A New Measure of Economic Distance

Eric O'N. Fisher
California Polytechnic State University and Utah State University
efisher@calpoly.edu

John Gilbert
Utah State University
jgilbert@usu.edu

Kathryn G. Marshall
California Polytechnic State University
kgmarsha@calpoly.edu

Reza Oladi*
Utah State University
reza.oladi@usu.edu

Abstract

This paper defines a new measure of economic distance. Using consistent cross-country data, we estimate local unit costs for 35 sectors in 39 countries. The distance between two countries is the largest percentage difference in unit costs among all sectors. If all goods are traded, this distance is the uniform *ad valorem* tariff that shuts down bilateral trade. The network induced by the closest 10% of these distances has a large component with two clusters, one corresponding to the advanced economies and another to the emerging economies. China, India, and several other countries are isolated components, indicating that their unit costs are idiosyncratic. We also introduce a new measure of revealed comparative advantage. (JEL Codes: F1)

1 Introduction

This paper presents a new measure of economic distance: two economies are "close" if their local unit costs for every good are similar. This measure was developed with empirical applications in mind, and it takes full advantage of the recent flowering of internationally consistent macroeconomic data on the production sets of different countries. This distance has an elegant interpretation as a trade cost.

There are many underlying reasons for trade costs. Some foreign countries are far away, and it costs a lot to ship goods to a distant market. It is harder to sell a good abroad than at home because language is a barrier. Perhaps shipping across the border involves unfamiliar bureaucratic impediments. There may be an

---

explicit tariff or non-tariff barrier. The foreign culture may well seem alien to any firm trying to expand sales by exporting. It is not easy to obtain the usual financing that facilitates the relations between local suppliers and wholesalers. The foreign legal system makes it much more difficult to enforce the norms that allow for transacting business in a repeated relationship. Exchange rate risk makes foreign sales less attractive. This list is far from exhaustive.

Still, we exploit a deep insight. *All these impediments to trade will be reflected in differences in unit costs.* Any departure from the law of one price must reflect an implicit trade cost. In fact, the only reason that two physically homogeneous goods sell for different prices in disparate markets is that it's costly to trans-ship them. It doesn't matter if it costs money to bribe a customs officer, to pay for airfreight, or to hire a translator, if an apple sells for $1 in New York and a *pomme* sells for $2 in Paris, then you can be sure that the generalized cost of shipping an apple across markets is at least $1.

This simple insight is the essence of our measure of bilateral economic distance. Assume that the estimated local unit cost of a *pomme* in Paris is 50% above the world price and that of an *Apfel* in Berlin is 30% above it. Then this cost differential is 20%. Do the same calculation for every good recorded in the input-output tables of France and Germany. Of course, these goods include traded and non-traded goods, but we are agnostic about this difference.[1] The maximum of the absolute value of these differences is the economic distance between France and Germany. In fact, it is 68%. So a uniform tariff of 68% would shut down all trade between France and Germany; also, since we are including non-traded goods in our calculation, a 68% discriminatory tax on foreigners would discourage every German from renting an apartment in Paris and *vice versa*.

Because we want to apply our measure, we are motivated by the data in hand. The production structure of an economy is a technology matrix that records the unit values of direct and indirect resource requirements for each sector. Our data have thirty-five sectors and four factors. These data are actually an overdetermined system of 35 equations in the 4 unknown factor prices. Our measure exploits this insight fully.

Let there be n goods and f factors. Local unit costs lie in a cone in n-dimensional space spanned by the f vectors that describe factor uses. Consider, for example, a simple Ricardian economy where it takes one hour of labor to make Good 1 and two hours of labor to make Good 2. All possible local unit costs are spanned by the vector $(1,2)^T$. What would local unit costs be if this economy faced world prices $p = (1,1)^T$? The usual answer is that the economy would specialize completely in the first good, and the second sector would shut down. National income accounts would show only one active local sector.

One does not see this paucity of active sectors in the data. One actually observes almost every good produced in practically every country; this phenomenon cannot be explained by simple applied general equilibrium theory. An explanation is that every national economy is somehow protected by price wedges that allow many local industries to be active. Think of iceberg transportation costs as uniform percentage price markups for every good. What are the smallest such costs that would allow both sectors to be active in this case? The least squares project of world prices onto local unit costs is $(0.6, 1.2)^T$; this is the point closest to world prices that is consistent with both sectors operating. Imports of the second good drive down local wage from 1.0 to 0.6 and thus lowers the unit costs of both goods. If iceberg transportation costs were at least 40%, importing the second good and exporting the first good would no longer make sense. Our distance measure would record this country as being 0.4 from any technology that could produce at world prices. This simple example captures the essence of the projection matrices that estimate local unit costs.

The simple Ricardian model is more general than many economists realize. Macroeconomic accounts record many more goods than factors of production. In this case, the observed vector of outputs occurs generically in the interior of a flat on the production possibility frontier. Small exogenous changes on the demand side will have no effect on the supporting price vector as long as all sectors remain active. Prices

---

[1]If the pound sterling gets strong enough and there is a price war on flights between Heathrow and JFK, a Broadway play eventually becomes a traded good for the right kind of Londoner.

are determined by fixed marginal rates of technical substitution, and quantities adjust to equilibrate supply and demand.

Our measure deals with an important subtlety in the data. Physical technology matrices are not observable; one cannot measure hours of labor per unit of Good 1 because of the way intermediate inputs are measured in input-output tables. Also, national accounts constrain empirical work in an important way. Only the value of output of Good 1 is observed; neither price nor quantity is observable separately. One can indeed record hours of labor used in the sector. Thus the ratio of hours of labor to the value of output is a canonical element in a unit-value matrix. The full employment condition for labor implies that one observes the physical input requirement per dollar of output, but *neither the physical technology nor the price of output is identified without an ancillary assumption.* Our identifying assumption is that every country's unit value technology matrix defines *unobservable* physical units by the *unobservable* world price. The definition of a unit of Good 1 is an international dollar's worth of it. This assumption implies that different unit-value matrices record disparate physical technologies. This assumption also lends a huge advantage; the unit value of every good at world prices is unity! Then the distance of a country from world prices can be computed from the least squares projection of *the unit vector* onto its local technology as measured by its unit-value matrix. Its distance from the world is the absolute value of the largest residual from this regression.

The distance between Canada and the United States is 0.66, the largest absolute value of the differences of the thirty-five estimated local unit costs in our data. This difference occurs in "Renting of machinery and equipment and other business activities". The smallest difference in estimated unit costs occurs in "Hotels and restaurants", where it is 0.02. Since world unit costs are one for every good, each of these numbers has an interpretation as a percent; renting and leasing of equipment is estimated to cost 66% more in Canada than in the United States.

Since we are constructing a measure of distance, we must use the largest of the absolute values of these estimated cost differences. This measure is conservative because it has to satisfy the triangle inequality. The average absolute value of the bilateral cost differences is only 21% for the United States and Canada. We interpret the maximal cost difference as a lower bound for a uniform bilateral prohibitive tariff; if NAFTA were abrogated and the United States put a 66% tariff or higher Canada retaliated likewise, then all bilateral trade would shut down. We think of our measure as capturing generalized economic distance, since some goods are traded and others are not. The first sixteen goods recorded in our data are traded sectors, and we also apply our measure to those sixteen sectors only. The bilateral distance based on traded goods only is 0.41 because Canada has a 41% cost advantage in "Wood and products of wood and cork".

Our main contribution is to define this economic distance and to bring it to the data. We define 741 bilateral economic distances among 39 countries. Using a technique from graph theory, we show that a network consisting of the 10% of closest links has a large connected component that breaks into two clusters. The first contains most of the advanced industrial countries, and the second has most of the emerging economies; the bridge between the two is the link between the Czech Republic and Poland. China, India, and Turkey are all isolated components. When we restrict our attention to the distance based on traded goods only, we see that the network described by the closest 10% of links has a single component that amalgamates the two former clusters and includes China. India, Russia, and Mexico are now isolated components. It seems that NAFTA has not equalized traded goods prices between Mexico and the United States; otherwise, this edge would be among the 10% of closest links.

Our distance measures are only as good as our projected unit costs. We examine these in detail in three ways. First, we corroborate that our costs are significantly correlated with those of other scholars who have worked with these data in a very different way. Second, we explore the real exchange rates that are implied by our projected unit costs, and show that rich countries have a high real cost of non-traded goods. Finally, we explore innovative measures of revealed comparative advantage for China, Mexico, and the United States. China has strongest revealed comparative advantage in "Basic and fabricated metals" and

strong comparative disadvantage in "Agriculture, forestry, and fisheries".

We hope that this brief summary of our empirical results has whetted your appetite for our theory and its applications. The rest of this paper is structured as follows.

The second section gives a review of the literature on trade costs and some discussion of the recent work in empirical trade that draws out the effects of differences in technology. The third section is the heart of our theoretical contribution; we develop our pseudo-metric and give some examples that flesh out one's economic intuition. In that section, we emphasize that one can measure physical inputs per dollar of output, but neither can be measured separately in the data; we make the identifying assumption each country's unit values are defined in terms of an unobservable vector of international prices. The fourth section gives a brief description of our data; we are very lucky to have the World Input Output Database's internationally consistent and detailed data on factor uses in thirty-five sectors in thirty-nine countries. Our measure of distance was designed precisely to take full advantage of these data.

The fifth section brings our measure to the data in four different ways. First, we show all the bilateral distances among our sample of thirty-nine countries as the graph of a network. We also give the network's minimum cost spanning tree; it is suggestive of a taxonomy of the world economy. Since our distance measures are only as good as our price projections, we then show that these are reasonable. We show that they are statistically significantly correlated with prices computed by other scholars who have used these data. Then we show that they give rise to reasonable measures of the real exchange rate, measured as the relative price of non-traded goods. Finally, we present some interesting new measures of revealed comparative advantage that arise from our price projections. The sixth section gives our conclusions and some suggestions for future research.

2   Review of the literature

If all goods are traded, then our distance measure is the uniform *ad valorem* equivalent cost that shuts down bilateral trade. It is now widely recognized that trade costs are large and consequently have a significant impact on international trade flows and economic welfare. It is also well accepted that trade costs are comprised in only small part by direct policy measures such as tariffs and the tariff equivalents of quotas (Anderson and van Wincoop, 2004), and that other costs of selling goods in foreign markets, such as transportation and freight, time costs, information costs, regulatory costs, and local distribution costs, are much more significant barriers to international trade. Indeed, of Anderson and van Wincoop's famous headline estimate of 170% average developed economy trade costs, only 8% reflects the cost of direct tariff and non-tariff barriers. The significance of broadly defined international trade costs is well-recognized in the policy world also, as the emphasis on trade facilitation in the Bali Package which resulted from the Ninth Ministerial Conference of the WTO in late 2013 indicates.

While the economic importance of trade costs is not in dispute, the measurement of trade costs remains difficult. There are two broad approaches that have been adopted in the literature. The first uses direct measures to construct estimates of the various components of trade costs; in principle, they can then be aggregated into a total measure. The second approach is indirect, using data on traded quantities and prices to infer international trade costs; Chen and Novy (2012) are a good example. As Anderson and van Wincoop (2004) note in their extensive survey of the early literature, the former approach is plagued by data inadequacies, while the latter inevitably involves the use of economic theory.

Tariff barriers are the easiest component of trade costs on which to obtain direct measures of incidence. Applied tariff barriers are available through UNCTAD's TRAINS database, and bound rates are accessible through the WTO's Consolidated Tariff Schedules. Information on the prevalence–but not the impact–of non-tariff barriers is also available through TRAINS. However, the data are far from complete, especially for developing economies. Although the MAcMap dataset does provide cross-sectionally complete applied

tariff data, it does so for limited years. An extensive literature has considered the problem of aggregating the data into economically meaningful measures, following Anderson and Neary (1996). An excellent recent discussion is Kee et al. (2009).

By contrast there is no uniform source of direct measures of transportation and freight costs, and these are not widely available, even from national sources. Because trade flow data are widely available at aggregated levels, through the IMF's DOTS, and at disaggregated ones, through the UN's COMTRADE, it is possible in principle to exploit the dual reporting of each flow in FOB from the export side and CIF from the importing side to impute transportation costs. For an example see the database constructed by CEPII and described in Gaulier et al. (2008). Unfortunately, by comparing the measures obtained in this manner with directly measured freight costs for two countries where such data are available (the United States and New Zealand), Hummels and Lugovskyy (2006) find that the constructed measures are badly error-ridden and contain no useful time-series or cross-commodity variation. Hummels (2007) reviews the available sources of data on direct freight costs, and documents the changes in the cost of international transport over time.

Comprehensive data on other components of trade costs are even more difficult to obtain. While some authors like Burstein et al. (2003) and Bradford (2005) construct local distribution costs from input-output data, it is common to use proxy data to construct indices for the components such as regulations, standards, and customs procedures that are under study. For example, one could uses counts of the average number of days that are needed for a good to cross the border or frequency counts and coverage ratios of prevailing WT0 standards that require notification. Chen and Novy (2012) summarize data availability, along with recent studies, in the area of standards and regulation. These proxies are then typically used as covariates in a gravity model, along with other variables such as distance, border and FTA controls, and common language. Making some assumptions about a trade cost function and using knowledge of the model's elasticity of substitution,one can back out *ad valorem* tax equivalents of the relative impact of a particular component of trade cost. Anderson and van Wincoop (2004) give the details.

Using this approach, Chen and Mattoo (2008) considers the role of standards and regulations. Hummels and Schaur (2013) use data on air and maritime transportation of imports to the US to estimate the significance of time as a trade cost, finding that each day in transit is equivalent to an *ad valorem* tariff of between 0.6 and 2.3 percent. In addition to data availability issues, weaknesses of this general approach include the arbitrariness of the assumed trade cost function, and the potential for omitted variable bias, given the unobserved nature of many aspects of trade costs.

Given the difficulties inherent in obtaining direct measures of trade cost components and then converting them into a usable measure of incidence, a number of researchers, beginning with Head and Ries (2001), have turned to indirect measures. In essence, this approach turns gravity on its head, inferring trade costs from the trade data without specifying a trade cost function.[2] It does so using a neat algebraic trick. The gravity equation is solved for the unobservable trade costs as a function of bilateral trade flows and the multilateral resistance variables. The latter are unobservable, but can be canceled out using trade flows in the opposite direction, and domestic trade flows in each trading partner. Given knowledge of the elasticity of substitution, it is possible to back out *ad valorem* tax equivalents. Chen and Novy (2012) and Novy (2013) provide the details.

The indirect technique can be applied to a much wider range of countries and time periods, but it has some disadvantages. By its nature, the technique yields aggregate trade costs, not information on any particular component, and can only determine the geometric average of bilateral trade costs between any country pair. Moreover, because it is based on a calibration of trade flow data, any measurement error is passed through. Recent applications include Chen and Novy (2011), who introduce a correction to the measure used by Head and Ries (2001) for heterogeneity across industries by using industry-specific substitution elasticity estimates. Jacks et al. (2011) construct the measure over a long time span, and show that it has

---

[2]In this sense, the technique is closely related to the earlier trade potentials literature.

significant power in explaining changes in trade flows. Novy (2013) shows that the indirect trade cost measure is consistent with a wide variety of underlying trade models. The technique has also recently been used to construct a new World Bank database on trade costs for a large group of developed and developing economies (Arvis et al., 2013).

The popularity of the gravity equation is due in part to the perceived empirical limitations of the Heckscher-Ohlin-Vanek (HOV) theory, although these two perspectives are not mutually exclusive. In the HOV framework, persistent differences in prices of the same good in different countries is in itself indirect evidence of high trade costs. To the extent that these trade costs are proportional to geographical distance and size, the gravity model of trade is empirically verified. See Deardorff, 1998 for an example.

The HOV model in its simplest textbook form has been discredited by a lengthy literature. The main reason for the model's failings has to do with the lack of a common technology. In their study of ten OECD economies using data from 1985, Davis and Weinstein (2001) incorporate a gravity specification to the demand side of the HOV model and show that it improves factor content predictions. However, this specification is their seventh alteration to the standard model in a meager sample of countries, and on the margin it has a relatively small impact on several of the empirical tests they report. Measuring endowments using the value of factor services, Fisher and Marshall (2013) show strong support for the HOV model. The value of factor services predicts the factor content of trade fares well because these measures already incorporate technical differences in the local factor prices inherent in them.

On the other hand, the evidence against a common technology, especially among countries at different stages of development, is overwhelming. Recent studies, such as Maskus and Nishioka (2009) and Marshall (2012), consider how these differences between developed and developing countries explain the failure of the HOV predictions. Once differences in technology are acknowledged, a thorny question arises on how to best adjust HOV factor content predictions. We rely on an underlying equilibrating force in the HOV framework: in the long-run countries' technologies adjust to local factor market conditions to make them as competitive as possible on world markets. If there are differences in local unit costs for the same good in two different countries, then these must reflect generalized trade costs.

Our bilateral distance measures give rise to a complex network indicating Heckscher-Ohlin similarity. This is exactly what one would expect in a world where trade costs matter. In response to the growing evidence on the significance of trade costs, Deardorff (2014) gives the first thorough theoretical treatment of the implications of these costs for some measures of comparative advantage. In an elegant and parsimonious specification based on bilateral trade costs, he demonstrates that in both the Ricardian and Heckscher-Ohlin models simple production costs are no longer the key determinants of the pattern and volume of trade.

3   The theory

This section will describe the ideas underlying our measure of economic distance. Our theory is grounded in what can be observed in the data. Since our measure is novel, we will also include a didactic exposition of increasingly realistic examples. The bottom line of our measure is this: if two countries have local unit costs that are identical for every market in which they are active, the bilateral economic distance between them is zero.

This section consists of four subsections. First, we describe what can be measured in the data; we make a sharp distinction between physical technology matrices and unit-value matrices. Second, we argue that the only consistent measure of units for an international economist is an (unobservable) international dollar's worth of output. Third, we define our pseudo-metric; this subsection is the crux of our theoretical contribution. Fourth, we give a series of illustrative examples.

3.1 Unit values in the closed economy

Consider an $n \times f$ matrix of direct and indirect unit input requirements:

$$A(w)$$

where w is the $f \times 1$ vector of local factor prices. Its canonical element

$$a_{ij}(w)$$

is the direct and indirect input requirement of factor j per unit of output of sector i. These are physical units, such as hours of unskilled labor per kilograms of apples.

Under the assumption of constant returns to scale and no joint production, this matrix is a complete description of the supply side of an economy. It is unfortunate for empirical researchers $A(w)$ is *not* observable. There is a very good practical reason for this fact: input-output data are recorded as flows of dollars, and the only natural definition of of a unit of good i is actually a dollar's worth of that good. Almost every empiricist who works with these matrices actually observes a point on the *unit-value* isoquant, not the unit-quantity isoquant.

For the closed economy, this point is moot. It amounts to rescaling the rows of the matrix $A(w)$. If both factor prices w and the physical technology matrix $A(w)$ were observable, the unit-value isoquants could be constructed from these data. Local unit costs are the $n \times 1$ vector

$$p = A(w)w.$$

Write $P = diag(p)$. Then the *observable* unit-value matrix is:

$$V(w) = P^{-1}A(w)$$

This unit-value matrix still depends upon factor prices w, and it is routinely calculated in many empirical applications in international trade and development economics. The units of the canonical element $v_{ij}(w)$ are now physical inputs per dollar's worth of apples.

For the moment, let us fix factor prices and drop the dependence of the observable $V(.)$ on w. The unit-value matrix satisfies an elegant property:

$$1_{n \times 1} = Vw.$$

Multiplying this matrix on the right by factor prices yields a vector of ones (by definition of course). In other words, as long as one defines physical units exactly according to local units costs $p = A(w)w$, then the unit vector is in the column space of V. Since V is observable, so is its column space. The consistency of the local input-output matrix can be checked, if one is willing to maintain the ancillary assumption of homogeneous factors that are mobile between sectors.[3]

This approach to input-output accounting has an added bonus. If local factor prices are not observable, then they can be calculated using the Moore-Penrose pseudo-inverse:

$$w = V^+ 1_{n \times 1} + (I - V^+ V)z \tag{1}$$

where $z \in \mathbb{R}^f$ is arbitrary. Equation (1) actually gives the set of all factor prices that are consistent with a given unit-value matrix v. It is empirically significant, since both V and V^+ are observable, whereas factor prices w may not be.

---

[3]For example, a simple unit value matrix describing a Ricardian economy where it takes one unit of labor to make a dollar's worth of the first good and two units of labor to make a dollar's worth of the second good is inconsistent because the unit vector $1_{2 \times 1}$ is not in that column space.

Let w be any solution to this equation. Then

$$1_{n \times 1} = Vw = VV^+ 1_{n \times 1} \tag{2}$$

where we have used the property of any generalized inverse that $V = VV^+V$. The $n \times n$ projection matrix

$$VV^+$$

is crucial to our analysis. It is symmetric by definition of the Moore-Penrose pseudo-inverse, and it takes any vector in $\mathbb{R}^n$ and projects it into the column space of V. In fact, it is the least squares projection. For any vector $p \in \mathbb{R}^n$

$$\hat{p} = VV^+p$$

is the vector in the column space of V that is closest in the Euclidean norm to p.

If local unit costs $p = A(w)w$ are measured correctly, then the unit-value matrix V is consistent. In this important benchmark, then (2) states that $1_{n \times 1}$ is an eigenvector of VV^+ corresponding to the eigenvalue 1. This matrix is a projection of course, so its eigenvalues are either zero or one. In the empirically important case where $n > f$ and V has full rank, this projection matrix is singular and has $n - f$ eigenvalues that are zero.

3.2 Measuring unit values consistently across countries

We have emphasized how important it is to define physical units properly, but there is no simple way to do so when local unit costs for identical commodities differ across countries. In international economics, an important issue of consistency of measurement arises.

Now let

$$A_i(w_i)$$

be the *unobservable* physical technology matrix for country i evaluated at local factor prices w_i. In this paper, we will be completely general and assume that the mappings $A_i(\cdot)$ may differ from $A_j(\cdot)$ and also that factor prices w_i need not equal w_j. Now this question occurs: *What is the only way to measure physical units consistently across countries?* The answer is that we must define unit value matrices using international prices, not local prices!

The empirical trade economist works with a unit-value matrix:

$$V_i(w_i) = P^{-1}A_i(w_i) \tag{3}$$

where the diagonal matrix P records *world prices*, not local prices, for each of the n commodities. At first blush, it may seem that we are being picayune in emphasizing this fact, but (3) has implications that are central to our analysis. They may have been overlooked in the literature. We are simply emphasizing the following fact: a kilogram of physically homogenous apples is identical in France or Thailand, but a euro's worth of apples is quite different from a baht's worth. Also, it may be cheaper to produce apples locally in Nancy than in Nan, even when one adjusts their marginal costs using a common currency. If one wants to define physically identical units, then the unit isoquants must be measured in every country the same way. So if the world price of a kilogram of apples is $2, the the scaled units in both Paris and Bangkok must be 500 grams of apples. The local prices of apples may differ from world prices, but consistency in measurement compels us to define physical units identically. *The natural scaling factor is the common world price.*

Perhaps there are lingering doubts about what is measured in a unit-value matrix derived from local national accounts. Its canonical element has units such as hours of unskilled labor per dollar's worth of an apple. As we have explained, the physical unit input coefficient (hours of unskilled labor per kilogram of apple) cannot possibly be measured because of the way input-output accounts are computed. Still, the

number of hours of unskilled labor used in the apple sector in a year are recorded. So is the gross dollar sales of apples in a given year. One can observe price times quantity, but neither is identified alone. Hence, it is impossible to observe the local price of an apple either. In essence, only the ratio of hours of unskilled labor per dollar's worth of apples is identified, and neither the numerator nor the denominator of this fraction can be determined independently.

The trade economist, who is concerned with making comparisons across countries, must make an identifying assumption. The relevant postulate depends upon the theory being examined. If one follows a simple version of Heckscher-Ohlin theory and assumes that the (unobservable) physical coefficients are equal in France and Thailand, then one can identify a ratio of local prices, and one has a part of a theory of international factor prices. If one takes a strong stand on the law of one price and assumes than the (unobservable) local prices in France and Thailand are identical, one has part of a theory of technological differences across countries. Since we place great emphasis on measuring physically homogeneous quantities of apples consistently in all countries, we impose the auxiliary assumption that the unit-value matrices in every country are measured using *unobservable* world prices. Hence we attribute differences in these matrices to international differences in technologies.

3.3 A pseudo-metric that defines economic distance

Let $\mathcal{V}$ denote the set of $n \times f$ unit value matrices. We can now define a pseudo-metric on $\mathcal{V}$ that will be at the heart of our analysis. For V and V' both in $\mathcal{V}$

$$d(V, V') \doteq \|(VV^+ - V'V'^+)1_{n \times 1}\|_{\infty} \tag{4}$$

This distance function projects the unit vector onto the column spaces of V and V'; then it chooses the largest absolute value of the difference between the two projections. This distance has an interpretation as a prohibitive bilateral tariff, a point that we will illustrate below.

It satisfies the four properties of a pseudo-metric. First, it is non-negative

$$d(V, V') \geq 0.$$

Second,

$$d(V, V) = 0,$$

although it is possible that $d(V, V') = 0$ for some $V' \neq V$. The fact that we have defined a pseudo-metric and not a metric has an economic intuition. Two unit value matrices are in the same equivalence class if local factor prices can adjust sufficiently so that each economy is competitive for bilateral trade in every sector. Third, this pseudo-metric is obviously symmetric:

$$d(V, V') = d(V', V).$$

Fourth, it satisfies the triangle inequality. If V, V', and V'' are all in $\mathcal{V}$, then

$$d(V, V'') \leq d(V, V') + d(V', V'')$$

by the properties of the row sums of any matrix and the fact that the infinity norm on vectors automatically satisfies the triangle inequality.

Write

$$p_i = V_i V_i^+ 1_{n \times 1}$$

These are the local unit costs in country i that are nearest in the Euclidean norm to world unit costs. The economic distance between the country i and country j is

$$d(V_i, V_j) = \|p_i - p_j\|_{\infty} = max_k |p_i(k) - p_j(k)| \tag{5}$$

where $p_i(k)$ is the unit cost of good k in country i. This equation shows that the economic distance between two countries is the smallest possible uniform bilateral tariff that would allow for all sectors to be active in both countries. Write $t_{ij} = d(V_i, V_j)$. Then

$$p_j(k) - t_{ij} \leq p_i(k) \leq p_j(k) + t_{ij}$$

for every good k. Hence if uniform (bilateral) iceberg trade costs or uniform bilateral tariffs were at least as large as t_{ij}, then the differences in unit costs between countries i and j are sufficiently narrow so that all bilateral trade is cut off because of high trade costs.

Here is an important advantage of our definition: we have a natural numéraire. Since each $|p_i(k) - p_j(k)|$ is a difference in local costs, it is akin to a real specific tariff. As such, it would not be comparable across goods. But we have defined our distance as the projection of the unit vector, assuming that each good has world unit cost of unity. Hence each element of $p_i = V_i V_i^+ 1_{n \times 1}$ has an interpretation as a relative cost. Each element of p_i is then a scalar and it has no units. This is exactly the way that an international economist thinks of an *ad valorem* tariff, and it is we why chose to define our economic distance in this simple way.

Here is what this pseudo-metric does not imply: if the economic distance between two countries is zero, then it is not true that they necessarily have the same physical technology. For example, if one country has workers that are one-tenth as productive as those in a trading partner, then its unit-value matrix will use ten times more workers in every sector. But its local wage will be ten times less, and it will have equal unit costs with its trading partner, no matter what the world prices are. For any diagonal matrix Λ and unit-value matrix V, both having full rank, the projection matrices VV^+ and $V\Lambda(V\Lambda)^+$ are identical.[4] Our measure of economic distance is quite appropriate in this case because both countries are equally competitive on world markets. Factor prices adjust exactly correctly for factor-specific technical differences, and that is why tests of Heckscher-Ohlin theory define factors in efficiency units, once they move away from the assumption of identical technologies in every country.

This pseudo-metric allows for the most representative case in the real world economy: countries have very different physical technologies and also very different factor prices. But any two countries that are equally competitive in every market have no economic distance between them. Any country that can compete with world prices in every market has a technology that is representative of the world. Its economic distance from the world would be zero because $p_i = V_i V_i^+ 1_{n \times 1} = 1_{n \times 1}$. We consider this state of affairs prototypical because Heckscher-Ohlin theory is concerned with the long run. One would expect local production techniques to adjust to long-run differences in factor prices so that local goods prices are as competitive with world prices as possible.

3.4 Some illustrative examples

This subsection develops a series of illustrative examples. Our measure of economic distance is new, and we hope that these examples will show its value. In each example, we describe an *unobservable* physical technology matrix assume a list of *unobservable* world prices, and show how our measure works.

A simple Ricardian economy with two goods and one factor allows us to depict our pseudo-metric exactly. Figure 1 shows the geometry of our projection. The first country has $A_1 = (2, 1)^T$, the second country has $A_2 = (1, 3)^T$, and world prices are $p = (1, 1)^T$. Hence $V_1 = (2, 1)^T$, and $V_2 = (1, 3)^T$. Projected unit costs in the first country are $p_1 = V_1 V_1^+ 1_{2 \times 1} = (1.2, 0.6)^T$; likewise, $p_2 = V_2 V_2^+ 1_{2 \times 1} = (0.4, 1.2)^T$. The distance between the two countries is $d(V_1, V_2) = max\{|1.2 - 0.4|, |0.6 - 1.2|\} = 0.8$.

These local costs have this interpretation: the predicted wage in the first country is $w_1 = V_1^+ 1_{2 \times 1} = 0.6$, and that in the second country is $w_2 = V_2^+ 1_{2 \times 1} = 0.4$. These wages allow both countries to be as competitive

---

[4]In this case, the Moore-Penrose inverses have explicit formulae: $V^+ = (V^T V)^{-1} V^T$ and $(V\Lambda)^+ = ((V\Lambda)^T V\Lambda)^{-1} (V\Lambda)^T$, and the result follows from simple matrix algebra. The intuition is that the least squares projection is independent of the scale of any one explanatory variable.

as possible on world markets in *both goods*. The first country is 0.4 from world markets, while the second is 0.6 from them. An 80% bilateral tariff would prohibit trade between these two countries, and a 40% iceberg transport cost would isolate the first country from the world, while a 60% iceberg costs would isolate the second country from the world. Even though the first country is two times less productive than the world in the first good, a 40% transport cost would allow it to produce both goods locally. In particular, it could not use its low wage to specialize and export the second good to the world.

Consider the classic 2×2 case in Heckscher-Ohlin theory with a slight twist: the countries have different physical technologies. Let

$$A_1(\cdot) = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$$

and

$$A_2(\cdot) = \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}.$$

The rows list goods, and the columns denote capital and labor inputs respectively. World prices are $p = (2,3)^T$. Hence, the observable unit-value matrices are:

$$V_1(\cdot) = \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \end{bmatrix}$$

and

$$V_2(\cdot) = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 1/3 \end{bmatrix}.$$

$V_1 V_1^+ = I_2$ and also $V_2 V_2^+ = I_2$, since both unit-value matrices have full rank. Thus $d(V_1, V_2) = 0$, and the slightest transportation cost will shut down trade. Both countries are competitive in both goods because factor prices have adjusted to long-run differences in technology; indeed $w_1 = (1,1)^T$ and $w_2 = (0.5, 1, 5)^T$ are the only factor prices that would allow for the input-output accounts to record positive local outputs in every sector in both countries. Indeed, in any pair of 2×2 economies, both having full rank, there is no economic distance between the pair precisely because local factor prices will adjust fully! [5]

Now consider an economy with three goods and two factors. The physical technology matrices are identical:

$$A_1(\cdot) = A_2(\cdot) = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}.$$

Again, the first column is inputs of capital and the second is inputs of labor. The sectors are listed in order of increasing capital intensity. If world prices are

$$p = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix},$$

then the identical *observable* unit-value matrices are

$$V_1(\cdot) = V_2(\cdot) = \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 3/4 & 1/4 \end{bmatrix}.$$

---

[5]In the interesting case where a country has a unit-value matrix of deficient rank, there may be distance between these countries. Thus if national income accounts artificially create a 2×2 technology matrix by simple repeating the national capital-labor ratio twice in each row, our pseudo-metric would depend upon that country's capital-labor ratio.
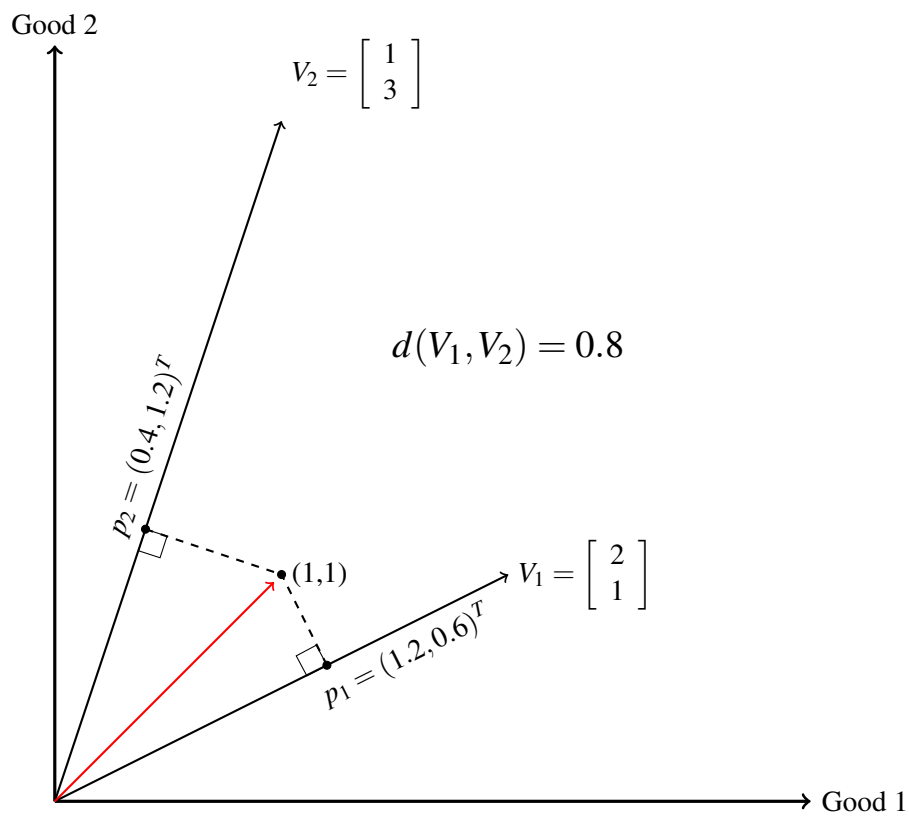
Good 2

$V_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

$P_2 = (0.4, 1.2)^T$

$d(V_1, V_2) = 0.8$

$(1,1)$

$V_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$p_1 = (1.2, 0.6)^T$

Good 1

Figure 1: Projecting world prices onto local unit values

Local factor prices are $w_1 = w_2 = (1,1)^T$. The countries have identical technologies, factor prices are equalized, and there are no impediments to trade. This is the canonical case in our textbooks. Since $V_1 = V_2$, $d(V_1, V_2) = 0$, and any uniform tariff greater than zero woulds block all bilateral trade. This is the well known paradox in Heckscher-Ohlin theory; when countries have identical technologies and equal factor prices, then the slightest transportation cost will stop all goods trade.

Consider the identical situation, but now *exogenous* world prices are:

$$p = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix},$$

Since these world prices are not in the column space of either country's technology matrix, local costs cannot possibly be equal to world prices. (Still, both countries will have the same local costs!) The identical *observable* unit-value matrices are now

$$V_1(\cdot) = V_2(\cdot) = \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 3/5 & 1/5 \end{bmatrix}.$$

The local unit costs that allow each economy to be as competitive on world markets as possible are

$$p_1 = p_2 = \begin{bmatrix} 0.97 \\ 1.09 \\ 0.92 \end{bmatrix}.$$

These costs were calculated using $V_i V_i^+ 1_{3\times1}$. The corresponding local factor prices are $w_i = V_i^+ 1_{3\times1} = (1.34, 0.6)^T$.[6] Again, since $V_1 = V_2$, the economic distance between theses countries is zero. This is almost the textbook model. The two countries have identical technologies, they have equal factor prices, and they produce goods at the same local costs. Both countries are equally isolated from the world market, and iceberg transportation costs of 10% (to ship to a distant world market) would cut off both from third parties. Any other minor bilateral trade impediment would force both economies into complete autarky.

Now consider

$$A_1(\cdot) = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$$

and

$$A_2(\cdot) = \begin{bmatrix} 2 & 10 \\ 4 & 10 \\ 6 & 10 \end{bmatrix}.$$

Let world terms of trade p be anything. Then $V_1(.) \neq V_2(.)$, but still $d(V_1, V_2) = 0$ because rents in the second country would be half those in the first and wages in the second country would be 10% of those in the first. In a neoclassical world with flexible factor prices, factor-specific technical differences are neither an impediment nor a boost to local competitiveness, a lesson we wish that policy makers might take closer to heart.

Now consider

$$A_1(\cdot) = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$$

---

[6]Since the world price of the capital intensive good has risen, it makes sense that the rent has risen and the wage has fallen. These are the Stolper-Samuelson effects in Fisher and Marshall (2011).

and
$$A_2(\cdot) = \begin{bmatrix} 1 & 1/2 \\ 2 & 0 \\ 2 & 1 \end{bmatrix}$$

Let world prices again be $p = (2,3,4)^T$. Then the unit-value matrices are:

$$V_1(\cdot) = \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 3/4 & 1/4 \end{bmatrix}$$

and

$$V_2(\cdot) = \begin{bmatrix} 1/2 & 1/4 \\ 2/3 & 0 \\ 1/2 & 1/4 \end{bmatrix}$$

Now the projection matrices themselves $V_1V_1^+$ and $V_2V_2^+$ are really quite different. But each one projects the unit vector onto itself: $V_1V_1^+1_{3\times1} = V_2V_2^+1_{3\times1} = 1_{3\times1}$. So $d(V_1,V_2) = 0$, and the countries have no economic distance between them. These two countries can be equally competitive and can match world prices in every market because local factor prices are different: $w_1 = (1,1)^T$ and $w_2 = (1.5,1)^T$. Again, these (possibly unobservable) factor prices can be computed directly using $w_i = V_i^+1_{3\times1}$.

We conclude this series of examples with the type of case that characterizes the data. It shows how the world really works. Let

$$A_1(\cdot) = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$$

and

$$A_2(\cdot) = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 5 & 1 \end{bmatrix}.$$

Let world prices again be $p = (2,3,5)^T$. These prices are chosen so that they lie in the column space of neither country's technology matrix, and the price of the third good is between the two local unit costs. Now the unit-value matrices are:

$$V_1(\cdot) = \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 3/5 & 1/5 \end{bmatrix}$$

and

$$V_2(\cdot) = \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 1 & 1/5 \end{bmatrix}.$$

Now $p_1 = (0.97, 1.09, 0.92)^T$, $p_2 = (1.03, 0.94, 1.02)^T$, and $d(V_1,V_2) = 0.15$. Local factor prices are $w_1 = (1.34, 0.6)^T$ and $w_2 = (0.77, 1.29)^T$. A 15% bilateral tariff would be prohibitive because it would make it impossible for the second country to sell its cheapest good in the first country, where it is quite expensive. The first country had a distance of 9% from the world (because its worst cost disadvantage occurs in the second good), and the second country has a distance of 6% (because its strongest cost advantage occurs in the second good too). In this situation, local factor prices have adjusted in the long run so that both countries are as competitive as possible in all goods, given that both economies remain in a neighborhood of the unobservable technologies $A_i(.)$. Even though the second country uses 67% more capital per unit of output in the third sector, factor prices adjust so that its economic distance from its neighbor is only 15%.

14

Direct and Indirect Inputs per Dollar in Australia

	High-skilled (hours)	Medium-skilled (hours)	Low-skilled (hours)	Capital (dollars)
Agriculture	0.0030	0.0102	0.0173	2.79
Mining and quarrying	0.0017	0.0047	0.0051	2.84
Food and beverages	0.0033	0.0099	0.0128	2.11
Textiles and textile products	0.0049	0.0157	0.0175	1.89
Leather and footwear	0.0037	0.0106	0.0127	3.42

Table 1: The first five rows of Australia's unit-value matrix

4 The data

Exploring the theory requires internationally consistent data on detailed factor uses for many industries across a diverse set of countries. We are fortunate that the recently released World Input-Output Database (WIOD) provides such data for 39 countries representing 80% of world GDP. Table A1 lists the countries in our sample, and it gives GDP per hour worked in each country. It includes most OECD countries and also major developing economies such as China, India, and Indonesia. A novel feature of this database is the combination of consistent input-output tables with the direct uses of three types of labor and physical capital in each of the 35 sectors. Table A2 lists the sectors in each country. Past empirical studies such as Bowen et al. (1987) had to cobble together input-output data and direct factor uses by sector from disparate sources. Our distance measure takes full advantage of this unique new consistent and highly detailed data source.

The social and economic accounts of the WIOD describe four factors that include three skill levels of labor, distinguished by level of education, and physical capital. High-skilled workers are those with some college-level education, medium-skilled workers are high school graduates, and low-skilled workers did not complete high school. In each sector, labor input is measured by total hours supplied by each skill level. Physical capital is computed by sector in a detailed fashion that distinguishes different asset classes and corresponding rates of depreciation. The 35 sectors in each country's input-output table encompass 14 distinct manufacturing activities and a wide range of other goods and services. All currency values in the original input-output tables are in millions of United States dollars. The extensive documentation provided by the WIOD includes the market exchange rates used to convert local currency values from each country's supply and use tables into US dollars. We only used these exchange rates to convert the physical capital stock, reported in local currency units, into US dollar values.[7]

To construct unit-value input coefficients, we follow the convention in the literature of dividing total factor inputs in each sector by that sector's gross output, giving direct factor usage per dollar of output. Next, we construct direct and indirect factor inputs per dollar of output using the Leontief matrix. The matrix of intermediate inputs includes domestic production and imported inputs. Of course, our key identifying assumption is that the physical unit of a good is an international dollar's worth of output. This premise allows us to explain differences in local unit costs entirely as variations in technology.

We close this section with a brief look at the data for the first five sectors of the first country in our sample in Table 1. We hope this small taste gives a sense of the numbers that are at the heart of our empirical work. One-thousandth of an hour is 3.6 seconds. The first row shows that it takes about ten seconds of high-skilled labor, 36 seconds of medium-skilled labor, a minute of low-skilled labor, and $2.79 of capital to make one dollar's worth of Australian agricultural output. Local factor prices can be computed using the relation $w_i = V_i^+ 1_{35 \times 1}$. The high skilled wage is $38.65, the medium-skilled wage is $26.34, the low-skilled wage

---

[7]See Timmer (2012) for complete details.

is \$21.18, and the rent on capital is 12%.[8] The first five elements in our estimated local unit costs p_i show that Australia is 8% above world unit cost in agriculture, 36% below that in mining, 9% below that in food, 20% above that in textiles, and 10% above that in leather and footwear. These numbers are reassuring, and they may lend credence to our empirical results.

5 The results

This section applies our measure in four different ways. First, we present all the bilateral distance measures and show that their magnitudes confirm what others have found using completely different techniques. We also present our findings within the framework of network theory. Second, we verify that our projected local unit costs accord roughly with those of other scholars who have worked with these data. Third, we explore real exchange rates by distinguishing traded and non-traded goods prices. Fourth, we use our projected local costs to give some new measures of revealed comparative advantage for China, Mexico, and the United States.

5.1 Bilateral distances and networks

Figure 2 shows histograms of all the $741 = 38 \times 39/2$ bilateral distances. The top panel gives the general economic distances, and the bottom shows those based upon traded goods only. It is obvious that trade tends to equalize unit costs, and it is also apparent that countries' productive structures differ most widely in the non-traded sectors. The sample average is 1.02, its median is 1.11, and its standard deviation is 0.47. The bilateral trade distances based upon traded goods are generally smaller. The mean is 0.46, the median is 0.51, and the standard deviation is 0.22. The distribution of bilateral distances based upon traded goods is much more skewed.

It is perhaps most useful to depict these bilateral distances using graph theory. Consider a weighted network with 39 nodes, one for each country. Since every bilateral distance is well defined, this network consists of one single completely connected component with the symmetric edge weights corresponding to the inverse of the bilateral distances.

Figure 3 shows the 74 closest links; these are the 10% of country pairs that show the highest degree of similarity in all their local unit costs. None of the links is greater than 0.54. The world economy consists of one large component that breaks into two clusters, and several isolated components, including China and India. It is interesting that the first cluster consists of 18 countries are mostly the advanced industrialized economies of Europe and North America. The second cluster contains recently industrialized economies (Japan and Korea) and those which are large exporters of primary products (Australia, Brazil, and Russia). The bridge between these two clusters is the link between the Czech Republic in the advanced group and Poland in the newly industrialized group. The countries in isolated components are sufficiently dissimilar from their nearest neighbor to indicate that their trade is mostly Ricardian and not due to the forces described in Heckscher-Ohlin theory. India, Indonesia, and China are in isolated components.

Figure 4 shows the 10% of closest links based upon traded goods only. If the law of one price held for every traded good, then all these bilateral distances would be zero, and the network would be one large densely connected component. Instead, we see that the closest 74 links have a bilateral distance smaller than 0.28. There is one large connected component, but several of the isolated components may seem anomalous. For example, why is the Netherlands not a part of the large component that includes most of the countries in the European Union? Please remember that a distance measure must, by its very nature, be conservative. The Netherlands has a very high estimated unit cost in agriculture, and this one deviation is enough to set it apart. Mexico likewise has a very low estimated unit cost in coke and refined petroleum. It is reassuring that
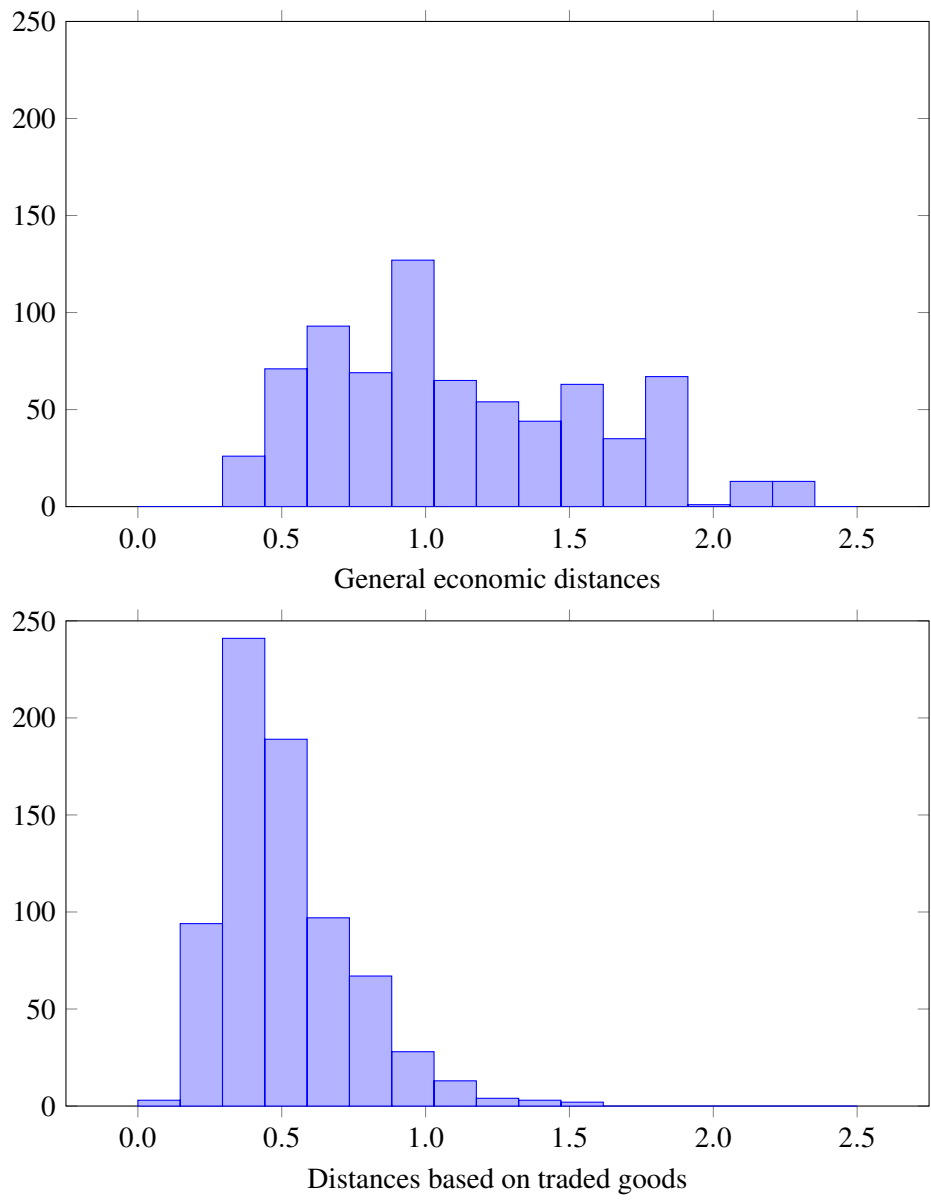
---

[8]Thus \$2.79 of capital rents for \$0.33.

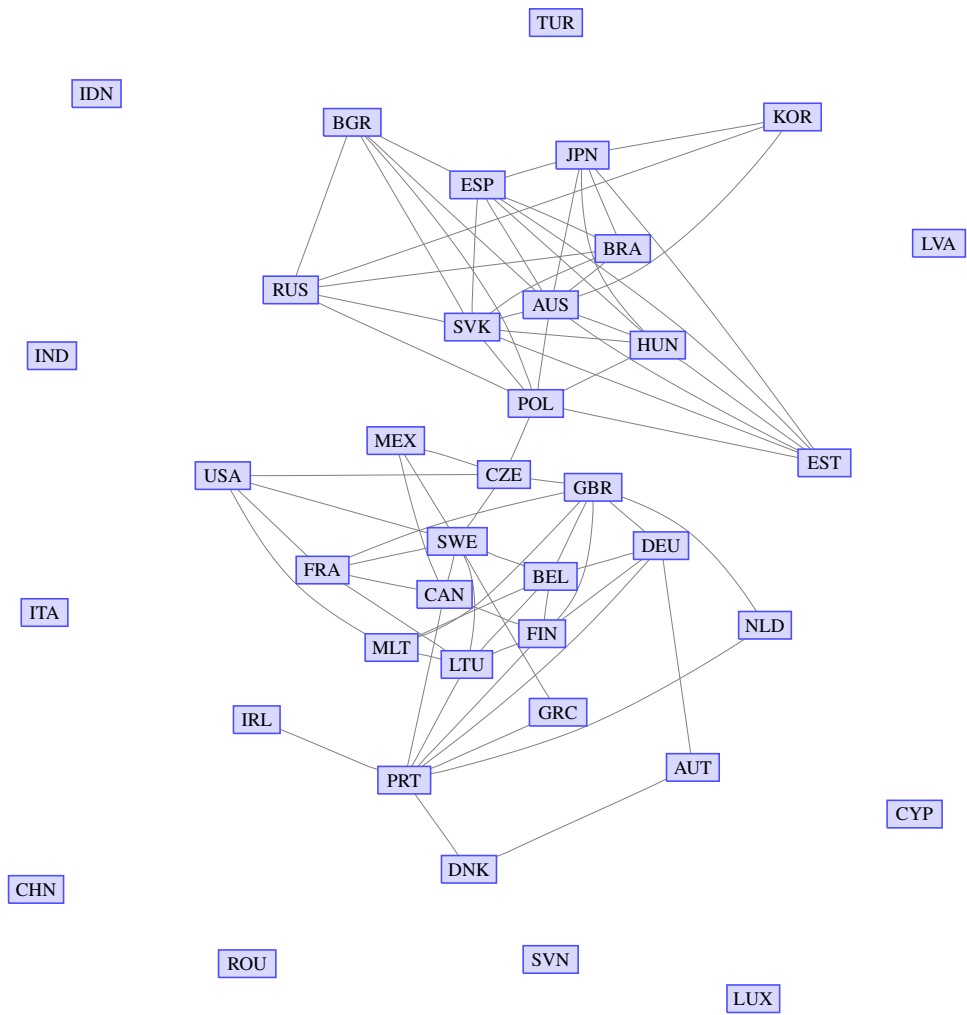Figure 2: Bilateral economic distances, as uniform *ad valorem* tariffs, $N = 741$

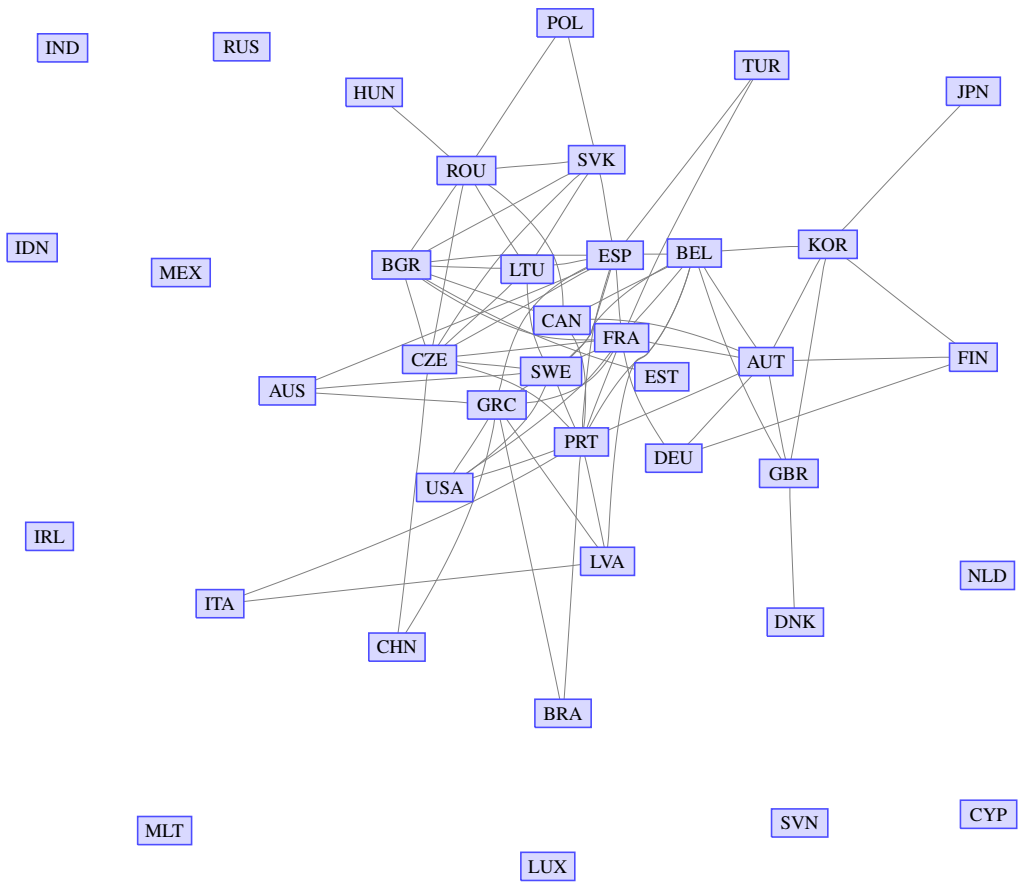Figure 3: Closest 10% of links in economic distance

Figure 4: Closest 10% of links in traded goods distance

China is a part of the large component because of the importance it plays in the world markets for traded goods.

Both of these graphs were based upon local network characteristics because they considered only bilateral similarities. We now turn our attention to a global characteristic of a network. A *tree* is an undirected graph in which any two vertices are connected by exactly one simple path. A completely connected network without any cycle is a tree. A *spanning tree* is one that connects all the vertices of a network. If one assigns a cost to each link, then a *minimum spanning tree* is the spanning tree that has the minimal cost among all trees that span the network. Such a tree is useful in engineering applications,[9] in taxonomy in the biological sciences, and in cluster analysis.

Figure 5 depicts the minimum spanning tree associated with the network based upon general economic distances. The lengths of the edges in the figure *do not* correspond to the economic distances; they are drawn for visual convenience. The nodes in this tree have degrees range from one to six. The diameter of this network is 14 because the longest path is from Denmark to Korea (among others). Slovakia has the highest degree, indicating that it is the most "typical" economy in some sense.

We are drawn to a taxonomic interpretation of this tree. Let Slovakia be the root of the "tree of Heckscher-Ohlin similarity." Six branches lead from this root. Four branches lead directly to idiosyncratic economies of Estonia, Brazil, Russia, and the dyad of Australia and China, a fifth goes through Hungary to Japan or Korea, and the sixth moves through the Czech Republic to most of the advanced European countries. The structures of the Canadian, Mexican, and American economies seem to be fairly disparate, when viewed from the perspective of a global network.

The minimum spanning tree for the network based only on traded goods Figure 6 has the same diameter of fourteen but now Portugal has the highest degree of five. The path from Portugal through Belgium leads to many of the advanced industrial economies, and the there seems to be a branch of the tree that links many of the former Soviet-bloc economies. Again, Canada, Mexico, and the United States do not seem to have similar unit costs for traded good, a fact that indicates that NAFTA, after its first decade,[10] may not have brought traded goods costs as close together as one might have hoped.

The network of general economic distance has very different properties from that based upon traded goods distance only. This fact is a strong indication that real exchange rates matter. The differences in unit costs for non-traded goods creates the greatest share of the dispersion of bilateral distances among these economies. We will return to this issue in greater depth.

5.2 How good are our projected unit costs?

Our distance measures are based upon unit value matrices that are purposely constructed to be consistent across a wide sample of countries. That these matrices are good measures of local technologies is not controversial; our projections of the unit vector on the local technologies may well be. How well do our vectors of unit costs $p_i = V_i V_i^+ 1_{n \times 1}$ actually match up with the data? These local unit costs are the foundation for our distance measure, and if they are wrong, then so is the crux of our work.

Our novel approach–using these simple projection matrices–also constitutes a computationally elegant technique for estimating unit costs. Of course, we utilize the World Input-Output Database extensively. It is important to compare our results with other international price comparisons.

The International Comparison Program (ICP) is an ongoing worldwide project, under the aegis of the World Bank and other national and international statistical offices, to estimate prices in 199 countries. Its primary focus is to construct internationally comparable measures of real GDP using purchasing power parity (PPP) indices based on prices of final goods. The ICP provides the most extensive database of international prices, but it focuses on expenditures and thus final goods prices. Hence, studies of the production side of

---

[9]The early mathematics in this area was based upon trying to connect the electrical grid among the cities of Czechoslovakia.

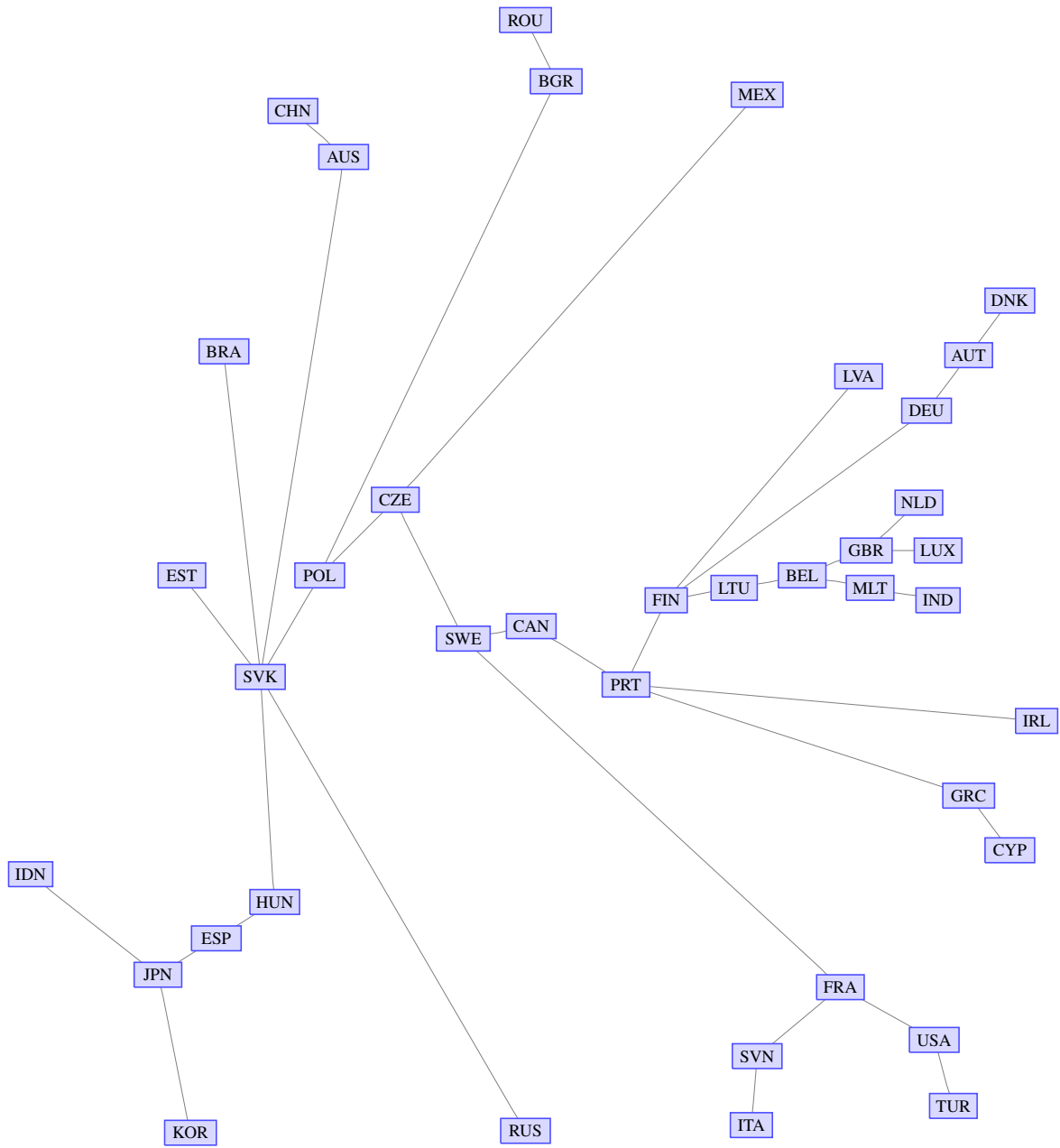[10]Our data are benchmarked in 2005; NAFTA began in 1994.

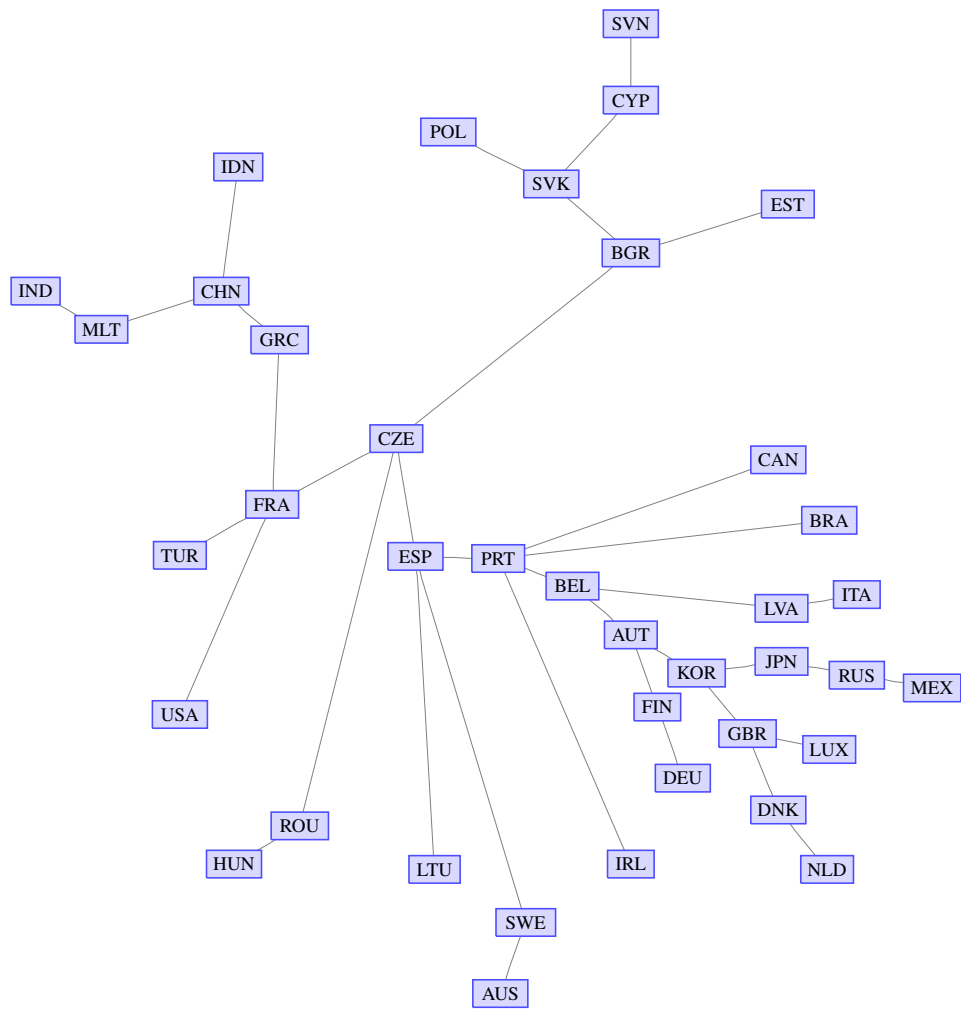Figure 5: Minimum spanning tree for the network based on economic distance

Figure 6: Minimum spanning tree for the network based on traded goods distance

GDP must make significant adjustments to the ICP price data. Using the WIOD production structure of thirty-five sectors in the benchmark year 2005, Inklaar and Timmer (2013) provide by far the most extensive effort to align the ICP data with producer prices. They provide the most appropriate comparison to our own unit cost estimates.

Inklaar and Timmer (2013) apply the Geary-Khamis (GK) method to constructing a real measure of local value-added and a concomitant PPP index for value-added in five major sectors. As part of this effort, they have published relative prices for the 35 WIOD sectors. The GK system constructs an international reference price for each good based on a system of simultaneous equations that involves local prices and quantities in all sample countries. The treatment in Inklaar and Timmer (2013) is very intricate, since they estimate separate prices for exports and imports and adjust the ICP price data to accord with base producer prices by sector. We will abstract from these details and focus on the essentials needed to compare their producer prices with our unit costs.

Here's how their technique works. First, all local prices are converted to U.S. dollars using market exchange rates. Second, Inklaar and Timmer's relative price for good k in country i is $p_i(k)/\pi_k$, where π_k is the international reference price for good k. Let $y_i(k)$ be the quantity of good k produced in country i, and let $z_i(k,l)$ represent the quantity of intermediate input l used in the production of good k in country i. Inklaar and Timmer's system of equations is:

$$\pi_k = \frac{\sum_i p_i(k) y_i(k)/P_i}{\sum_i y_i(k)}$$

and

$$P_i = \frac{\sum_k [p_i(k) y_i(k) - \sum_l p_i(k.l) z_i(k,l)]}{\sum_k [\pi_k y_i(k) - \sum_l \pi_l z_i(k,l)]}.$$

Each of the thirty-five reference prices in the first set of equations captures an international price for the relevant good. Each of the thirty-nine purchasing power parity indices described in the second set of equations captures a local producer price index based upon value added by sector.

We solved the GK system of equations using our own unit costs for each country $p_i = V_i V_i^+ 1_{n \times 1}$, and we used the WIOD input-output tables for the quantities of gross output and intermediate inputs in each sector. From the solution to this system, we find thirty-five world reference prices for the goods, and we determine thirty-eight free producer price indices, using the normalization that the U.S. index is unity.

Figure 7 presents the scatter plots of $p_i(k)/\pi_k$ for all sectors and countries[11] and also the local producer price indices P_i for the 39 countries. A regression of our calculations for $p_i(k)/\pi_k$ on theirs and a constant has an $R^2 = 0.08$ and an estimated slope of 0.06, with a standard error of 0.02. A regression of our calculated P_i on theirs and a constant has an $R^2 = 0.11$, and an estimated slope of 0.13 with a standard error of 0.06. Our simple projections, using absolutely no external price data, are weakly correlated with the computationally intricate ones that Inklaar and Timmer construct using idiosyncratic adjustments from the ICP project. These partial correlations are statistically significant.

We have imposed a theoretical framework on the data that constrains our estimates of unit costs. In contrast, price survey data offer no theoretical explanation for price differences, and they confront serious measurement problems such having to do with the quality of goods across international borders and the local prices of services. Our costs give rise to prices that are weakly but significantly partially correlated with theirs. Also, we will show strong support for the Balassa-Samuelson effect in the next subsection. This finding is in line with the the primary focus of Inklaar and Timmer's study. Taken as a whole, our theoretically grounded measures of unit costs are substantiated by Inklaar and Timmer's very different approach to international price comparisons.

---

[11]We excluded twenty-one data points because Inklaar and Timmer recorded no price in those cases; hence there are $N = 39*35 - 21 = 1344$ pieces of data displayed.
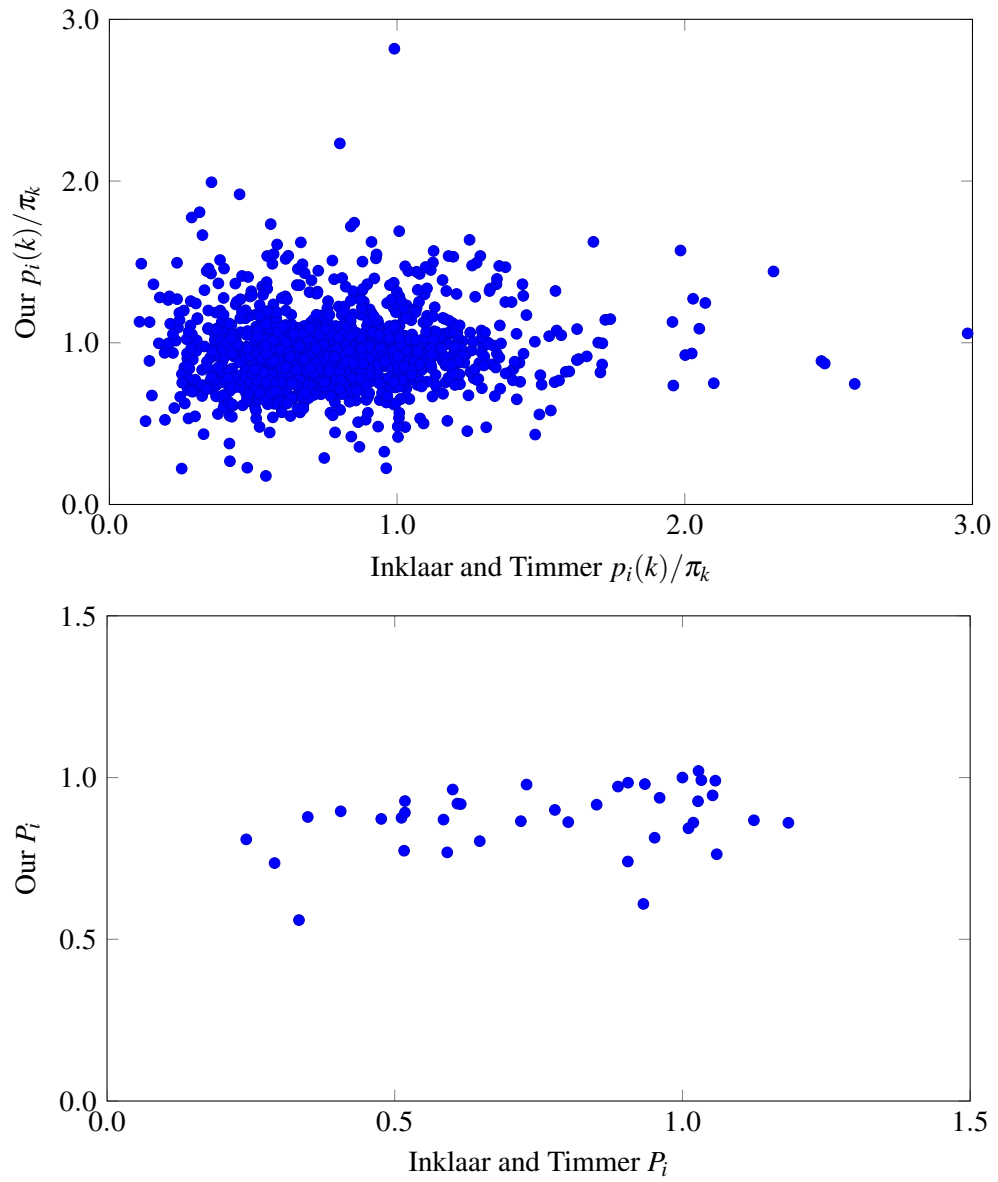
Figure 7: Calculations based on our unit costs versus those from Inklaar and Timmer

5.3 Balassa-Samuelson effects

We begin by exploring unit costs by country. Figure 8 gives the average unit cost across the thirty-five goods in each country. The horizontal axis shows dollars per hour worked in 2005. The diameter of the bubbles shows the standard deviation of each country's 35 estimated unit costs. The dispersion of unit costs are roughly comparable for each country, and some very poor countries have lower estimated unit costs. Rich and middle income countries show no discernible pattern of average unit costs. These averages aggregate the prices of all locally produced goods, and they may obscure important effects of the real exchange rate.

We have an ideal theory and excellent data for exploring the relative price of non-traded goods that is at the heart of the defintion of the real exchange rate (Salter, 1959, Balassa, 1964, and Samuelson, 1964). Balassa and Samuelson emphasized that faster productivity growth in traded goods than in services tends to make the relative price of services higher in rich countries.

The first sixteen goods in the WIOD data are traded, and the last nineteen are non-traded. We define the price of trade goods as the average unit costs of the first sixteen sectors and the price of non-traded goods as that average taken across the last nineteen. These prices are best understood as the average deviation from the unobservable unit cost of unity that characterizes every unit cost on world markets. Salter's definition of the real exchange rate then is this ratio. Figure 9 plots this ratio against GDP per hour worked. It is lower for countries with low levels of productivity, such as China, India and Turkey, than it is for Belgium, Denmark, or Ireland.

The source of the variation in the real exchange rates is the fact that unit costs for services are lower in poor countries. Figure 10 plots average price of traded goods against GDP per hour worked. Local unit costs in traded goods sectors show no correlation with aggregate productivity, as Balassa and Samuelson hypothesized. One may conclude tentatively that the results shown in Figure 8 occur because the estimated unit costs of non-traded goods in poor countries are low indeed.

An alternative way to evaluate the Balassa-Samuelson proposition is to construct an index of real exchange rate and see how it change with respect to changes in relative productivity, as in Rogoff (1996). We constructed an index of the real exchange rate:

$$\rho_i = \frac{p_i^{NT}/p_i^{T}}{p_{US}^{NT}/p_{US}^{T}}$$

where p_i^{NT} and p_i^{T} are unit-cost averages for the non-traded and traded goods in country i. We ran a simple regression of this index against relative productivity measured as the ratio of GDP per hour worked in country i relative to that in the United States. The estimated intercept was 1.007, the slope was 0.102, and the $R^2 = 0.195$. The estimated standard error for the slope is 0.034. The Balassa-Samuelson effect is statistically significant in our data. A \$10 increase in a country's GDP per hour worked is associated with a 1% appreciation of the real exchange rate. For example, the difference between the productivity of Romania and Luxembourg is about \$60 per hour worked; this estimate indicates that Luxembourg's relative price of non-traded goods is predicted to be 6% higher than Romania's. These results are another indication that our unit cost projections are not off base, and they give further confidence to our distance measure.

5.4 Revealed comparative advantage

There has recently been renewed interest in understanding the sources of comparative advantage. Chor (2010) is a good example. Measures of revealed comparative advantage are of considerable practical importance, and our work makes a unique contribution in this regard.

Standard measures of revealed comparative advantage are based on the observed trade flows, and they are usually constructed following Balassa (1965). This measure is has a numerator and a denominator. The top of the fraction is the ratio of the share of the value of exports in a given product category to the value of
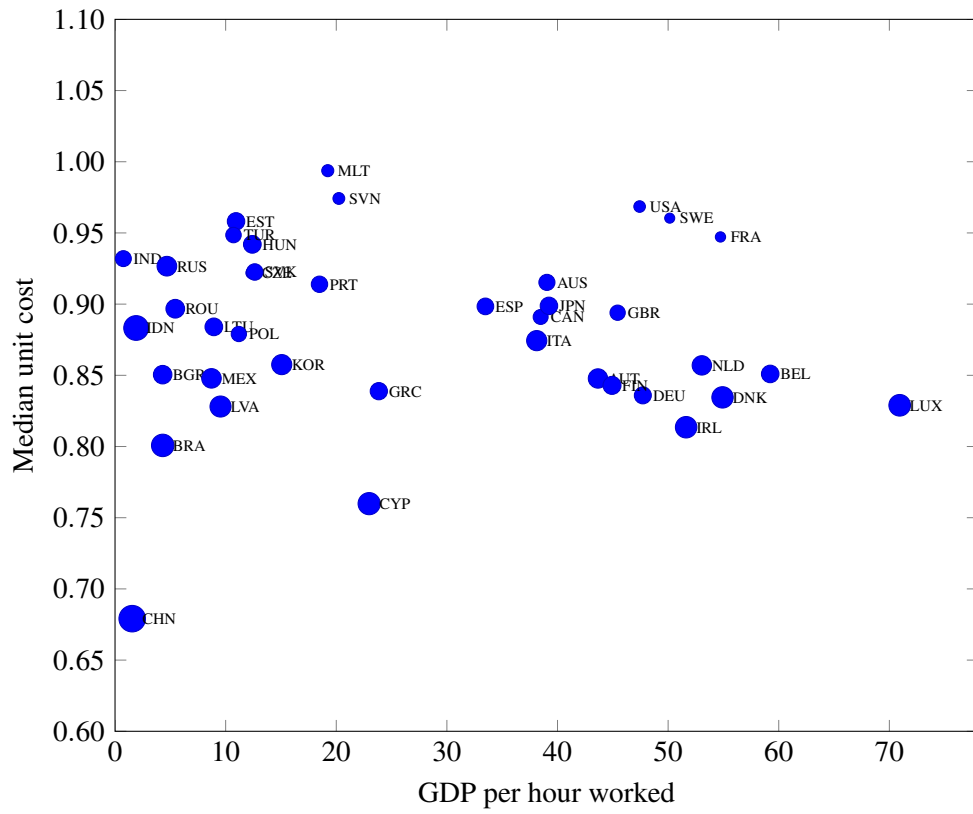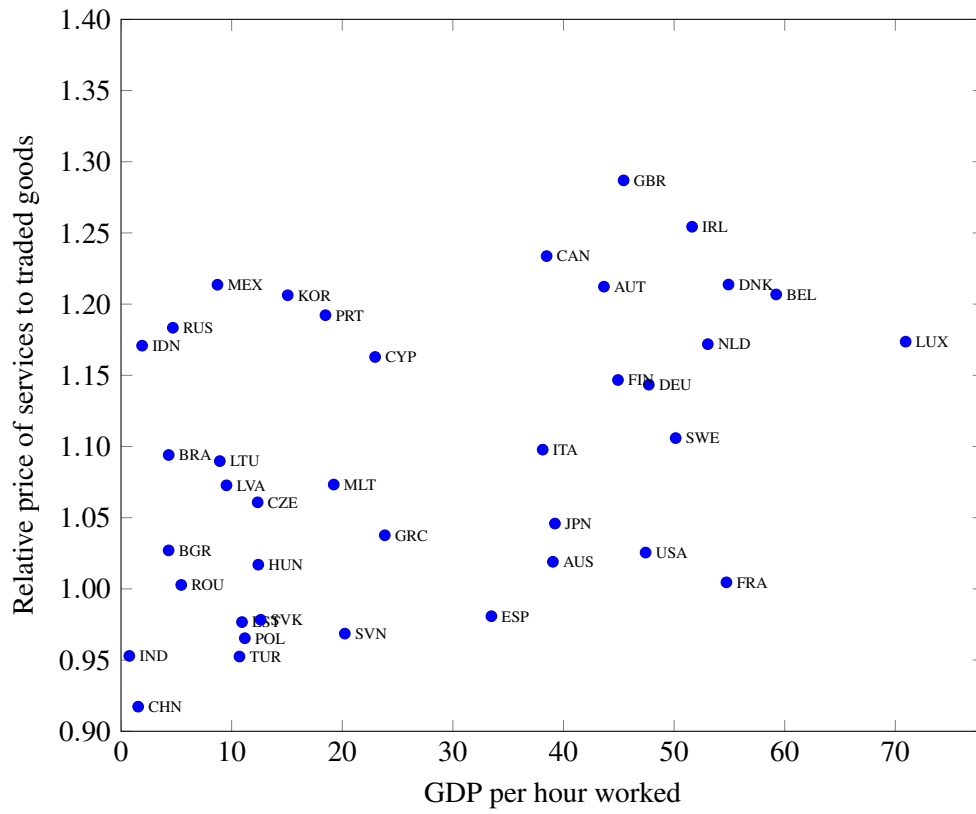
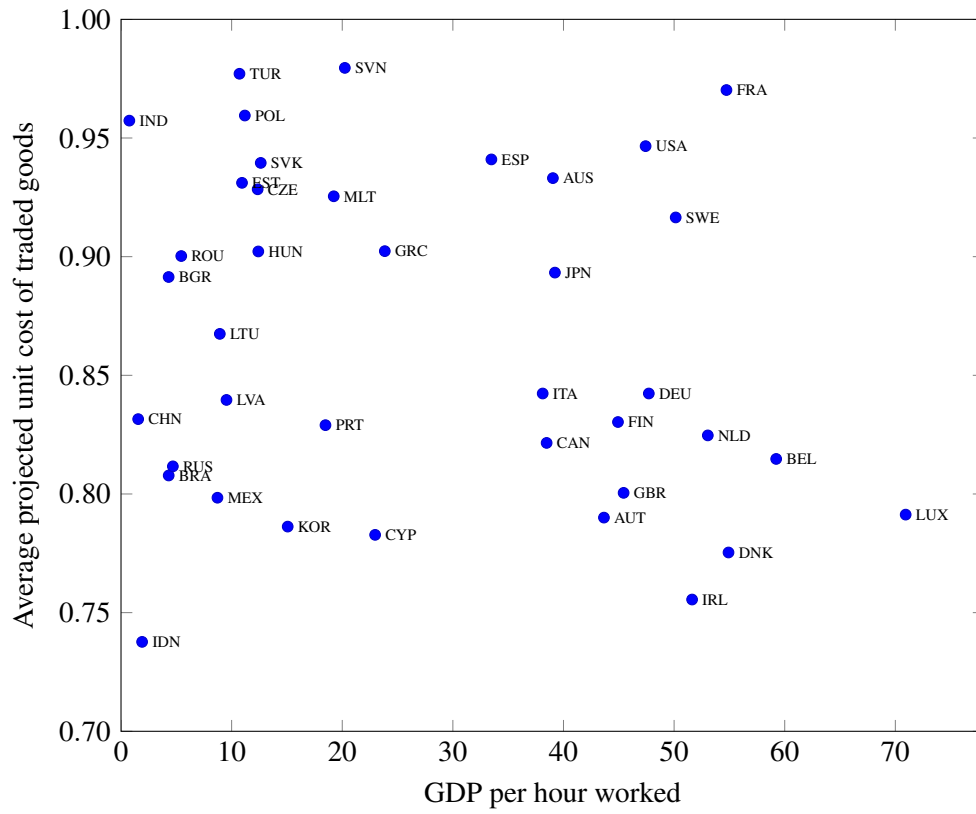Figure 8: Unit costs by country

Figure 9: Balassa-Samuelson effects

Figure 10: Unit costs of traded goods are uncorrelated with productivity

its total exports; the bottom of the fraction is the same ratio for the total exports of the world. It measures the ratio of the actual trade pattern to the expected. A value greater than unity is taken as 'revealing' a comparative advantage, while a value less than unity indicates comparative disadvantage.

While popular and simple to construct, this metric has been heavily criticized for a number of reasons. First, Hinloopen and Marrewijk (2001) show that it has poor distributional properties, being both asymmetric and unstable over time.[12] Moreover, although the index is generally treated as ordinal rather than cardinal, it is not even clear that the index can provide an ordinal ranking of comparative advantage, since the distributions vary across industries, as Yeats (1985) emphasizes.

Here is an important point: the standard measure of the Balassa index is poorly related to the theoretical concept of comparative advantage, which is based on differences in relative autarkic prices, not on the equilibrium volume of trade. A recent paper by Costinot et al. (2012) shows how a theoretically grounded measure of comparative productivity can be derived from the model of Eaton and Kortum (2002) and then estimated from observed trade flows if one has knowledge of intra-industry productivity heterogeneity. Our technique, by way of contrast, provides a means to extract estimates of the unobservable local unit costs directly from the observed unit value matrix; hence, we can also use our projections as an indicator of revealed comparative advantage. These differences in unit costs are the essence of Ricardian comparative advantage. For a given sector, a country has a revealed comparative advantage if its projected local unit cost is below the median for that of all of the sample countries. Our measure has a substantial advantage over the traditional one because it satisfies both ordinal and *cardinal* properties.[13] Our measure is ordinal because one can say that China's revealed comparative advantage in "Basic and fabricated metals" is stronger that its measure for "Transport equipment." This statement is equivalent to saying that China's estimated unit cost is about 22% below the median among all countries for the former, but its estimated unit cost for the latter is only about 17% below the median. It is cardinal because we know what this difference means in a Ricardian model with many goods: if there an 18% appreciation of the real exchange rate, China will no longer be as competitive as the median country in the world in "Transport equipment", but it will still be so in "Basic and fabricated metal".

Figure 11 presents our computations for the United States, Mexico, and China for the sixteen traded goods sectors. Positive values indicate a comparative advantage relative to the reference countries, while negative values signify a comparative disadvantage. Because the index is cardinal, it makes sense to order it; hence, the chart lists the sectors from top to bottom in terms of increasing comparative advantage for China. The patterns are broadly in accord with the typical stylized facts.[14] China has a revealed comparative advantage in metals, machinery, electrical equipment and textiles, and a comparative disadvantage in products based upon natural resources such as agricultural and wood-based products. Mexico tends to have a comparative advantage in mining and quarrying and related products, such as minerals and metals, and also in food and wood products. The results for the US are a little more perplexing, with a revealed comparative advantage in only in "Electrical equipment", "Textiles and textile products", and "Chemicals and chemical products." These manufacturing sectors are coarsely aggregated, and the United States may have revealed comparative advantage only in high-tech manufacturing processes.

---

[12]A number of variations, such as the symmetric RCA index, have been proposed to deal with the former issue. See Dalum et al. (1998) and Vollrath (1991).

[13]The main disadvantage of the measure has to do with the data. Consistent macroeconomic have coarser commodity classifications than trade flow data. Also, the input-output data on which our measure is based is currently available for only 39 countries. Moreover, like any revealed measure, including those of (Balassa, 1965) and (Costinot et al., 2012), it may reflect existing distortions to the trade. The usual interpretative cautions therefore apply.

[14]The most recent implementation of the World Bank's WITS database allows for straightforward construction of the Balassa index for comparison purposes.
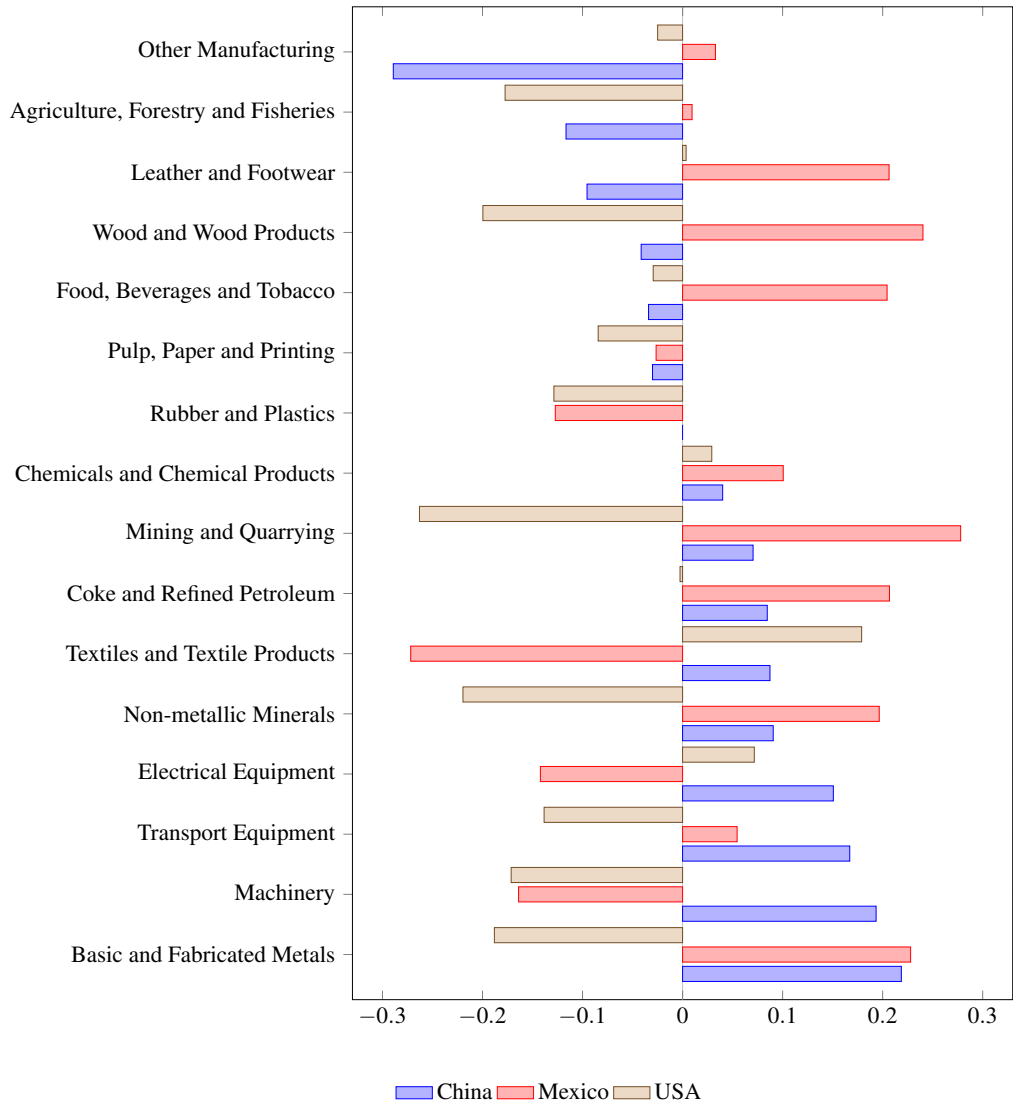
Figure 11: Revealed comparative advantage and disadvantage for 16 traded goods

6 Conclusion

We have introduced a new measure of bilateral economic distance, the maximal percentage difference of unit costs across all sectors of two comparable economies. Since local unit costs are not observable, one must estimate them. We are the first to have observed that the least squares projection of the unit vector onto a country's unit-value technology matrix is the natural estimator for these costs.

Neither a country's physical technology nor the world prices that clear international markets are observable. Using the identifying assumption that a physical unit of a good is an international dollar's worth of it, the theory attributes differences in unit-value matrices to disparate physical technologies. This insight allows one to use the projection of the unit vector onto a local technology to estimate local unit costs. This projection ia nothing more than the least squares estimator, and its coefficients are the factor prices that give rise to the local unit costs that are as near to world prices as possible. This main theoretical result is mathematically elegant, strikingly simple, and powerful. The maximal absolute deviation of this regression is the distance of the local economy from any country with a technology that can produce every good at world prices.

The theory was designed to explore the World Input-Output Database (WIOD) that describes the technologies of 39 countries that account for 80% of world GDP. The general economic distance between two countries assumes that every good is traded; it has the interpretation as a uniform *ad valorem* tariff that shuts down all bilateral trade. The economic distance based upon traded goods is a restriction of the general distance onto its first sixteen coordinates because the last nineteen goods in the WIOD are considered not traded by most economists. Because we have 741 bilateral distances, we presented our results as a network theory of the tightest links. We think of this network as highlighting "Heckscher-Ohlin-Vanek" similarity; two vertices that are close have production structures that are similar, and trade between them occurs mostly because of differences in endowments. The taxonomic interpretation of the minimum spanning tree showed how countries evolve from root to branch. Having the highest degree, Slovakia was identified as the root of the tree of "Heckscher-Ohlin-Vanek" similarity.

These distance measures are only as good as the projected unit costs that underly them. Three robustness checks were done. First, these costs were used to compute Geary-Khamis prices and the compared with those of authors who had worked with these data in a completely different way. Second, these costs define real exchange rates that exhibit strong the Balassa-Samuelson effects. Third, we defined a new measure of revealed comparative advantage that seemed to corroborate elements of the trade patterns of China, Mexico, and the United States.

There are at least four avenues for future research. First, a measure of generalized economic distance may be more relevant for the estimation of a gravity equation than a measure of the geographic distance between the economic centers of two countries. So it would be very interesting to re-estimate the classical gravity equation using this new measure. Second, this technique allows for an endogenous determination of the margin for traded and non-traded goods. If physical impediments to trade determine this margin, then this measure of bilateral distance shows the bound below which decreased trade costs makes a formerly not-traded good traded. Third, this distance measure points towards a quantitative theory of the difference between Heckscher-Ohlin and Ricardian trade. If distant countries have a large volume of bilateral trade, then it is likely due to Ricardian forces. Fourth, this measure lays the foundation for the study of the international transmission of business cycles. If a primary source of macroeconomic business cycles is the local propagation of sectoral shocks, then countries with similar production structures should have highly correlated business cycles. We hope that this measure has laid the groundwork for a lot of future inquiry.

Table A1: Sectors

ISIC Rev. 3	Description
A,B	Agriculture, hunting, forestry and fishing
C	Mining and quarrying
15,16	Food, beverages, and tobacco
17,18	Textiles, textile products, leather and footwear
19	Leather and footwear
20	Wood and products of wood and cork
21,22	Pulp, paper, paper products, printing and publishing
23	Coke, refined petroleum products and nuclear fuel
24	Chemicals and chemical products
25	Rubber & plastics
26	Other non-metallic mineral products
27,28	Basic metals and fabricated metal
29	Machinery, Nec
30-33	Electrical and optical equipment
34,35	Transport equipment
36,37	Manufacturing, Nec; Recycling
E	Electricity, gas, and water supply
F	Construction
50	Sale, maintenance and repair of motor vehicles; Retail sale of fuel
51	Wholesale trade
52	Retail trade
H	Hotels and restaurants
60	Inland transport
61	Water transport
62	Air transport
63	Other transport activities
64	Post and telecommunications
J	Financial intermediation
70	Real estate activities
71-74	Renting of machinery & equipment and other business activities
L	Public administration and defence
M	Education
N	Health and social work
O	Other community, social and personal services
P	Private households with employed persons

Table A2: Countries

Name	Abbreviation	GDP per hour worked
Australia	AUS	$39.04
Austria	AUT	$43.66
Belgium	BEL	$59.23
Brazil	BRA	$4.30
Bulgaria	BGR	$4.31
Canada	CAN	$38.48
China	CHN	$1.55
Cyprus	CYP	$22.97
Czech Republic	CZE	$12.35
Denmark	DNK	$47.71
Estonia	EST	$10.94
Finland	FIN	$44.93
France	FRA	$54.73
Germany	DEU	$54.92
Greece	GRC	$23.85
Hungary	HUN	$12.41
India	IND	$1.91
Indonesia	IDN	$0.75
Ireland	IRL	$51.63
Italy	ITA	$38.12
Japan	JPN	$39.23
Korea	KOR	$15.07
Latvia	LVA	$9.54
Lithuania	LTU	$8.93
Luxembourg	LUX	$70.93
Malta	MLT	$19.23
Mexico	MEX	$8.72
Netherlands	NLD	$53.05
Poland	POL	$11.19
Portugal	PRT	$18.48
Romania	ROU	$5.45
Russia	RUS	$4.69
Slovak Republic	SVK	$12.63
Slovenia	SVN	$20.23
Spain	ESP	$33.48
Sweden	SWE	$50.14
Turkey	TUR	$10.71
United Kingdom	GBR	$45.44
United States	USA	$47.42

Bibliography

Anderson, J.E. and J.P Neary (1996), "A new approach to evaluating trade policy." *Review of Economic Studies*, 63, 107–25.

Anderson, J.E. and E. van Wincoop (2004), "Trade costs." *Journal of Economic Literature*, 42, 691–751.

Arvis, J-F., Y. Duval, B. Shepherd, and C. Utoktham (2013), "Trade costs in the developing world: 1995-2010." World Bank Policy Working Paper Number 6309.

Balassa, B. (1964), "The purchasing-power parity doctrine: A reappraisal." *Journal of Political Economy*, 72, 584–596.

Balassa, B. (1965), "Trade liberalization and 'revealed' comparative advantage." *Manchester School of Economic and Social Studies*, 33, 92–123.

Bowen, H.P., E.E. Leamer, and L. Sveikauskas (1987), "Multicountry, multifactor tests of the factor abundance theory." *American Economic Review*, 77, 791–809.

Bradford, S. (2005), "The welfare effects of distribution regulations in OECD countries." *Economic Inquiry*, 43, 795–811.

Burstein, A., J.C. Neves, and S. Rebelo (2003), "Distribution costs and real exchange rate dynamics during exchange-rate-based stabilizations." *Journal of Monetary Economics*, 50, 1189–214.

Chen, M.X. and A. Mattoo (2008), "Regionalism in standards: Good or bad for trade?" *Canadian Journal of Economics*, 41, 838–63.

Chen, N. and D. Novy (2011), "Gravity, trade integration, and heterogeneity across industries." *Journal of International Economics*, 85, 206–21.

Chen, N. and D. Novy (2012), "On the measurement of trade costs: Direct vs. indirect approaches to quantifying standards and technical regulations." *World Trade Review*, 11, 401–14.

Chor, D. (2010), "Unpacking sources of comparative advantage: A quantitative approach." *Journal of International Economics*, 82, 152–67.

Costinot, A., D. Donaldson, and I. Komunjer (2012), "What goods do countries trade? A quantitative exploration of ricardo's ideas." *Review of Economic Studies*, 79, 581–608.

Dalum, B., K. Laursen, and G. Villumsen (1998), "Structural change in OECD export specialisation patterns: De-specialisation and 'stickiness'." *International Review of Applied Economics*, 12, 423–443.

Davis, D.R. and D.E. Weinstein (2001), "An account of global factor trade." *American Economic Review*, 91, 1423–1453.

Deardorff, A.V. (1998), "The regionalization of the world economy." chapter Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?, University of Chicago Press and NBER.

Deardorff, A.V. (2014), "Local comparative advantage: Trade costs and the pattern of trade." *International Journal of Economic Theory*, 10, 9–35.

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

Fisher, E. O'N. and K. G. Marshall (2011), "The structure of the American economy." *Review of International Economics*, 19, 15–31.

Fisher, E. O'N. and K.G. Marshall (2013), "Testing the Heckscher-Ohlin-Vanek paradigm in a world with cheap foreign labor." Unpublished manuscript.

Gaulier, G., D. Mirza, S. Turban, and S. Zignago (2008), "International transportation costs around the world: A new cif/fob rates dataset." Technical report, CEPII.

Head, K. and J. Ries (2001), "Increasing returns versus national product differentiation as an explanation for the pattern of US-Canada trade." *American Economic Review*, 91, 858–76.

Hinloopen, J. and C. Van Marrewijk (2001), "On the empirical distribution of the Balassa index." *Weltwirtschaftliches Archiv*, 137, 1–35.

Hummels, D. (2007), "Transportation costs and international trade in the second era of globalization." *Journal of Economic Perspectives*, 21, 131–54.

Hummels, D. and V. Lugovskyy (2006), "Are matched partner trade statistics a usable measure of transportation costs?" *Review of International Economics*, 14, 69–86.

Hummels, D. and G. Schaur (2013), "Time as a trade barrier." *American Economic Review*, 103, 2935–59.

Inklaar, Robert and Marcel P. Timmer (2013), "The relative price of services." *Review of Income and Wealth*, 1–20.

Jacks, D.S., C.M. Meissner, and D. Novy (2011), "Trade booms, trade busts, and trade costs." *Journal of International Economics*, 83, 185–201.

Kee, H-L., A. Nicita, and M. Olarreaga (2009), "Estimating trade restrictiveness indices." *Economic Journal*, 119, 172–99.

Marshall, K.G. (2012), "International productivity and factor price comparisons." *Journal of International Economics*, 87, 386–390.

Maskus, K.E. and S. Nishioka (2009), "Development-related biases in factor productivities and the HOV model of trade." *Canadian Journal of Economics*, 42, 519–553.

Novy, D. (2013), "Gravity redux: Measuring international trade costs with panel data." *Economic Inquiry*, 51, 101–21.

Rogoff, K. (1996), "The purchasing power parity puzzle." *Journal of Economic Literature*, 34, 647–668.

Salter, W.E.G. (1959), "Internal and external balance: The role of price and expenditure effects." *Economic Record*, 35, 226–238.

Samuelson, P.A. (1964), "Theoretical notes on trade problems." *Review of Economics and Statistics*, 46, 145–154.

Timmer, M.P. (2012), "The world input-output database (WIOD): Contents, sources and methods." WIOD Working Paper Number 10.

Vollrath, T. (1991), "A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage." *Weltwirtschaftliches Archiv*, 127, 265–80.

Yeats, A.J. (1985), "On the appropriate interpretation of the revealed comparative advantage index: Implications of a methodology based on industry sector analysis." *Weltwirtschaftliches Archiv*, 121, 61–73.