c_{i_1j}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}} = \frac{\Pr\left\{p_j^k(\omega) \leq p, c_{i_2j}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}}{\Pr\left\{c_{i_2j}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}},$$

for all $p > 0$ and any $1 \leq i_1, i_2 \leq I$. Using Assumptions A1 and A3, we have

(30)
$$\Pr\left\{c_{i_1j}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\} = \int_{-\infty}^{+\infty} f(u) \prod_{i' \neq i_1} \left[1 - F(\ln c_{i_1j}^k - \ln c_{i'j}^k + u)\right] du$$

and

(31)
$$\Pr\left\{p_j^k(\omega) \leq p, c_{i_1j}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'j}^k(\omega)\right\}$$
$$= \int_{-\infty}^{\ln p - \ln c_{i_1j}^k} f(u) \prod_{i' \neq i_1} \left[1 - F(\ln c_{i_1j}^k - \ln c_{i'j}^k + u)\right] du,$$

with similar expressions for i_2. So the condition in Equation (29) is equivalent to

$$\frac{\int_{-\infty}^{\ln p - \ln c_{i_1j}^k} f(u) \prod_{i' \neq i_1} \left[1 - F(\ln c_{i_1j}^k - \ln c_{i'j}^k + u)\right] du}{\int_{-\infty}^{+\infty} f(u) \prod_{i' \neq i_1} \left[1 - F(\ln c_{i_1j}^k - \ln c_{i'j}^k + u)\right] du}$$
$$= \frac{\int_{-\infty}^{\ln p - \ln c_{i_2j}^k} f(u) \prod_{i' \neq i_2} \left[1 - F(\ln c_{i_2j}^k - \ln c_{i'j}^k + u)\right] du}{\int_{-\infty}^{+\infty} f(u) \prod_{i' \neq i_2} \left[1 - F(\ln c_{i_2j}^k - \ln c_{i'j}^k + u)\right] du},$$

for all $p > 0$ and any $1 \leq i_1, i_2 \leq I$. Differentiating the above equality with respect to $\ln p$ and using the fact that $f(x) > 0$ and hence $F(x) < 1$ for all $x \in \mathbb{R}$, this in turn implies

$$\frac{f(\ln p - \ln c_{i_1 j}^k)\left[1 - F(\ln p - \ln c_{i_2 j}^k)\right]}{f(\ln p - \ln c_{i_2 j}^k)\left[1 - F(\ln p - \ln c_{i_1 j}^k)\right]} = \frac{\int_{-\infty}^{+\infty} f(u) \prod_{i' \neq i_1} \left[1 - F(\ln c_{i_1 j}^k - \ln c_{i' j}^k + u)\right] du}{\int_{-\infty}^{+\infty} f(u) \prod_{i' \neq i_2} \left[1 - F(\ln c_{i_2 j}^k - \ln c_{i' j}^k + u)\right] du},$$

for all $p > 0$ and any $1 \leq i_1, i_2 \leq I$. Since the right-hand side of the above equality does not depend on p, we necessarily have that

(32) $\dfrac{h_F(p/c_{i_1 j}^k)}{h_F(p/c_{i_2 j}^k)}$ only depends on $c_{i_1 j}^k, c_{i_2 j}^k$,

where $h_F(\cdot)$ is a modified hazard function of $F(\cdot)$, i.e. $h_F(x) \equiv [1 - F(\ln x)]^{-1} f(\ln x)$ for any $x > 0$. We now make use of the following Lemma:

Lemma 6. *If for any positive constants c_1 and c_2, $h_F(x/c_1)/h_F(x/c_2)$ only depends on c_1, c_2, then necessarily $h_F(x)$ is of the form $h_F(x) = \mu x^\theta$ where $\mu > 0$ and θ real.*

Proof of Lemma 6. Let $U(t, x) \equiv h_F(tx)/h_F(x)$ for any $x > 0$ and any $t > 0$. Consider $t_1, t_2 > 0$: we have

$$
\begin{aligned}
U(t_1 t_2, x) &= \frac{h_F(t_1 t_2 x)}{h_F(x)} = \frac{h_F(t_1 t_2 x)}{h_F(t_1 x)} \cdot \frac{h_F(t_1 x)}{h_F(x)} \\
&= U(t_2, t_1 x) \cdot U(t_1, x).
\end{aligned}
$$

(33)

If the assumption of Lemma 6 holds then $U(t, x)$ only depends on its first argument t and we can write it $U(t)$. Hence the Equation (33) becomes

$$U(t_1 t_2) = U(t_2) \cdot U(t_1).$$

So, $U(\cdot)$ solves the Hamel equation on $\mathbb{R}_*^+$ and is of the form $U(t) = t^\theta$ for some real θ. This implies that

(34) $h_F(xt) = x^\theta h_F(t).$

Consider $t = 1$ and let $\mu \equiv h_F(1) > 0$; Equation (34) then gives

$$h_X(x) = \mu x^\theta,$$

which completes the proof of Lemma 6.

Proof of Theorem 3 (continued). The result of Lemma 6 allows us to characterize the class of distribution functions $F(\cdot)$ that satisfy Property (9). For any $u \in \mathbb{R}$, we have

$$(35) \qquad \frac{f(u)}{1 - F(u)} = \mu \exp(\theta u).$$

Note that when $u \to -\infty$ we have $f(u), F(u) \to 0$ so that necessarily $\theta > 0$. We can now integrate Equation (35) to obtain for any $u \in \mathbb{R}$

$$(36) \qquad F(u) = 1 - \exp\left[-\exp\left(\theta u + \ln(\frac{\mu}{\theta})\right)\right] \text{ with } \mu > 0 \text{ and } \theta > 0,$$

which belongs to the (negative) Gumbel family. Noting the expected value of the (negative) Gumbel distribution in Equation (36) equals $-\theta^{-1}\left(\ln(\mu/\theta) + \mathsf{e}\right)$, where e is the Euler's constant, we necessarily have, by Assumptions A1 and A4(ii),

$$F(u) = 1 - \exp\left[-\exp\left(\theta u - \mathsf{e}\right)\right] \text{ with } \theta > \sigma - 1 \text{ for any } u \in \mathbb{R},$$

which completes the proof of Theorem 3. □

Proof of Theorem 5. We make use of the following Lemma:

Lemma 7. *Suppose that Assumption A2(ii) holds. Then, for all countries i and goods k,*

$$(37) \qquad \Omega_i^k = \left\{\omega \,\middle|\, c_{ii}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'i}^k(\omega)\right\}$$

Proof of Lemma 7. We proceed by contradiction. Fix an exporter j, and suppose there exists a variety ω_0 of good k and a country $l \neq j$ such that:

$$\begin{cases} c_{jl}^k(\omega_0) = \min_{1 \leq i' \leq I} c_{i'l}^k(\omega_0) \\ c_{jj}^k(\omega_0) \neq \min_{1 \leq i' \leq I} c_{i'j}^k(\omega_0) \end{cases}$$

Then, there must be an exporter $i \neq j$ such that

$$\begin{cases} d_{jl}^k \cdot a_j^k(\omega_0) \leq d_{il}^k \cdot a_i^k(\omega_0) \\ d_{jj}^k \cdot a_j^k(\omega_0) > d_{ij}^k \cdot a_i^k(\omega_0) \end{cases}$$

Since $d_{jj}^k = 1$, the two previous inequalities imply

$$d_{il}^k > d_{ij}^k \cdot d_{jl}^k$$

which contradicts Assumption A2(ii). This completes the proof of Lemma 7.

Proof of Theorem 5 (continued). Without loss of generality, we focus on the case $i = 1$. By definition, we know that $c_{11}^k(\omega) = a_1^k(\omega) \cdot \mathbf{w}_1 \cdot d_{11}^k$. Using Lemma 7 then yields

$$(38) \quad E\left[\ln a_1^k(\omega) \,|\omega \in \Omega_1^k\right] = E\left[\ln c_{11}^k(\omega)\,\mathbb{I}\left\{c_{11}^k(\omega) = \min_{1 \le i' \le I} c_{i'1}^k(\omega)\right\}\right] - \ln(d_{11}^k \cdot \mathbf{w}_1) - \ln \pi_{11}^k$$

where we have let $\pi_{11}^k \equiv \Pr\left\{c_{11}^k(\omega) = \min_{1 \le i' \le I} c_{i'1}^k(\omega)\right\}$. Now, consider

$$G_1(c_{11}^k, ..., c_{I1}^k) \equiv E\left[c_{11}^k(\omega)\,\mathbb{I}\left\{c_{11}^k(\omega) = \min_{1 \le i' \le I} c_{i'1}^k(\omega)\right\}\right]$$

The expressions for $G_1(c_{11}^k, ..., c_{I1}^k)$ and π_{11}^k are readily available from the proof of Theorem 1 when the result in Equation (22) is evaluated at $\sigma = 0$ and $\sigma = 1$, respectively. Hence,

$$\ln \widehat{a}_1^k = \frac{\mathsf{e}}{\theta} + \ln \Gamma\left(\frac{\theta + 1}{\theta}\right) - \frac{1}{\theta} \ln \sum_{i=1}^{I} (c_{i1}^k)^{-\theta} - \ln(d_{11}^k \cdot \mathbf{w}_1).$$

In addition, from Assumption A1 and $c_{11}^k = a_{11}^k \cdot d_{11}^k \cdot \mathbf{w}_1$ we know that

$$(39) \quad\quad\quad\quad \ln E\left[a_1^k(\omega)\right] = \ln c_{11}^k - \ln(d_{11}^k \cdot \mathbf{w}_1) + \ln E[\exp u]$$

so that $\Delta_1^k \equiv \ln E\left[a_1^k(\omega)\right] - \ln \widehat{a}_1^k$ satisfies

$$(40) \quad\quad\quad \Delta_1^k = \ln E[\exp u] - \frac{\mathsf{e}}{\theta} - \ln \Gamma\left(\frac{\theta + 1}{\theta}\right) - \frac{1}{\theta} \ln\left(\frac{(c_{11}^k)^{-\theta}}{\sum_{i=1}^{I}(c_{i1}^k)^{-\theta}}\right)$$

Now, from Equation (23), we know that for every i,

$$x_{i1}^k = e_1^k \cdot \frac{(c_{i1}^k)^{-\theta}}{\sum_{i'=1}^{I}(c_{i'1}^k)^{-\theta}}.$$

Using the definition of the import penetration ratio m_1^k in country 1 and industry k given in Theorem 5, we then get

$$(41) \quad\quad\quad\quad\quad\quad m_1^k = \frac{\sum_{i \ne 1}(c_{i1}^k)^{-\theta}}{\sum_{i=1}^{I}(c_{i1}^k)^{-\theta}}$$

In addition, under the Gumbel assumption, we have that

$$(42) \quad\quad\quad\quad\quad\quad E[\exp u] = \exp\left[\frac{\mathsf{e}}{\theta}\Gamma\left(\frac{\theta + 1}{\theta}\right)\right]$$

so combining Equations (40), (41), and (42) yields $\Delta_1^k = -\frac{1}{\theta}\ln(1 - m_1^k)$ as desired. $\square$

APPENDIX B: BERTRAND COMPETITION

Instead of Assumption A3, we now consider:

A3'. *In any country j, the price $p_j^k(\omega)$ paid by buyers of variety ω of good k is*

$$p_j^k(\omega) = \min\left\{\min_{i'\neq i^*}\left[c_{i'j}^k(\omega)\right], \overline{m}c_{i^*j}^k(\omega)\right\},$$

where $c_{i^*j}^k(\omega) = \min_{1\leq i'\leq I} c_{ij}^k(\omega)$ and $\overline{m} = \sigma/(\sigma-1)$ is the monopoly markup.

This is in the spirit of Bernard, Eaton, Jensen, and Kortum (2003): the producer with the minimum cost may either charge the cost of its closest competitor or the monopoly price. We then have the following result:

Theorem 8. *Suppose that Assumptions A1, A2, A3', and A4 hold. In addition, assume that the number of varieties N^k of any good k is large, and that technological differences across exporters are small: $c_{1j}^k \simeq ... \simeq c_{Ij}^k$. Then, for any exporter i, any importer $j \neq i$, and any good k,*

(43) $$\ln x_{ij}^k \simeq \widetilde{\gamma}_{ij} + \widetilde{\gamma}_j^k - \widetilde{\gamma}\ln a_i^k.$$

where $\widetilde{\gamma} > -(\sigma-1)/(I-1)$.

Under Bertrand competition, the qualitative insights of Theorem 2 remain valid, albeit in a weaker form. We obtain new importer–exporter and importer–industry fixed effects, $\widetilde{\gamma}_{ij}$ and $\widetilde{\gamma}_j^k$, and a new parameter $\widetilde{\gamma}$ constant across countries and industries. However, the restriction $\widetilde{\gamma} > -(\sigma-1)/(I-1)$ is less stringent than in the case of perfect competition. When $\sigma \to 1$, that is when varieties become perfect substitutes, or when $I \to +\infty$, that is when the number of exporters is very large, this collapses to: $\widetilde{\gamma} \geq 0$.

Proof of Theorem 8. Compared to the proof of Theorem 2, the only difference comes from the expression of $H_i(c_{1j}^k,...,c_{Ij}^k) = E\left[p_j^k(\omega)^{1-\sigma}\cdot \mathbb{I}\left\{c_{ij}^k(\omega) = \min_{1\leq i'\leq I} c_{i'j}^k(\omega)\right\}\right]$. Assumptions

A1, A3' and straightforward computations now yield

$$(44) \quad H_i(c_{1j}^k, ..., c_{Ij}^k) = \left(c_{ij}^k\right)^{1-\sigma} \int\limits_{-\infty}^{+\infty} f(u_1) du_1 \int\limits_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} \cdot$$

$$\sum_{i'\neq i} \left\{ \prod_{i''\neq i,i'} \left[1 - F(\ln c_{ij}^k - \ln c_{i''j}^k + u_2)\right] f(\ln c_{ij}^k - \ln c_{i'j}^k + u_2) \right\} du_2$$

where we let $f(u) \equiv F'(u)$.

As previously, we approximate $\ln \tilde{H}_i(c_{1j}^k, ..., c_{Ij}^k) \equiv \ln H_i(c_{1j}^k, ..., c_{Ij}^k) - (1-\sigma)\ln c_{ij}^k$, obtained from Equation (44), by its first order Taylor series around the symmetric case $\ln c_{1j}^k = ... = \ln c_{Ij}^k = 0$. We have

$$\tilde{H}_i(c_{1j}^k, ..., c_{Ij}^k)\Big|_{(0,...,0)} = \int\limits_{-\infty}^{+\infty} f(u_1) du_1 \int\limits_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} \cdot$$

$$(45) \qquad\qquad (I-1)\left[1 - F(u_2)\right]^{I-2} f(u_2) du_2,$$

$$\frac{\partial \tilde{H}_i(c_{1j}^k, ..., c_{Ij}^k)}{\partial \ln c_{ij}^k}\Bigg|_{(0,...,0)} = -(I-1)\int\limits_{-\infty}^{+\infty} f(u_1) du_1 \int\limits_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} \cdot$$

$$(46) \qquad\qquad \left\{-f'(u_2)\left[1 - F(u_2)\right]^{I-2} + (I-2)f^2(u_2)\left[1 - F(u_2)\right]^{I-3}\right\} du_2,$$

and, for $i' \neq i$,

$$\frac{\partial H_i(c_{1j}^k, ..., c_{Ij}^k)}{\partial \ln c_{i'j}^k}\Bigg|_{(0,...,0)} = \int\limits_{-\infty}^{+\infty} f(u_1) du_1 \int\limits_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} \cdot$$

$$(47) \qquad\qquad \left\{-f'(u_2)\left[1 - F(u_2)\right]^{I-2} + (I-2)f^2(u_2)\left[1 - F(u_2)\right]^{I-3}\right\} du_2.$$

Let then

$$(48) \quad \kappa \equiv (I-1)\int\limits_{-\infty}^{+\infty} f(u_1) du_1 \int\limits_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} \left[1 - F(u_2)\right]^{I-2} f(u_2) du_2,$$

and

$$\delta \equiv \kappa^{-1} \int_{-\infty}^{+\infty} f(u_1)\, du_1 \int_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} \cdot$$

(49) $$\left\{-f'(u_2)\left[1 - F(u_2)\right]^{I-2} + (I-2) f^2(u_2)\left[1 - F(u_2)\right]^{I-3}\right\} du_2.$$

Combining Equations (45), (46), and (47), we get

$$\ln H_i(c_{1j}^k, ..., c_{Ij}^k) \;=\; \ln\kappa + (1-\sigma)\ln c_{ij}^k - (I-1)\,\delta \ln c_{ij}^k + \delta \sum_{i'\neq i} \ln c_{i'j}^k + o\left(\left\|\ln c_j^k\right\|\right)$$

$$=\; \ln\kappa - (\delta I + \sigma - 1)\ln c_{ij}^k + \delta \sum_{i'=1}^{I} \ln c_{i'j}^k + o\left(\left\|\ln c_j^k\right\|\right),$$

where $\left\|\ln c_j^k\right\|^2 = \sum_{i'=1}^{I} \left[\ln c_{i'j}^k\right]^2$ as previously, and δ only depends on $f(\cdot)$, $F(\cdot)$, σ and I. Let

$$\widetilde{\gamma} \equiv \delta I + \sigma - 1$$

It remains to be shown that $\widetilde{\gamma} > -(\sigma - 1)/(I-1)$.

For this, let $I(u_1) \equiv \int_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} f'(u_2)\left[1 - F(u_2)\right]^{I-2} du_2$. We can rearrange $I(u_1)$ as

(50) $$I(u_1) = \int_{u_1}^{u1+\ln\overline{m}} \left[\exp u_2\right]^{1-\sigma} f'(u_2)\left[1 - F(u_2)\right]^{I-2} du_2$$

$$+ \left[\overline{m}\exp u_1\right]^{1-\sigma} \int_{u1+\ln\overline{m}}^{+\infty} f'(u_2)\left[1 - F(u_2)\right]^{I-2} du_2$$

$$= - \left[\exp u_1\right]^{1-\sigma} f(u_1)\left[1 - F(u_1)\right]^{I-2}$$

$$- (1-\sigma) \int_{u_1}^{u1+\ln\overline{m}} \left[\exp u_2\right]^{1-\sigma} f(u_2)\left[1 - F(u_2)\right]^{I-2} du_2$$

$$+ (I-2) \int_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} f^2(u_2)\left[1 - F(u_2)\right]^{I-3} du_2$$

where the second equality uses a simple integration by parts. Combining Equations (49) and (50), we then get

$$\delta = \kappa^{-1} \int\limits_{-\infty}^{+\infty} f\left(u_1\right) du_1 \left\{ \left[\exp u_1\right]^{1-\sigma} f(u_1) \left[1 - F(u_1)\right]^{I-2} \right.$$

(51)
$$\left. - (\sigma - 1) \int\limits_{u_1}^{u1+\ln\overline{m}} \left[\exp u_2\right]^{1-\sigma} f(u_2) \left[1 - F(u_2)\right]^{I-2} du_2 \right\}.$$

Using Equations (48) and (51), we then have

$$(I - 1)\delta + \sigma - 1$$

$$= (I - 1)\kappa^{-1} \int\limits_{-\infty}^{+\infty} \left[\exp u_1\right]^{1-\sigma} f^2(u_1) \left[1 - F(u_1)\right]^{I-2} du_1$$

$$- (I - 1)(\sigma - 1)\kappa^{-1} \int\limits_{-\infty}^{+\infty} f\left(u_1\right) du_1 \int\limits_{u_1}^{u1+\ln\overline{m}} \left[\exp u_2\right]^{1-\sigma} f(u_2) \left[1 - F(u_2)\right]^{I-2} du_2$$

$$+ (I - 1)(\sigma - 1)\kappa^{-1} \int\limits_{-\infty}^{+\infty} f\left(u_1\right) du_1 \int\limits_{u_1}^{+\infty} \left[\min\left(\exp u_2, \overline{m}\exp u_1\right)\right]^{1-\sigma} \left[1 - F(u_2)\right]^{I-2} f(u_2) du_2$$

$$= (I - 1)\kappa^{-1} \int\limits_{-\infty}^{+\infty} \left[\exp u_1\right]^{1-\sigma} f^2(u_1) \left[1 - F(u_1)\right]^{I-2} du_1$$

$$+ (I - 1)(\sigma - 1)\kappa^{-1} \int\limits_{-\infty}^{+\infty} f\left(u_1\right) du_1 \int\limits_{u_1+\ln\overline{m}}^{+\infty} \left[\overline{m}\exp u_1\right]^{1-\sigma} \left[1 - F(u_2)\right]^{I-2} f(u_2) du_2,$$

which is positive by inspection. Hence, writing $\widetilde{\gamma} = I(I-1)^{-1}[(I-1)\delta+\sigma-1]-(I-1)^{-1}(\sigma-1)$ and using $(I - 1)\delta + \sigma - 1 > 0$ yields the desired result: $\widetilde{\gamma} > -(I - 1)^{-1}(\sigma - 1)$. □

Appendix C: Revealed Comparative Advantage

This Appendix illustrates how our theoretical framework may be used to reveal the pattern of comparative advantage. The basic idea is to follow the three-term decomposition offered by Theorems 1 and 2 and consider a panel model of the form

$$(52) \qquad \ln x_{ij}^k = \delta_{ij} + \delta_j^k + \delta_i^k + \varepsilon_{ij}^k,$$

where δ_{ij}, δ_j^k, and δ_i^k are treated as importer–exporter, importer–industry, and exporter-industry fixed effects, respectively, and ε_{ij}^k is an error term.[22] In the absence of ε_{ij}^k, there would be, for any pair of exporters, a unique ranking of relative exports by industry, as suggested in Corollary 4. Furthermore, this ranking would be entirely determined by the cross-industry and cross-country variation of the third term, δ_i^k. If $\varepsilon_{ij}^k = 0$, then $\frac{x_{i_1 j}^{k_1}}{x_{i_2 j}^{k_1}} > \frac{x_{i_1 j}^{k_2}}{x_{i_2 j}^{k_2}}$ if and only if $\delta_{i_1}^{k_1} - \delta_{i_2}^{k_1} > \delta_{i_1}^{k_2} - \delta_{i_2}^{k_2}$. Hence, the estimates of δ_i^k can be interpreted as a *revealed measure*—up to a monotonic transformation—of the fundamental productivity levels, a_i^k, that determine the Ricardian chain of comparative advantage.

Table 3 reports the ranking of the OLS estimates of $\left(\delta_{US}^k - \delta_i^k \right)$ across industries for all exporters $i \neq United\ States$ in 1997, from the highest to the lowest value. According to our estimates, "Aircraft" always is the first industry in the chain of comparative advantage of the United States. Compared to any other country in our sample, the United States tend to export more in the aircraft industry than in any other industry. The industries at the bottom of the US chain of comparative advantage tend to be "Basic Metals" and "Textile," depending on the identity of the other exporter. A notable exception is Germany for which "Automobile" is the bottom industry.

Note that there is a close connection between Balassa (1965) and the present paper. Like Balassa (1965), we offer a methodology that uses data on relative exports to infer the pattern of comparative advantage across countries and industries. In his well-known paper, the revealed comparative advantage of country i in industry k is defined as

$$\left(\frac{x_{iWorld}^k}{\sum_{k'=1}^K x_{iWorld}^{k'}} \right) \Bigg/ \left(\frac{\sum_{i'=1}^I x_{i'World}^k}{\sum_{i'=1}^I \sum_{k'=1}^K x_{i'World}^{k'}} \right)$$

---

[22]We thank Stephen Redding for suggesting that approach.

Table 3: Ranking of Industries in the Chain of Comparative Advantage of the United States

Industry	Other Exporting Country															
	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	NLD	PRT	SWE	Mean	Balassa
Aircraft	.	1	1	.	1	1	1	.	.	.	1	1	.	1	1	1
Office	1	2	2	1	2	3	2	.	1	14	2	7	.	2	2	3
Medical	4	4	6	7	3	7	3	.	4	10	5	5	.	6	3	4
Telecom	3	6	4	3	4	15	6	.	2	11	4	4	.	16	4	2
Shipbuilding	2	3	3	8	8	14	5	.	.	.	3	6	.	4	5	18
Transport	11	5	9	2	5	2	10	.	.	2	12	3	.	5	6	8
Wood	10	7	7	12	7	17	4	1	5	3	8	2	7	17	7	9
Other	6	13	8	10	12	5	7	3	10	6	17	9	1	7	8	12
Machinery	8	9	12	13	6	12	9	2	7	5	14	8	2	10	9	6
Automobile	12	15	19	4	17	9	16	.	3	1	11	14	.	14	10	7
Electrical	7	10	13	11	10	16	14	.	8	8	9	10	.	12	11	5
Paper	14	8	10	5	9	19	8	5	9	15	7	11	3	18	12	10
Food	5	12	5	18	13	4	11	4	13	13	6	19	4	3	13	16
Chemicals	9	18	11	17	11	6	15	.	6	16	10	17	.	15	14	15
Plastic	13	14	14	14	14	10	13	6	12	9	13	13	5	11	15	13
Metal products	16	11	17	16	15	13	12	.	11	7	15	15	.	13	16	11
Minerals	15	16	15	6	19	11	19	7	14	4	18	12	6	9	17	17
Textile	18	17	16	15	18	8	17	8	15	12	19	16	8	8	18	19
Basic metals	17	19	18	9	16	18	18	.	.	.	16	18	.	19	19	14

where x_{iWorld}^k are the total exports of country i in industry k. The ranking of industries in terms of their Balassa's (1965) revealed comparative advantage for $i = United\ States$ is reported in the last column of Table 4.

There are, however, two important differences between Balassa's (1965) approach and ours. First, our empirical strategy is theoretically grounded. The ranking of the OLS estimates of $\left(\delta_{US}^k - \delta_i^k\right)$ is the empirical counterpart to the ranking of $\left(\ln a_i^k - \ln a_{US}^k\right)$ in our model. Second, our approach fundamentally is about pairwise comparisons. Our fixed effects aim to uncover which of Portugal and England is the country relatively better at producing wine than cloth. They do not try to find out whether Portugal is good at producing wine compared to an intuitive but ad-hoc benchmark. Unlike Balassa (1965), we do not aggregate exports across countries and industries, which—according to our model—allows us to separate the impact of technological differences from transport costs and demand differences.

APPENDIX D: SELECTION ERROR WITH UNKNOWN DISTRIBUTIONS

In this Appendix, we derive a result similar to that of Theorem 5 but in the case of unknown distributions.

Theorem 9. *Under the assumptions of Theorem 2, the selection error is given by*

$$\Delta_1^k \simeq \chi - \frac{\nu}{\gamma} \ln\left(1 - m_1^k\right),$$

where γ is the constant from Theorem 2, χ and $\nu \in \mathbb{R}$, and $m_i^k \equiv \sum_{i' \neq i} x_{i'i}^k \Big/ \sum_{i'=1}^{I} x_{i'i}^k$.

Proof of Theorem 9. Without loss of generality, we again focus on the case $i = 1$. As in the proof of Theorem 5, we have

$$(53)\ \ln E\left[a_1^k(\omega)\,|\omega \in \Omega_1^k\right] = \ln E\left[c_{11}^k(\omega)\mathbb{I}\left\{c_{11}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'1}^k(\omega)\right\}\right] - \ln(d_{11}^k \cdot \mathtt{w}_1) - \ln \pi_{11}^k$$

where $\pi_{11}^k \equiv \Pr\left\{c_{11}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'1}^k(\omega)\right\}$, as before. As previously, we let $G_1(c_{11}^k, ..., c_{I1}^k) \equiv E\left[c_{11}^k(\omega)\mathbb{I}\left\{c_{11}^k(\omega) = \min_{1 \leq i' \leq I} c_{i'1}^k(\omega)\right\}\right]$. Now, consider a first order Taylor development of $\ln G_1(c_{11}^k, ..., c_{I1}^k)$ around $\ln c_{11}^k = ... = \ln c_{I1}^k = 0$. The latter is readily available from the proof of Theorem 2: we only need to set $\sigma = 0$ in Equations (24)-(26). By letting

$$\lambda \equiv \int_{-\infty}^{+\infty} [\exp u] f(u) \left[1 - F(u)\right]^{I-1} du$$

and

$$\mu \equiv \lambda^{-1}\left[\int_{-\infty}^{+\infty} [\exp u] f^2(u) \left[1 - F(u)\right]^{I-2} du\right]$$

we get

$$(54)\qquad \ln G_1(c_{11}^k, ..., c_{I1}^k) = \ln \lambda - (\mu I - 1) \cdot \ln c_{11}^k + \mu \cdot \sum_{i=1}^{I} \ln c_{i1}^k + o\left(\left\|\ln c_1^k\right\|\right),$$

where $\|\cdot\|^2$ is the L_2-norm as previously. We can follow the same approach for $\ln \pi_{11}^k$. By setting $\sigma = 1$ in Equations (24)-(26), we get

$$(55)\qquad \ln \pi_{11}^k = -\ln I - \tau \cdot I \cdot \ln c_{11}^k + \tau \cdot \sum_{i=1}^{I} \ln c_{i1}^k + o\left(\left\|\ln c_1^k\right\|\right),$$

where

$$\tau \equiv I \cdot \left[\int_{-\infty}^{+\infty} f^2(u) \left[1 - F(u)\right]^{I-2} du \right].$$

Combining Equations (53) with (54) − (55), we then obtain

(56)

$$\ln E\left[a_i^k(\omega) \,|\, \omega \in \Omega_i^k\right] = \ln(\lambda I) + \left[(\tau - \mu)I + 1\right] \cdot \ln c_{11}^k - \ln(d_{11}^k \cdot \mathtt{w}_1) + (\mu - \tau) \sum_{i=1}^{I} \ln c_{i1}^k + o\left(\left\|\ln c_1^k\right\|\right).$$

Now, combining Assumption A1 with $c_{11}^k = a_{11}^k \cdot d_{11}^k \cdot \mathtt{w}_1$ gives

(57)
$$\ln E\left[a_1^k(\omega)\right] = \ln c_{11}^k - \ln(d_{11}^k \cdot \mathtt{w}_1) + \ln E[\exp u]$$

Combining Equations (56) − (57) yields the following expression for Δ_1^k

(58)
$$\Delta_1^k = \ln E[\exp u] - \ln(\lambda I) - (\mu - \tau) \sum_{i=1}^{I} \ln\left(\frac{c_{i1}^k}{c_{11}^k}\right) + o\left(\left\|\ln c_1^k\right\|\right).$$

Now, fix any constant γ, and note that in the neighborhood of $\ln c_{11}^k = ... = \ln c_{I1}^k = 0$ we have:

$$\ln\left(\sum_{i=1}^{I} \left(\frac{c_{i1}^k}{c_{11}^k}\right)^{-\gamma}\right) = \ln I + \frac{1}{I}\sum_{i=1}^{I} \left(\frac{c_{i1}^k}{c_{11}^k}\right)^{-\gamma} - 1 + o\left(\left\|\ln c_1^k\right\|\right),$$

and, in addition, for any i,

$$\left(\frac{c_{i1}^k}{c_{11}^k}\right)^{-\gamma} = 1 - \gamma \ln\left(\frac{c_{i1}^k}{c_{11}^k}\right) + o\left(\left\|\ln c_1^k\right\|\right).$$

Hence,

$$\sum_{i=1}^{I} \ln\left(\frac{c_{i1}^k}{c_{11}^k}\right) = \frac{I}{\gamma}\left[\ln I - \ln\left(\sum_{i=1}^{I} \left(\frac{c_{i1}^k}{c_{11}^k}\right)^{-\gamma}\right)\right] + o\left(\left\|\ln c_1^k\right\|\right).$$

From the proof of Theorem 2 and its Equation (27) recall that

(59)
$$\sum_{i=1}^{I} \left(\frac{c_{i1}^k}{c_{11}^k}\right)^{-\gamma} = \left[\sum_{i=1}^{I} \frac{x_{i1}^k}{x_{11}^k}\right]\left[1 + o\left(\left\|\ln c_1^k\right\|\right)\right]$$

where we let γ be as defined in the proof of Theorem 2. >From Equations (58) and (59) we then get

$$\Delta_1^k = \ln E[\exp u] - \ln(\lambda I) - (\mu - \tau)\frac{I}{\gamma}\left[\ln I + \ln\left(1 - m_1^k\right)\right] + o\left(\left\|\ln c_1^k\right\|\right),$$

where m_1^k is the import penetration ratio in country 1 and industry k defined in Theorem 5. Let then $\nu \equiv (\mu - \tau)I$, so

$$\nu = I \left(\frac{\int_{-\infty}^{+\infty} [\exp u] f^2(u) \left[1 - F(u)\right]^{I-2} du}{\int_{-\infty}^{+\infty} [\exp u] f(u) \left[1 - F(u)\right]^{I-1} du} - \frac{\int_{-\infty}^{+\infty} f^2(u) \left[1 - F(u)\right]^{I-2} du}{\int_{-\infty}^{+\infty} f(u) \left[1 - F(u)\right]^{I-1} du} \right)$$

and $\chi \equiv \ln E[\exp u] - \ln(\lambda I) - \gamma^{-1}(\mu - \tau)I \ln I$, i.e.

$$\chi = \ln E[\exp u] - \ln(\lambda I) - \frac{\nu}{\gamma} \ln I$$

Then,

$$\Delta_1^k \simeq \chi - \frac{\nu}{\gamma} \ln\left(1 - m_1^k\right),$$

as desired. Note that under the assumptions of Theorem 1, the parameter γ equals θ and we have the following equalities:

$$\nu = 1 \text{ and } \chi = 0$$

so the results of Theorem 9 reduce to those of Theorem 5. $\qquad\square$

APPENDIX E: MULTIPLE FACTORS OF PRODUCTION

In order to introduce multiple factors of production into the Eaton and Kortum's (2002) model, we follow Costinot's (2005b) chapter 3.[23] Suppose that there are $f = 1, ..., F$ factors of production, which are perfectly mobile across industries and immobile across countries. Further, assume that the production technology is Cobb-Douglas in all industries and countries so that the constant unit cost of variety ω of good k in country i equals

$$(60) \qquad c_i^k(\omega) = a_i^k(\omega) \cdot \prod_{f=1}^{F} w_{if}^{\alpha_f^k},$$

where w_{if} is factor f's reward in country i; and $0 < \alpha_f^k < 1$ is the intensity of factor f in the production of good k. Compared to Section 2, $a_i^k(\omega)$ now is the inverse of *total factor productivity* in the production of variety ω of good j in country i. Combining Assumption A1 and Equation (60), we get

$$\ln c_i^k(\omega) = \ln a_i^k + \sum_{f=1}^{F} \alpha_f^k \ln w_{if} + u_i^k(\omega).$$

Following the same reasoning as in Section 3, we may now generalize Theorems 1 and 2:

Theorem 10. *Suppose that the assumptions of Theorem 1 or 2 hold. Then, for any importer j, any exporter $i \neq j$, and any product k,*

$$(61) \qquad \ln x_{ij}^k \simeq \eta_{ij} + \eta_j^k - \eta \ln a_i^k - \eta \sum_{f=1}^{F} \alpha_f^k \ln(w_{if}/w_{i1}).$$

where $\eta > 0$.

The interpretation of the two fixed effects is the same as in Section 3. The additional term $\eta \sum_{f=1}^{F} \alpha_f^k \ln w_{if}$ captures the impact of cross-country differences in relative factor prices—and therefore, cross-country differences in factor endowments—on the pattern of trade. It could be estimated with data on factor prices and factor shares.

---

[23]Independently, Chor (2006) also incorporates multiple factors of production—together with institutional differences—into the Eaton and Kortum's (2002) model.