Unpacking Sources of Comparative Advantage: A Quantitative Approach*

Davin Chor†
Singapore Management University

First version: November 2006
This version: July 2010

Comments welcome

Abstract

This paper develops an approach for quantifying the importance of different sources of comparative advantage, by extending the Eaton and Kortum (2002) model to predict industry trade flows. In this framework, comparative advantage is determined by the interaction of country and industry characteristics, with countries specializing in industries whose production needs they can best meet with their factor endowments and institutional strengths. I estimate the model parameters using: (i) OLS; and (ii) a simulated method of moments procedure that accounts for the prevalence of zeros in the bilateral trade data. I apply the model to explore various quantitative questions, such as how much distance, Ricardian productivity, factor endowments, and institutions each matter for country welfare in the global trade equilibrium.

Keywords: Comparative advantage, gravity, Ricardian model, factor endowments, institutional determinants of trade, simulated method of moments

JEL Classification: C15, F11, F15, F17

---

*I am grateful to the editor, Daniel Trefler, as well as to Pol Antràs, Gary Chamberlain, Elhanan Helpman, and Marc J. Melitz for their patient advice. Thanks also to Leon Berkelmans, Filipe Campante, Arnaud Costinot, Jonathan Eaton, Doireann Fitzgerald, Hiroyuki Kasahara, Samuel Kortum, Kalina Manova, Brent Neiman, Nathan Nunn, Adrian Ow, Parag Pathak (who started as a co-author), Natalia Ramondo, Roberto Rigobon, Kartini Shastry, and Jeffrey Williamson, as well as seminar audiences at Harvard, Southern Methodist, Western Ontario, National University of Singapore, Singapore Management University, the Midwest International Economics Group Meeting (Michigan), the North American Winter Meetings and the Far East Meetings of the Econometric Society (New Orleans, Singapore), and the Empirical Investigations in Trade and Investment meeting (Keio). Stephen F. Lin, Kalina Manova, Marc J. Melitz, Nathan Nunn, and Stanley Watt generously shared data and code. Gurmeet Singh Ghumann provided excellent research assistance. All errors are my own.

†School of Economics, Singapore Management University, 90 Stamford Rd, Singapore 178903. Tel: +(65) 6828-0876. Fax: +(65) 6828-0833. E-mail: davinchor@smu.edu.sg

1 Introduction

The past few years have seen a resurgence in empirical work on sources of comparative advantage, namely those forces, such as country differences in productivity or factor endowments, that determine patterns of specialization and trade. On the Ricardian model, it is only recently that Eaton and Kortum (2002) showed how an appropriate parametrization for the underlying distribution of productivity levels can deliver a tractable expression for trade flows in a general multi-country setup. The good fit of their model to the manufacturing trade data delivered an important piece of evidence on the role of productivity differences in determining comparative advantage (Eaton and Kortum, 2002; Costinot and Komunjer, 2007).[1] Separately, several studies have reaffirmed the role of factor endowments and the Heckscher-Ohlin framework, showing in particular that countries' relative endowments are informative of their pattern of trade (Debaere, 2003; Romalis, 2004).[2] Moving beyond this neoclassical focus, a recent cluster of work has identified the influence of country institutions on international trade. Such institutional sources of comparative advantage include: financial development (Beck, 2003; Manova, 2008), the security of contract enforcement (Levchenko, 2007; Nunn, 2007; Costinot, 2009), and labor market flexibility (Cuñat and Melitz, 2009).[3]

This paper seeks to quantify the importance of these different sources of comparative advantage for country welfare within a common framework. I present an extension of the Eaton-Kortum (EK) model that goes beyond aggregate trade volumes to explain the cross-country pattern of specialization and industry trade flows. In the model, the productivity level of firms is composed of a systematic and a stochastic component, where the systematic component is driven by the interaction between country and industry characteristics. The motivation for this is intuitive: Industries vary in the factors and institutional conditions needed for production, and countries differ in their ability to provide for these industry-specific requirements. Comparative advantage therefore stems from such country-industry matches.

At heart, this specification draws on a recent body of work that identifies comparative advantage

[1]Ricardian models of an earlier vintage, such as Dornbusch et al. (1977), could not easily be taken to empirical work, largely because these were two-country models featuring complete specialization. Most previous studies instead tested reduced-form implications, such as whether countries tend to export more in industries where domestic productivity is higher; see for example Section 3 of Deardorff (1984), and Golub and Hsieh (2000).

[2]There is also an established related literature seeking to explain the *absolute* levels of the factor content of trade; see for example Trefler (1993, 1995), and Davis and Weinstein (2001).

[3]Strictly speaking, these institutional determinants of trade can be viewed as a subset of the Ricardian model, insofar as the mechanism through which institutions operate is to influence domestic industry productivity.

from such interactions between country and industry characteristics. Romalis (2004) applied this logic to test for Heckscher-Ohlin forces: By interacting countries' relative factor abundance with an industry measure of factor intensities in production, he showed that countries capture a larger US market share in industries that use their abundant factors more intensively.[4] The literature on institutional determinants of trade has also adopted this empirical strategy, by applying or constructing novel measures of an industry's dependence on specific institutional conditions. Beck (2003) and Manova (2008) interacted country measures of private credit availability with an industry measure of external capital dependence, to show that countries with better financial development export more in industries that rely heavily on external financing.[5] Similarly, several studies have shown that countries with better rule of law export relatively more in industries that are more exposed to holdup problems or other institutional frictions, as measured by input concentration (Levchenko, 2007), the share of customized inputs (Nunn, 2007), or job task complexity (Costinot, 2009).[6] Cuñat and Melitz (2009) have further demonstrated that countries with flexible labor markets facilitate exports in more volatile industries, these being the industries that benefit most from being able to adjust employment margins regularly.

The model that I present in Section 2 provides a common interpretation for the estimation being performed in this recent literature, by embedding these specifications within the multi-country setting of the EK model. Conveniently, the model delivers an analytic expression for trade flows at the industry level that resembles a gravity equation, which incorporates a role for distance barriers, Ricardian forces, Heckscher-Ohlin forces, and institutional determinants in explaining trade volumes. The model can thus be readily taken to the data.

Section 3 estimates the trade flow expressions using a fixed effects ordinary least squares (OLS) specification. For the empirical implementation, I assembled a dataset of bilateral industry trade flows, distance measures, as well as country and industry characteristics for a sample of 83 countries and 20 manufacturing industries, including a comprehensive set of all the country-industry

[4] See Baldwin (1971, 1979) for early work on the correlation between industry factor intensities and net exports.

[5] This builds on Rajan and Zingales (1998), who showed that countries with better financial development experienced higher growth rates in industries that are more dependent on external financing. The link between financial development and trade has also been explored by Beck (2002), Wynne (2005), Svaleryd and Vlachos (2005), Hur et al. (2006), and Becker and Greenberg (2007). For related theoretical work, see Kletzer and Bardhan (1987), and Matsuyama (2005).

[6] The effect of the institutional rule of law on trade flows has also been studied by Anderson and Marcouillier (2002), Berkowitz et al. (2006), and Ranjan and Lee (2007). For theoretical work formalizing the role of contract enforcement as a source of comparative advantage in an incomplete contracts framework, see Acemoglu et al. (2007).

interaction terms from the papers cited above. I find strong evidence for the importance of factor endowments, financial development, legal institutions, and labor market regimes as sources of comparative advantage, even when all interaction terms are run in one regression. This represents a first exercise (to the best of my knowledge) at jointly verifying the significance of this extensive a list of trade determinants, while also facilitating a first-pass comparison of their quantitative importance.

While OLS provides a useful baseline, it nevertheless suffers from the drawback that zero trade observations are dropped when the log of trade flows is the dependent variable. These zeros constitute about two-thirds of the data, and discarding this information could systematically bias the OLS coefficients (see Santos-Silva and Tenreyro, 2006, and Helpman et al., 2008, among others).[7] To address this, Section 4 pursues a minimal modification of the model to generate zero trade predictions. Specifically, I impose a bounded support on the distribution of the stochastic productivity component, so that a country with a low systematic productivity may thus never receive a large enough productivity shock to be able to export a good to a given market. As we however lose closed-form expressions, I instead implement a simulated method of moments (SMM) procedure to obtain a separate set of parameter estimates, by matching key statistical moments in the actual data with those simulated from the model (Pakes and Pollard, 1989).[8]

Using the SMM estimates, I briefly explore two sets of quantitative exercises in Section 5. A first set of counterfactuals relates to distance and geography. The model implies a sizeable average increase in country welfare (15.7%) from a hypothetical reduction of distance measures to their minimum value, comparable to what EK (2002) find for their OECD sample (16.1%-24.1%). Second, the framework allows us to assess the relative importance of the various sources of comparative advantage for country welfare. I explore this by shutting down the relevant terms in the empirical model that capture each comparative advantage force, namely the ability of a country to benefit from relative producer price differences across industries vis-à-vis other countries stemming from the source of comparative advantage in question. The simulations indicate that each of the individual Ricardian, Heckscher-Ohlin and institutional determinants are comparably important, with the impact of neutralizing each being similar to doubling the distance markup faced by the country.

[7]Haveman and Hummels (2004) and Anderson and van Wincoop (2004) point out that the traditional gravity equation is inconsistent with the presence of zeros in the trade data. EK (2002) were not affected by this criticism, since their dataset of aggregate OECD trade flows contains no zeros.

[8]SMM methods have previously been applied to estimate variants of the EK model by Bernard et al. (2003), Eaton et al. (2008), and Ramondo (2008).

This paper falls within a broader body of research seeking to quantify the importance of different determinants of trade flows. Past work has examined for example the welfare effects of moving to a zero-gravity world (EK, 2002), border effects (Anderson and van Wincoop, 2003), and tariff liberalization (Lai and Trefler, 2002; Lai and Zhu, 2004; Alvarez and Lucas, 2007). The approach I develop here goes a step further in enabling the researcher to explore the role of country and industry characteristics in influencing the pattern of comparative advantage. In doing so, I build on recent studies which have sought a more holistic view on the determinants of trade by incorporating both Ricardian and Heckscher-Ohlin forces within a common setting (Harrigan, 1997; Morrow, 2008; Burstein and Vogel, 2009). In this regard, a closely related paper is Shikher (2010), who also extends the EK model to the industry level. Empirically, Shikher calibrates the technology parameters to fit the output and trade data, whereas the approach that I take will instead link these productivity parameters to observable country and industry characteristics.

The roadmap of the paper is as follows. Section 2 extends the canonical EK model to the industry level. Section 3 presents the OLS results. I modify the model in Section 4 to incorporate the zero trade flows, and re-estimate it with the SMM procedure. Section 5 briefly explores some counterfactuals. Section 6 concludes. The Data Appendix at the end of this paper provides an overview of the dataset. Further details on the data variables and the SMM estimation are documented in an online Supplementary Appendix for interested readers.

2 A Benchmark Model of Industry Trade Flows

2.1 The basic setup

Consider a world with $n = 1, \ldots, N$ countries. There are $K + 1$ industries, indexed by $k = 0, 1, \ldots, K$. Industry 0 denotes non-tradables, which is a homogeneous good sector. The tradable sectors ($k \geq 1$) feature differentiated products, where the continuum of varieties in each industry is indexed by $j^k \in [0, 1]$. (The measure of varieties in each industry is normalized to 1.) I proceed to build the model in stages.

Utility: The utility of a representative consumer in country n is given by:

$$U_n = \left(Q_n^0\right)^{1-\eta} \left(\sum_{k \geq 1} \left(\int_0^1 (Q_n^k(j^k))^\alpha \, dj^k \right)^{\frac{\beta}{\alpha}} \right)^{\frac{\eta}{\beta}}, \quad \alpha, \beta, \eta \in (0, 1) \tag{1}$$

where $Q_n^k(j^k)$ denotes the quantity of variety j^k from industry k consumed in country n. (In what follows, I suppress the superscript k for varieties unless there is cause for confusion.) Utility from tradables is aggregated via a nested constant elasticity of substitution (CES) function. Define $\varepsilon = 1/(1-\alpha) > 1$ to be the elasticity of substitution between any two varieties from the same industry, and $\phi = 1/(1-\beta) > 1$ to be the corresponding elasticity between varieties drawn from different industries. I assume that $\varepsilon > \phi$, so that varieties from the same industry are closer substitutes than varieties from different industries. Total utility is a Cobb-Douglas aggregate over the consumption of tradables and non-tradables, with the share of income spent on tradables equal to η.

The representative consumer in country n maximizes (1) subject to the budget constraint:

$$p_n^0 Q_n^0 + \sum_{k\geq 1}\left(\int_0^1 p_n^k(j) Q_n^k(j)\, dj\right) = Y_n \tag{2}$$

where Y_n is total income in country n, $p_n^k(j)$ is the price in country n of variety j from industry k, and p_n^0 is the price of country n's non-tradable good. Solving this optimization problem, it is straightforward to show that the demand for each tradable variety is:

$$Q_n^k(j) = \frac{\eta Y_n\ (P_n^k)^{\varepsilon-\phi}}{\sum_{\kappa\geq 1}(P_n^\kappa)^{1-\phi}}\ p_n^k(j)^{-\varepsilon}, \quad k \geq 1 \tag{3}$$

Here, $(P_n^k)^{1-\varepsilon} = \int_0^1 (p_n^k(j))^{1-\varepsilon} dj$ is the ideal price index for industry k faced by consumers in country n. On the other hand, the demand for the homogeneous good is simply: $Q_n^0 = (1-\eta)Y_n/p_n^0$.

Goods Prices: The market for each variety is perfectly competitive. Production technologies exhibit constant returns to scale, so all producers price at average cost. (There are no fixed costs in the model.) Consider the market for supplying an industry-k variety ($k \geq 1$) to country n. All N countries in the world are potential producers of this variety. Following EK's notation, let $p_{ni}^k(j)$ denote the price that country i would charge for exporting variety j to country n (the 'n' subscript identifies the importing country, while the 'i' subscript refers to the exporter). We have:

$$p_{ni}^k(j) = \frac{c_i^k d_{ni}^k}{z_i^k(j)} \tag{4}$$

Here, c_i^k is the unit production cost of the prospective exporter (country i) in industry k, while $d_{ni}^k \geq 1$ is the iceberg transport cost incurred due to distance or policy barriers. The $z_i^k(j)$ terms capture the Ricardian productivity of country i in the manufacture of variety j; formally, $z_i^k(j)$ is equal to the number of units of variety j that country i can produce using the same bundle of factors that would produce one unit under the baseline technology.

It is convenient to specify the unit production cost, c_i^k, to be a Cobb-Douglas aggregate over factor prices in country i, namely: $c_i^k = \prod_{f=0}^{F}(w_{if})^{s_f^k}$, where $f = 0, 1, \ldots, F$ indexes factors of production.[9] w_{if} is the local unit price of factor f, while $s_f^k \in (0,1)$ is the share of total factor payments in industry k that accrues to this factor. Under constant returns to scale, we have: $\sum_{f=0}^{F} s_f^k = 1$. Each producer takes the w_{if}'s as given, being too small to affect aggregate factor markets. These factor price terms capture the role of Heckscher-Ohlin forces, namely endowment-based production cost differences, in influencing trade patterns. It should be stressed that the presence of transport cost barriers in this model implies that factor price equalization (FPE) across countries does not hold in general.

For the distance markup, I further assume that $d_{ni}^k \leq d_{nn'}^k d_{n'i}^k$ for any three countries n, n' and i, so that it is cheaper to transport goods directly between two countries, rather than through a third country. I allow the iceberg cost to vary by industry, since some goods may be more costly to transport, for reasons such as heavier tonnage or industry tariffs.

Productivity: In order to relate productivity to observable characteristics, I specify the log productivity of country i in industry-k varieties to be:

$$\ln z_i^k(j) = \lambda_i + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} + \beta_0 \epsilon_i^k(j) \tag{5}$$

Productivity is thus composed of: (i) a systematic component, $\lambda_i + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km}$, that linearly shifts the average log productivity level of country i in this industry; and (ii) a stochastic term, $\beta_0 \epsilon_i^k(j)$, that generates idiosyncratic variation in productivity across varieties. While country i may on average be less productive than other exporters, it may nevertheless be the most productive exporter in those varieties for which it receives a good stochastic shock. The spread parameter β_0 therefore plays a key role in regulating the variance of these productivity shocks.

The systematic component of productivity is driven by a linear combination of country characteristics (L_{il}, indexed by l) and industry characteristics (M_{km}, indexed by m). This embeds the idea that it is precisely the interaction between pairs $\{l, m\}$ of country and industry attributes that determines a country's productivity position in that industry. As an example, countries where legal institutions securely enforce contracts will on average be more productive in industries that are more vulnerable to holdup problems between producers and input suppliers (Levchenko, 2007;

[9] As is well known, this is the unit cost function that emerges from the cost minimization problem when the production technology is Cobb-Douglas in the inputs, with factor shares equal to s_f^k. The constant factor shares also imply that there are no factor-intensity reversals.

Nunn, 2007). These $L_{il}M_{km}$ interaction terms will serve primarily to capture the role of institutional determinants of the pattern of trade, with the β_{lm} coefficients parameterizing how important each institutional channel is for generating a productivity edge. The λ_i and μ_k terms are exporter and industry fixed effects respectively.

As for the stochastic component of productivity, the $\epsilon_i^k(j)$'s are independent draws from the Type I extreme-value (Gumbel) distribution, with cumulative distribution function (cdf) $F(\epsilon) = \exp(-\exp(-\epsilon))$. This is the natural counterpart to EK's specification of a Fréchet distribution for productivity levels, since the natural log of a Fréchet random variable inherits a Gumbel distribution.[10] This probability specification yields a closed-form expression for trade flows, in much the same way that it delivers an explicit formula for product market shares in discrete choice models in industrial organization.

Substituting (5) into (4), the price presented by country i to country n for variety j in industry k is:

$$\ln p_{ni}^k(j) = \ln(c_i^k d_{ni}^k) - \lambda_i - \mu_k - \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} - \beta_0 \epsilon_i^k(j) \tag{6}$$

Not surprisingly, prices are increasing in unit production costs (c_i^k) and transport costs (d_{ni}^k), but a country's productivity position in variety j potentially lowers the price that it charges.

The distribution of the $\epsilon_i^k(j)$'s gives rise to a distribution of prices, $G_{ni}^k(p)$, presented by country i to country n for each industry-k variety. Using the Gumbel cdf in (6), it follows that:

$$G_{ni}^k(p) \;=\; Prob\{p_{ni}^k(j) < p\} \;=\; 1 - \exp\{-(c_i^k d_{ni}^k)^{-\theta} p^{\theta} \varphi_i^k\} \tag{7}$$

where $\theta = \frac{1}{\beta_0}$ and

$$\varphi_i^k = \exp\{\theta\lambda_i + \theta\mu_k + \theta \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km}\} \tag{8}$$

Note that φ_i^k is increasing in the systematic component of country i's productivity in industry k. Also, θ has the interpretation of an inverse productivity spread parameter.

[10] The micro-foundation that EK offer for this distributional choice thus also applies here: Suppose that firm productivity levels within a country follow a Pareto distribution. Then, the order statistic for the maximum productivity level across all firms is a Fréchet random variable. Costinot and Komunjer (2007) show that this distributional choice for the productivity shocks can be relaxed to some extent. In an earlier draft, Costinot (2009) also independently recognized this way of extending the EK model to the industry level.

2.2 Implications for trade flows

The remaining steps derive an expression for trade flows following EK (2002) closely. Countries procure each variety from the lowest-price provider, giving rise to the possibility of cross-border trade. Let $p_n^k(j) = \min\{p_{ni}^k(j) : i = 1, \ldots, N\}$ be the price actually paid by country n for variety j from industry k. The industry-k price distribution faced by country n (denoted by G_n^k) is therefore:

$$G_n^k(p) \;=\; 1 - \prod_{i=1}^{N}[1 - G_{ni}^k(p)] \;=\; 1 - \exp\{-(\sum_{i=1}^{N}(c_i^k d_{ni}^k)^{-\theta}\varphi_i^k)p^{\theta}\} \tag{9}$$

Also, let π_{ni}^k be the probability of country i being the lowest-price provider, and hence the unique exporter, of an industry-k variety to country n.[11] We have:

$$\pi_{ni}^k \;=\; \int_0^{\infty} \prod_{s \neq i} [1 - G_{ns}^k(p)]\; dG_{ni}^k(p) \;=\; \frac{(c_i^k d_{ni}^k)^{-\theta}\varphi_i^k}{\sum_{s=1}^{N}(c_s^k d_{ns}^k)^{-\theta}\varphi_s^k} \tag{10}$$

We now aggregate the trade flows across varieties in an industry. Denote by X_{ni}^k the value of industry-k exports from country i to n, with $X_n^k = \sum_{i=1}^{N} X_{ni}^k$ being country n's total consumption in this industry. It follows that:

$$\frac{X_{ni}^k}{X_n^k} \;=\; \frac{\pi_{ni}^k \int_0^{\infty}\int_0^1 p_n^k(j)Q_n^k(j)\; dj\; dG_n^k(p_n^k)}{\sum_{i=1}^{N}\pi_{ni}^k \int_0^{\infty}\int_0^1 p_n^k(j)Q_n^k(j)\; dj\; dG_n^k(p_n^k)} \;=\; \pi_{ni}^k \tag{11}$$

Note that to evaluate the total value of country n's industry-k consumption in the denominator, one needs to integrate over varieties j and the minimum price distribution, G_n^k. As shown by EK (2002), the distribution of prices in country n conditional on country i being the minimum price provider is in fact also G_n^k. Since this conditional price distribution does not depend on the identity of the exporter, it follows that the fraction of expenditures in industry k spent on imports from country i is precisely π_{ni}^k. From (10), we thus have a closed-form which expresses i's industry-k market share in country n as a function of underlying country and industry characteristics, as well as bilateral distance.

It is instructive to re-express (11) by normalizing it with respect to country n's expenditure share from a fixed reference country, u:

$$\frac{X_{ni}^k}{X_{nu}^k} = \frac{(c_i^k d_{ni}^k)^{-\theta}\varphi_i^k}{(c_u^k d_{nu}^k)^{-\theta}\varphi_u^k} \tag{12}$$

This last equation has an intuitive interpretation: Country i's market share in country n (normalized by the market share of the reference country) is decreasing in both i's relative unit cost of production

[11] I assume that there are no ties in prices.

(c_i^k/c_u^k) and in the relative bilateral distance barrier (d_{ni}^k/d_{nu}^k). Conversely, country i's market share rises in i's productivity edge in that industry $(\varphi_i^k/\varphi_u^k)$. As for the role played by the inverse spread parameter, θ, observe that (12) can be rewritten as: $\frac{X_{ni}^k}{X_{nu}^k} = \left(\frac{c_i^k d_{ni}^k/\tilde{\varphi}_i^k}{c_u^k d_{nu}^k/\tilde{\varphi}_u^k}\right)^{-\theta}$, where $\tilde{\varphi}_i^k = \exp\left\{\lambda_i + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km}\right\}$. It is convenient to view $(c_i^k d_{ni}^k/\tilde{\varphi}_i^k)$ as an "average" price for industry-k varieties exported from i to n. Suppose that this "average" price is higher for exporter i than for u, so that i exports less to market n than the reference country $(\frac{X_{ni}^k}{X_{nu}^k} < 1)$. A lower θ will then raise $\frac{X_{ni}^k}{X_{nu}^k}$, so a larger spread in the productivity shocks shifts market shares in favor of the initially smaller exporter. This stems from the fact that the Gumbel distribution features a thick right tail: A large spread parameter (low θ) increases the likelihood that a country with low systematic productivity will nevertheless get a good enough productivity shock in some varieties to emerge as the lowest-price provider.

Comparison with Eaton and Kortum (2002): It is useful to highlight the close link between the expressions for trade flows and those in EK (2002). To recapitulate, EK develop a model of aggregate trade flows in which the exporter i productivity terms, $z_i(j)$, are independent draws from a Fréchet distribution, with cdf $F_i(z) = \exp(-T_i z^{-\theta})$. Here, $T_i > 0$ is a country-specific parameter that reflects the technological position of the country, and $\theta > 1$ is an inverse spread parameter. (Industry superscripts no longer apply.) A similar derivation now yields the following expression for the share of n's expenditure that is imported from country i, which is equation (10) in EK:

$$\left(\frac{X_{ni}}{X_n}\right)^{EK} = \frac{(c_i d_{ni})^{-\theta} T_i}{\sum_{s=1}^{N} (c_s d_{ns})^{-\theta} T_s} \tag{13}$$

It follows that trade flows normalized with respect to the reference country u are:

$$\left(\frac{X_{ni}}{X_{nu}}\right)^{EK} = \frac{(c_i d_{ni})^{-\theta} T_i}{(c_u d_{nu})^{-\theta} T_u} \tag{14}$$

These are clearly direct analogues of equations (11) and (12) in this paper: Both pairs of equations explain trade shares as a function of factor costs, distance barriers, and productivity in a similar way, except that each term has now been replaced with its industry-specific counterpart. In particular, the more general productivity term, φ_i^k, takes the place of EK's technological parameter, T_i. This highlights the sense in which this paper unpacks sources of Ricardian comparative advantage, by positing a functional form for φ_i^k to relate productivity to observable characteristics that reflect how well countries are able to meet the requirements of industries along various technological and institutional dimensions.

2.3 Closing the Model

I close the model by specifying the factor market clearing conditions when factors are fully mobile across domestic industries, but immobile across borders. For each factor f and country i, this is done by equating factor payments across all $K + 1$ industries (including the non-tradable sector) to factor income:

$$s_f^0(1-\eta)Y_i + \sum_{k=1}^{K}\sum_{n=1}^{N} s_f^k X_{ni}^k = w_{if}V_{if} \tag{15}$$

where $Y_i = \sum_{f=0}^{F} w_{if}V_{if}$ and V_{if} denotes country i's endowment of factor f. This system of equations cannot be solved analytically for the w_{if}'s, since the X_{ni}^k's are non-linear functions of the factor prices, although the equilibrium w_{if}'s can in principle be computed numerically for given parameter values. (One of the w_{if}'s or one of the Y_i's will have to be designated as the global numeraire.)

Observe that summing both sides of equation (15) across all factors f, and using the fact that: $\sum_{f=0}^{F} s_f^k = 1$, yields: $(1-\eta)Y_i + \sum_{k=1}^{K}\sum_{n=1}^{N} X_{ni}^k = \sum_{f=0}^{F} w_{if}V_{if} = Y_i$. This implies:

$$\sum_{k=1}^{K}\sum_{n=1}^{N} X_{ni}^k = \eta Y_i \tag{16}$$

If we further subtract country i's consumption of differentiated goods from its own domestic industries from both sides of (16), this last equation reduces precisely to the trade balance condition for country i, with total exports on the left-hand side and total imports on the right-hand side. Thus, the factor market clearing conditions imply trade balance within the model.

3 Estimation by OLS

I turn now to the task of estimating the model. It turns out that the regression model for trade flows implied by (11) resembles closely that in existing empirical work based on standard OLS methods. The OLS exercise which follows thus provides a basis for comparison and corroboration with the current literature on sources of comparative advantage. We will address the potential bias from the omission of zero trade flows later in Section 4 using the SMM procedure.

3.1 Deriving the estimating equation

I need first to specify the empirical counterparts for several variables in the model. Following the extensive gravity literature, I write the distance markup between any country pair to be a log-linear function of observable distance measures:

$$d_{ni}^k = \exp\{\beta_d D_{ni} + \delta_k + \zeta_{ni} + \nu_{ni}^k\} \tag{17}$$

where $\beta_d D_{ni}$ is a linear combination of variables that impose an iceberg cost on trade. In the empirical implementation below, the vector of trade cost measures, D_{ni}, will include physical distance, as well as indicator variables for shared linguistic ties, colonial links, border relationships, and trade agreements.[12] The distance markup is allowed to vary by industry through the fixed effect, δ_k, since transport costs may vary with the nature of the goods being shipped. Finally, trade flows are subject to idiosyncratic shocks, $\zeta_{ni} + \nu_{ni}^k$, which include a country-pair specific component (ζ_{ni}); these are treated as iid draws from mean-zero normal distributions: $\zeta_{ni} \sim N(0, \sigma_\zeta^2)$, and $\nu_{ni}^k \sim N(0, \sigma_\nu^2)$.

We now substitute this distance term (17), as well as the definition of φ_i^k from (8), into (11). Further making use of the fact that $s_0^k = 1 - \sum_{f=1}^F s_f^k$, it is straightforward to derive the following:

$$\ln(X_{ni}^k) = -\theta \sum_{f=1}^F \left(\ln \frac{w_{if}}{w_{i0}}\right) s_f^k \;+\; \theta \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} - \theta\beta_d D_{ni} + I_i + I_{nk} - \theta\zeta_{ni} - \theta\nu_{ni}^k \tag{18}$$

Note that all the terms specific to the exporting country i (namely λ_i) have been collected in an exporter fixed effect, I_i. Likewise, all terms specific to each n-k pair have been collected in a corresponding importer-industry fixed effect, I_{nk}.

In practice, however, good data on factor prices is not readily available for a broad sample of developed and developing countries. Following the approach taken by Romalis (2004), I treat relative factor prices, $\ln \frac{w_{if}}{w_{i0}}$, as an inverse function of relative factor endowments, $\ln \frac{V_{if}}{V_{i0}}$. This leads to the estimating equation:

$$\ln(X_{ni}^k) = \sum_{f=1}^F \theta\beta_f \left(\ln \frac{V_{if}}{V_{i0}}\right) s_f^k \;+\; \theta \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} - \theta\beta_d D_{ni} + I_i + I_{nk} - \theta\zeta_{ni} - \theta\nu_{ni}^k \tag{19}$$

I therefore regress log bilateral industry trade flows, $\ln(X_{ni}^k)$, as a function of: (i) Heckscher-Ohlin forces, as captured by the interaction between country factor endowments, $\ln \frac{V_{if}}{V_{i0}}$, and industry factor intensities, s_f^k; (ii) institutional forces, through the interaction between country institutional measures, L_{il}, and industry measures of dependence, M_{km}; (iii) bilateral distance variables, D_{ni}; (iv) exporter fixed effects, I_i; and (v) importer-industry fixed effects, I_{nk}.[13] Standard errors are clustered

[12]See Anderson and van Wincoop (2004) for a survey of the many bilateral variables used to capture trade costs in gravity equations.

[13]Note that the estimating equation cannot include exporter-industry (i-k) fixed effects, since this would preclude identifying the $\theta\beta_f$ and $\theta\beta_{lm}$ coefficients.

by country pair, to allow for correlated shocks among bilateral observations ($-\theta\zeta_{ni}$). Equation (19) thus provides a neat decomposition of the determinants of trade flows, which embeds the empirical specifications in Romalis (2004) and the recent work on institutional sources of comparative advantage. Note that the $\theta\beta_f\left(\ln\frac{V_{if}}{V_{i0}}\right)s_f^k$ and $\theta\beta_{lm}L_{il}M_{km}$ terms are similar in spirit, in that both reflect how well conditions in country i provide for the production needs of industry k.

Before proceeding further, some discussion on this estimating equation is in order. Much as it is necessitated by data constraints, the use of relative factor endowments in place of relative factor prices is admittedly an *ad hoc* simplification. More typically in this literature, factor endowments would be introduced formally via a revenue function approach (Dixit and Norman, 1980; Harrigan, 1997), rather than as part of an empirical model for producer costs as has been done here. A related set of considerations lies in the extent to which relative factor prices are actually determined by relative factor endowments. On the one hand, standard Heckscher-Ohlin thinking would suggest that factor prices would in fact not depend on endowments if the latter lie within the FPE set. This is however arguably less of a concern here, since as briefly discussed in Section 2, the model in this paper does not feature FPE due to the presence of transport costs. On the other hand, forces other than factor endowments, such as institutional frictions in financial or labor markets, could in practice also influence prices in factor markets.[14] Ultimately though, if relative factor endowments are a noisy proxy for relative factor prices, I would in effect be biasing myself against finding significant effects.

One last concern lies in the prospect that pure productivity differences at the country-industry level might be picked up by the Heckscher-Ohlin or institutional terms, since these are the only terms on the right-hand side of (19) that vary at the country-industry level. Reassuringly though, there is now a body of evidence suggesting that most of the variation in productivities rests at the country-factor rather than at the country-industry level (see for example Maskus and Nishioka, 2009), which helps allay concerns over this issue.

3.2 Discussion of OLS regression results

The empirical analysis uses a large dataset of bilateral trade flows, distance measures, and country and industry characteristics. The sample consists of 83 countries, the largest number for which

[14] The framework here can certainly allow for such richer possibilities, for example by incorporating further interaction terms between country institutions and industry factor intensities. I have opted not to do so here for the sake of parsimony, as well as to keep the computational burden of the later SMM analysis manageable.

I could assemble a balanced dataset of all the country variables. For the differentiated products industries ($k \geq 1$), I work with the US 1987 Standard Industrial Classification (SIC-87) 2-digit manufacturing categories. This provides 20 industries, with SIC codes from 20 (food processing) to 39 (miscellaneous manufacturing). The analysis focuses on one year, 1990, the same year as in EK; I thus abstract from dynamic issues. (An overview of the key variables and a country list are included in the Data Appendix at the end of this paper. Further details on the data construction and some descriptive statistics can be found in the online Supplementary Appendix.)

The trade data are from Feenstra et al. (2005)'s World Trade Flows database. The original data are in the Standard Industrial Trade Classification (SITC), Revision 2 format. I convert this to SIC-87 format using detailed information on the composition of US exports to derive concordance weights.[15] As for the country and industry characteristics, these were drawn directly from or constructed following closely the methodology of existing studies. As is the standard practice in this literature, the industry characteristics are based on information on the US economy, given the data limitations for calculating these for a wide set of countries. While industry characteristics such as factor intensities may in principle differ across countries, this would not invalidate the empirical strategy so long as the relative ranking of the industries along each of these dimensions is similar across countries. As far as possible, I use data from the immediate years preceding 1990. When multiple years are available, I use averages over 1980-89 to help smooth out the data from any single year.

The sample of 83 countries accounts for 82.4% of all recorded manufacturing trade in 1990. While the total number of data points is $83 \times 82 \times 20 = 136,120$, only 45,034 (or 33.1%) of these contain a positive amount of trade. This pervasiveness of zeros is a common feature of bilateral trade data and it presents a challenge to consistent estimation of gravity equations, since the zeros are dropped from the regression when the dependent variable is log trade flows. The OLS results thus have to be interpreted strictly as estimates conditional on observing a positive trade flow.

Table 1 presents the OLS regressions of (19), where the explanatory variables are introduced in turn. The recent empirical papers on sources of comparative advantage have each worked with trade data that is in different industry classification formats and levels of aggregation. Table 1 thus verifies that the patterns identified in this preceding work are also present in the dataset here,

[15] This follows Cuñat and Melitz (2009), with the composition of US exports calculated from Feenstra et al. (2002); please see the Supplementary Appendix for details. As most of the industry characteristics used are constructed for SIC industries, it was most convenient to work with the SIC codes and concord the trade data into that format.

which works with bilateral flows at the relatively broad 2-digit level of industry aggregation.

[TABLE 1 ABOUT HERE]

Distance. Column 1 reports a basic specification that includes standard measures of trade barriers (D_{ni}). The OLS coefficients confirm the importance of distance in explaining trade patterns, although not all are statistically significant. Of note, physical distance has a negative and significant effect ($\theta\beta_{d1} = -1.152$) on trade flows, the magnitude of which implies large effects: A hypothetical halving of physical distance would slightly more than double bilateral trade, raising it by a factor of $(0.5)^{-1.152} = 2.22$. Sharing a common language ($\theta\beta_{d2}$) or colonial ties ($\theta\beta_{d3}$) promotes trade between countries. On the other hand, while the border effect ($\theta\beta_{d4}$) is positive, this is not statistically significant. I also include two commonly-used dummy variables to capture aspects of trade policy. Joint membership in an RTA ($\theta\beta_{d5}$) delivers a statistically significant boost to bilateral trade. However, I do not find a significant GATT effect ($\theta\beta_{d6}$) with OLS. These distance coefficients are remarkably stable even as more explanatory variables are subsequently included.

Heckscher-Ohlin. Column 2 demonstrates the relevance of Heckscher-Ohlin forces for the cross-country pattern of trade. Here, country measures of endowments per worker ($\log(H/L)_i$ and $\log(K/L)_i$, for human and physical capital respectively) are interacted with industry measures of factor intensity, where the latter are captured by the log factor usage per worker in the industry ($\log(H/L)^k$ and $\log(K/L)^k$).[16] The country endowment measures are from Hall and Jones (1999), while the industry factor intensities are derived from the NBER-CES dataset (Bartelsman et al., 2000). I find that countries which are more skill abundant indeed exhibit higher volumes of bilateral exports in more skill-intensive industries. Similarly, countries which have more physical capital per worker tend to export more in capital-intensive industries ($\theta\beta_{f1} = 4.148$ and $\theta\beta_{f2} = 0.056$, both significant at the 1% level). These echo the findings in Romalis (2004).[17]

Institutional determinants. The rest of Table 1 finds broad support for the hypotheses on institutional sources of comparative advantage advanced recently. Column 3 examines the role of country financial development, captured by the ratio of private credit to GDP in the economy

[16] I use this measure of industry factor intensities, instead of factor payment shares, as the former accounts for more of the variance in the trade data.

[17] This is also consistent with Trefler (1993, 1995), who finds that the fit of the data to the Heckscher-Ohlin-Vanek equations improves substantially when allowing for international productivity differences. In contrast, Davis and Weinstein (2001) need additional modifications beyond productivity differences to obtain a similarly improved fit. A likely explanation is that Trefler's dataset contains a broad range of developed and developing countries (as is the case in my paper), whereas the sample in Davis and Weinstein (2001) comprises rich OECD countries.

($FINDEV$), a measure taken from Beck et al. (2000). This is interacted against a measure of industry dependence on external finance ($CAPDEP$), calculated following the methodology of Rajan and Zingales (1998). I obtain a positive and highly significant coefficient on this interaction term ($\theta\beta_{lm1}$), confirming that financially-developed countries are more successful exporters in industries that depend more on external capital funding (Beck, 2003; Manova, 2008).

The next few columns turn to the role of the contracting and legal environment in facilitating production. Levchenko (2007) argued that industries that rely heavily on a few key inputs are more vulnerable to holdup problems from suppliers, and are hence more dependent on the legal system to enforce contracts. Column 4 examines this mechanism by interacting a Herfindahl index of input-use concentration in each industry (HI) calculated from US Input-Output Tables against a measure of the strength of legal systems in each country ($LEGAL$) from Gwartney and Lawson (2004). The positive and significant coefficient obtained ($\theta\beta_{lm2}$) indicates that countries with stronger legal systems are in a better position to specialize in goods with a high input concentration. Expanding on this incomplete contracting logic, Nunn (2007) developed an alternative measure of the extent to which holdup problems affect production, calculated as the value share of inputs that are classified as relationship-specific (RS). The interaction between this industry measure with the country variable $LEGAL$ yields a positive and significant effect ($\theta\beta_{lm3}$), providing further confirmation of the importance of contracting institutions in facilitating specialization and exports in contract-dependent industries.[18]

On a related note, Costinot (2009) proposed a different channel through which legal institutions can matter, by providing a contracting environment that facilitates the division of labor among work teams. It is argued that organizational frictions impeding the division of labor have more adverse productivity effects on industries where job tasks are more complex ($COMPL$), where complexity is measured by the length of time needed for a new worker to be fully trained for the job at hand. Column 5 shows that this logic does help in explaining trade patterns, as countries with stronger institutions do indeed export more in complex industries ($\theta\beta_{lm4}$). As in Costinot (2009), I also find that countries with a higher skill endowment are better-placed to export in more complex industries ($\theta\beta_{lm5}$), suggesting that skilled workers are indeed able to perform complex tasks more efficiently.

The final column in Table 1 considers the effect of labor market institutions, as captured by

[18]The results are similar under the alternative ways of classifying relationship-specific inputs discussed in Nunn (2007). I report results using the z^{rs2} measure in Nunn's notation, which is based on the liberal classification in Rauch (1999) and also treats inputs that are reference-priced in trade journals as relationship-specific.

an index of labor market flexibility based on Botero et al. (2004). Consistent with Cuñat and Melitz (2009), I obtain a positive coefficient ($\theta\beta_{lm6}$, significant at the 1% level) indicating that countries with flexible labor institutions ($FLEX$) export more in industries that experience greater sales volatility ($SVOL$); these are precisely the industries that rely most on being able to adjust employment to respond to changing market conditions.

Full model. All the above conclusions remain intact when I run all of these determinants jointly in a single specification, as evidenced by Table 2, Column 1. Of note, all the interaction terms for the Heckscher-Ohlin and institutional determinants are statistically significant, suggesting that the empirical literature has to date successfully identified largely independent channels through which country attributes influence the pattern of trade. To gauge the relative importance of these explanatory variables, Column 1a reports the standardized beta coefficients based on the specification in Column 1.[19] Physical distance is not surprisingly the most influential distance variable (beta coefficient = -0.319). That said, the Heckscher-Ohlin and the institutional terms collectively have a larger role in explaining trade flows than physical distance, with the sum of the beta coefficients for all eight interaction terms exceeding that for physical distance. In particular, the physical capital endowment and legal institutions appear to have the largest influence (as indicated by the beta coefficients of 0.491, 0.654, and 0.494 for $\theta\beta_{f2}$, $\theta\beta_{lm2}$, $\theta\beta_{lm3}$ respectively).

[TABLE 2 ABOUT HERE]

Column 1b provides an alternative summary of the magnitudes of these effects. For each interaction, this reports *ceteris paribus* how much larger export volumes would be for the exporting country at the 75th versus 25th percentile, in the 75th versus 25th percentile industry. To illustrate, consider $\theta\beta_{f1}$: The interquartile gap in the human capital distribution in this sample of 83 countries is 0.415, while the corresponding gap in the industry skill-intensity distribution is 0.494. The Column 1 estimate of $\theta\beta_{f1}$ then implies that trade flows would rise by a sizeable factor of $\exp(1.246 \times 0.415 \times 0.494) = 1.29$, namely a 29% increase, when moving from the 25th percentile country and industry to the 75th percentile. This exercise indicates that the country attributes with the largest roles as sources of comparative advantage are the physical capital endowment ($\log(K/L)^k \times \log(K/L)_i$, a 56% increase) and legal institutions ($HI \times LEGAL$, a 69% increase; $RS \times LEGAL$, 59%; $COMPL \times LEGAL$, 33%).

[19] The beta coefficient standardizes the OLS coefficient to capture the change in standard deviation units of the dependent variable in response to a one standard deviation increase in the right-hand side variable.

In sum, the above regressions confirm the usefulness of the model in explaining the intensive margin of trade, namely conditional on observing positive trade flows. However, a key concern with OLS is that two-thirds of the bilateral trade observations are in fact zeros, and these are dropped from the regression sample. Columns 2 and 2a clearly suggest that OLS does not provide the full picture: A probit regression based on equation (19) reveals that the same set of trade determinants also has a lot of explanatory power for the extensive margin of trade. (Column 2 reports marginal effects, while Column 2a standardizes these to report the approximate increase in the probability of observing a positive trade flow when the covariate is raised by one standard deviation.) For example, physical distance has a significant effect in deterring trade completely, while several of the sources of comparative advantage are also significant determinants of whether trade is non-zero.

As a consequence, it would be inappropriate to use the OLS estimates for a welfare exercise, without first accounting for the coefficient bias from dropping the zeros. One view here is that the zeros arise from measurement issues, either because small volumes are rounded down to zero, or because of the lack of reporting from less-developed countries (with zero assumed as a default). While this may explain some of the zeros, the fact that the probit regression has good predictive power suggests that more systematic economic forces are inhibiting trade flows.[20] Moreover, removing the countries with the lowest per capita income levels, and hence presumably the poorest quality data, has little effect on the OLS results (available upon request).[21]

4 Estimation by Simulated Method of Moments (SMM)

In keeping with the Ricardian spirit of the EK model, the approach I pursue is instead to view the zero trade flows as arising from large productivity gaps between countries, which prevent low productivity countries from exporting to particular markets. This requires a slight modification of the model to generate zero trade predictions. The underlying parameters can then be re-estimated by matching moments of trade flows simulated from the model with the corresponding moments from the actual data.[22]

[20] In this regard, neither a simple tobit regression nor an *ad hoc* fix of adding one US dollar to each trade flow are likely to be fully satisfactory (estimates from these procedures available on request). See however Eaton and Tamura (1994) for a more comprehensive tobit procedure that models the censoring value as a function of observables.

[21] I experimented with removing the 10 poorest countries, as measured by GDP per capita in 1990, all of which are African countries. The results are robust to removing slightly fewer or slightly more countries.

[22] Other approaches for dealing with the zeros-bias require further departures from the EK model. Santos-Silva and Tenreyro (2006) propose a Pseudo-Poisson maximum likelihood procedure, but this assumes a non-standard distribution for the regression error terms. Helpman et al. (2008) implement a two-stage estimation method, with a

4.1 Modifying the theory to generate zeros

In its present form, the model from Section 2 precludes any zeros, since equation (10) establishes that each country i has a strictly positive probability (albeit possibly tiny) of being the lowest-price supplier to any country n for each industry-k variety. Suppose instead that the productivity shocks, $\epsilon_i^k(j)$, are independent draws from a *truncated* Gumbel distribution with bounded support $[\underline{x}, \bar{x}]$. This has the cdf: $\tilde{F}(\epsilon) = \frac{F(\epsilon)-F(\underline{x})}{F(\bar{x})-F(\underline{x})}$, where $F(\epsilon) = \exp(-\exp(-\epsilon))$. The bounded support now makes zero predicted trade possible, as X_{ni}^k will equal zero if there exists another country, i', which is systematically more productive than i in this industry, to the extent that i cannot possibly become the lowest-price exporter even with the best productivity shock, $\bar{x}$. Formally, $X_{ni}^k = 0$ if and only if there exists a country $i' \neq i$ such that:

$$\ln(c_i^k d_{ni}^k) - \lambda_i - \mu_k - \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} - \beta_0 \bar{x} \ > \ \ln(c_{i'}^k d_{ni'}^k) - \lambda_{i'} - \mu_k - \sum_{\{l,m\}} \beta_{lm} L_{i'l} M_{km} - \beta_0 \underline{x}$$

In contrast, under the previous assumption of a Gumbel distribution with unbounded support, there would have been a positive probability of trade between every country pair in each industry, since even countries with a poor systematic productivity level would in expectation obtain large enough productivity shocks to become the lowest-price exporter of a positive measure of varieties.

Naturally, this modeling choice is not meant to imply that productivity gaps are the only reason giving rise to the zeros. Other features, most notably fixed cost barriers, would almost certainly play a role in practice. Note nevertheless that the introduction of fixed costs alone would not lead to zero trade predictions: One would still need to impose a bounded support assumption, to ensure that countries with low systematic productivity levels will never receive a large enough productivity shock to overcome the fixed cost barrier. Thus, truncating the distribution of the stochastic productivity component is a minimal extension of the model that is in fact necessary for the theory to be consistent with the zero trade flows. The further introduction of fixed costs would clearly be useful in enriching the model, but this is an extension which I leave for future work in order to first develop this perfectly competitive benchmark.

first-stage selection equation for the probability of positive trade. Their approach views fixed costs to exporting as a key reason for the zeros, whereas such fixed costs are absent from the EK model considered here.

4.2 SMM estimation procedure

Although we lose closed-form expressions for the trade flows under the bounded support assumption, we can nevertheless readily simulate a complete set of flows for given parameter values from the underlying model. I therefore pursue estimation via a SMM procedure, to search for parameter values that deliver predicted trade flows which match key statistical moments of the actual data as closely as possible (Pakes and Pollard, 1989). To implement this, I take a discrete approximation of the measure of varieties. With a slight abuse of notation, I index the varieties in each industry by $j = 1, 2, \ldots, J$. Using the price equation (6), and substituting in the distance and factor endowment terms following (19), the log price of each variety in industry k is given by:

$$\ln(p_{ni}^k)^{(j)} = \frac{1}{\theta}\left(\theta\beta_d \cdot D_{ni} - \sum_{f=1}^{F}\theta\beta_f \cdot s_f^k \ln\frac{V_{if}}{V_{i0}} - \sum_{\{l,m\}}\theta\beta_{lm} \cdot L_{il}M_{km} + \tilde{I}_i + \tilde{I}_k - (\epsilon_i^k)^{(j)}\right) \tag{20}$$

Here, $(\epsilon_i^k)^{(j)}$ is a random draw from the truncated Gumbel distribution with support $[\underline{x}, \bar{x}]$, while $\tilde{I}_i$ captures all exporter fixed effects (such as δ_i) and $\tilde{I}_k$ groups together all industry-specific terms (such as μ_k and δ_k). As a reminder, the expression for country producer prices in (20) is in turn based on an empirical model for log productivity that was spelt out in (5). Even though (5) is motivated by the specifications run in many recent papers on comparative advantage, it does mean that the procedure outlined below should be viewed strictly as a semi-structural rather than a fully structural approach.

For any given realization of the parameter values, one can simulate a full set of bilateral industry trade flows as follows:

1. For each variety j in industry k, compute the prices presented by all N countries to each importing country n using (20). This requires $N \times K \times J$ independent draws from the truncated Gumbel distribution for the productivity shocks, $(\epsilon_i^k)^{(j)}$. (Once drawn, these shocks are fixed throughout the SMM procedure.)

2. For each importing country n, identify the country that presents it with the lowest price for variety j from industry k. Denote this lowest price by $(p_{n,i(j)}^k)^{(j)}$, where $i(j)$ identifies the (unique) exporter of this variety to country n. Also, calculate the approximate ideal price indices:

$$(P_n^k)^{1-\varepsilon} \approx \frac{1}{J}\sum_{j=1}^{J}((p_{n,i(j)}^k)^{(j)})^{1-\varepsilon} \tag{21}$$

3. Using the ideal price indices from (21), calculate the quantity demanded, $(Q^k_{n,i(j)})^{(j)}$, for each variety in country n using (3). Here, the country GDP data for Y_n are taken from the World Development Indicators (WDI). The value of exports from country i to n in industry k is then obtained by summing over the relevant exporter subscripts:

$$(X^k_{ni})^{sim} = \frac{1}{J} \sum_{\{j:\ i(j)=i\}} (p^k_{n,i(j)})^{(j)} (Q^k_{n,i(j)})^{(j)} \tag{22}$$

In practice, the number of fixed effects to be estimated in (20) is large and could strain the reliability of conventional minimization algorithms. To reduce the number of parameters, I arrange the countries into five groups in ascending order of their aggregate export volumes in 1990, and assign the same exporter fixed effect to each group of countries (which I denote by $\tilde{I}_{i1}, \tilde{I}_{i2}, \ldots, \tilde{I}_{i5}$, in increasing order of observed trade). Similarly, I sort the SIC industries into three groups according to the magnitude of total trade in each industry, and assign the same industry fixed effect $(\tilde{I}_{k1}, \tilde{I}_{k2}, \tilde{I}_{k3})$ to each industry group. (The cutoffs between groups were selected at natural breakpoints in the pecking order of trade volumes by exporter or industry; see the Supplementary Appendix for the full list of cutoffs and groups.) Substituting the expression for quantity demanded (3) into (22), it is straightforward to verify that $(X^k_{ni})^{sim}$ is invariant to any constant additive term that shifts all the log prices in (20) by the same amount. I therefore normalize $\tilde{I}_{i1} = 0$ and $\tilde{I}_{k1} = 0$, since one of the country fixed effects and one of the industry fixed effects cannot be identified.

The parameter vector, Θ, to be estimated is thus:

$$\Theta = \{\beta_{d1}, \ldots, \beta_{d6}, \beta_{f1}, \beta_{f2}, \beta_{lm1}, \ldots, \beta_{lm6}, \tilde{I}_{i2}, \ldots, \tilde{I}_{i5}, \tilde{I}_{k2}, \tilde{I}_{k3}\}$$

This comprises 20 parameters, namely the distance, Heckscher-Ohlin and institutional coefficients, as well as the group fixed effects.

Before estimation, I set the remaining model parameters as follows. For the inverse spread parameter θ, EK (2002) present a range of values from 3.60 up to 12.86, depending on the estimation method. I set $\theta = 8.28$, which is the central value adopted by EK (2002).[23] While I have performed the estimation using the more extreme values of θ, these tend to lead to qualitatively similar implications on the relative importance of the different determinants of trade. (The Supplementary Appendix contains a more detailed discussion of this sensitivity analysis.) As for $\underline{x}$ and $\bar{x}$, the good

[23] When the productivity shock distribution is exactly Gumbel, it is well-known from the theory of discrete choice models that θ cannot be identified. Given the prior that the truncated productivity shock distribution covers most of the support of the Gumbel distribution, it is likely not feasible to estimate θ, hence the decision to calibrate it.

fit of the OLS regressions indicates that the closed-form expressions from the theory yield reasonable approximations for actual trade, which in turn suggests that the productivity shock distribution should include most of the relevant mass of the Gumbel distribution. While it is in principle possible to estimate $\underline{x}$ and $\bar{x}$ by introducing additional moments to match, it would nevertheless be difficult to identify these bounds precisely because the Gumbel distribution is very flat at its upper and lower tails. This weak identification of $\underline{x}$ and $\bar{x}$ is a difficult issue which I do not address. To make progress, I instead set these bounds to cover the central 99% of the mass of the (unbounded) Gumbel distribution, namely $\underline{x} = -1.667$ and $\bar{x} = 5.296$. For the elasticities of substitution, I take $\varepsilon = 3.8$ from Bernard et al. (2003), who estimate this from US firm-level data. I set $\phi = 2$ to satisfy $\varepsilon > \phi > 1$; the SMM estimates are not particularly sensitive to the use of different values for ϕ (results available on request). I follow EK in setting $\eta = 0.13$ for the consumption share of the manufacturing sector in total GDP. I opt not to formally estimate these parameters in order to put the focus squarely on estimating the β coefficients in Θ.

The estimation problem is then to determine the parameter vector, $\hat{\Theta}$, that minimizes the distance metric between selected moments, $b(\cdot)$, of the simulated trade flows, $(X_{ni}^k)^{sim}$, and the actual data, X_{ni}^k:

$$\min_{\hat{\Theta}} \; (b(\hat{\Theta}) - b(\Theta))' \; \Psi \; (b(\hat{\Theta}) - b(\Theta))$$

On the choice of moments to match, I include in $b(\Theta)$:

1. The OLS regression coefficients from (19). This gives 14 moments, which are informative for estimating $\beta_{d1}, \ldots, \beta_{d6}, \beta_{f1}, \beta_{f2}, \beta_{lm1}, \ldots, \beta_{lm6}$.

2. The share of total trade flows in each group for exporter groups 2-5 and industry groups 2-3. These should be informative for estimating $\tilde{I}_{i2}, \tilde{I}_{i3}, \tilde{I}_{i4}, \tilde{I}_{i5}$, $\tilde{I}_{k2}$ and $\tilde{I}_{k3}$.

Since the problem is exactly identified (as many moments as there are parameters), I set the optimal weight matrix Ψ to the identity matrix. In practice, I first use a Newtonian search algorithm to determine the relevant parameter subspace in which the minima lies, before using the Nelder-Mead (1965) simplex search to obtain the final SMM estimates, which I call $\hat{\Theta}^{SMM}$.[24] I set $J = 500$; experimenting with larger J raises computation time, without changing the value of the objective function substantially.

[24] I first use the `lsqnonlin` command in MATLAB for the Newtonian search, followed by `fminsearch` for the Nelder-Mead search. In practice, initializing the search at different starting values leads to similar regions of the parameter space.

I compute the standard errors from: $\Lambda = (\Gamma'\Gamma)^{-1}\Gamma' V \Gamma (\Gamma'\Gamma)^{-1}$, where $\Gamma = \frac{\partial}{\partial\Theta}(b(\hat{\Theta}^{SMM}) - b(\Theta))$, and V is the variance-covariance matrix of the moments $(b(\hat{\Theta}) - b(\Theta))$. Specifically, the standard errors are equal to $1/\sqrt{J}$ times the square root of the diagonal entries of Λ. The underlying stochastic shocks, $(\epsilon_i^k)^{(j)}$, are the only source giving rise to variation in the calculation of the moments. I thus estimate V through a Monte Carlo procedure, as the empirical variance-covariance matrix of $(b(\hat{\Theta}) - b(\Theta))$ based on 1000 sets of $N \times K \times J$ draws from the truncated Gumbel distribution, when $\hat{\Theta}$ is evaluated at $\hat{\Theta}^{SMM}$.

4.3 The SMM estimates

The $\hat{\Theta}^{SMM}$ estimates are presented in Table 2, Column 3. I report the values of $\theta\beta_d$, $\theta\beta_f$ and $\theta\beta_{lm}$ to ensure comparability with the OLS coefficients. Physical distance retains a negative and highly significant effect on trade, although the size of this coefficient is smaller than found with OLS ($\theta\beta_{d1}$ is now -0.919). This is a feature found with other bias correction methods related to the omission of zeros in the gravity equation, such as Santos-Silva and Tenreyro (2006) and Helpman et al. (2008). One explanation that has been offered is that the elasticity of trade volumes with respect to distance declines over longer distances. The exclusion of the zeros thus biases the magnitude of the OLS distance coefficient upwards, since the zeros tend to occur in high-distance country pairs where the associated distance elasticity is low (Anderson and van Wincoop 2004, p.730).

Turning to the Heckscher-Ohlin and institutional determinants, I find positive and significant effects of these interaction terms that echo the baseline OLS results.[25] Accounting for the zeros tends to reduce the SMM coefficients slightly compared to the corresponding OLS coefficients, suggesting that these sources of comparative advantage have quantitatively more explanatory power for the intensive margin of trade (how much countries trade) than the extensive margin (whether countries trade). This is consistent with Manova (2008), who finds using the Helpman et al. (2008) bias-correction method that about two-thirds of the effect of financial development ($CAPDEP \times FINDEV$) operates through the intensive margin and about one-third through selection into exporting.

How sensible are these SMM estimates in explaining the actual trade data? I illustrate here that the model delivers a fairly reasonable although admittedly imperfect fit for country GDPs and

[25]The small standard errors here are a reflection of the good fit of the model moments to the data moments, as documented in the Supplementary Appendix.

the accompanying trade patterns. One can first solve for the equilibrium country income levels (the Y_n's) implied by the SMM estimates, by applying the trade balance condition (16). With the discrete approximation for the measure of varieties, total manufacturing exports from country i, EXP_i, are given by:

$$EXP_i \approx \frac{1}{J} \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{\{j:\ i(j)=i\}} \frac{\eta Y_n \ (P_n^k)^{\varepsilon-\phi}}{\sum_{\kappa \geq 1} (P_n^\kappa)^{1-\phi}} \ ((p_{n,i(j)}^k)^{(j)})^{1-\varepsilon} \tag{23}$$

where the sum is taken over all varieties, industries and export destinations for which country i is the lowest-price exporter. Total exports from country i are thus a linear combination of the Y_n's, since import volumes are linear functions of country GDP under the Cobb-Douglas utility function. The GDP coefficients, namely the $\frac{\eta \ (P_n^k)^{\varepsilon-\phi}}{\sum_{\kappa \geq 1} (P_n^\kappa)^{1-\phi}} \ ((p_{n,i(j)}^k)^{(j)})^{1-\varepsilon}$ terms, can be evaluated at the SMM estimate $\hat{\Theta}^{SMM}$ via simulation (using $J = 500$). On the other hand, i's total imports, IMP_i, are equal to ηY_i. Setting $EXP_i = IMP_i$ for each country therefore gives a homogeneous system of N linear equations in the N income levels, Y_n. Setting income for the US to 1 (as the global numeraire), and inverting this system yields the implied equilibrium country GDPs.[26] Note that the 83 countries in the sample had a combined output equal to 92.7% of world GDP in 1990, so the country income levels computed should be close approximations for actual GDP.

Figure 1 plots the implied values for country GDP obtained in this way against the actual nominal income levels in 1990 (from the WDI). There is a tendency for the model to under-predict GDP, particularly for low-income countries which cluster under the 45-degree line. The Spearman rank correlation between the two variables is nevertheless fairly high (0.54, significant at the 1% level), so the model does capture approximately the rank order of country income levels. Likewise, the Pearson linear correlation between the two log income series is equal to 0.58, significant at the 1% level.

[FIGURE 1 ABOUT HERE]

Using these implied Y_n's, one can then compute a full set of simulated bilateral industry trade flows that are based on the SMM estimates, $\hat{\Theta}^{SMM}$. Comparing these to the actual trade data, the model does capture some key trends. The linear correlation between the simulated and actual log

[26] To operationalize this procedure when the matrix to be inverted is sparse and thus close to singular, I add to any zero entries in the matrix a small positive quantity (less than half the smallest non-zero entry); I subtract the relevant quantity from the coefficient that corresponds to each country's imports from itself, to ensure that the sum of the coefficients for each country's imports remains equal to $\eta = 0.13$.

trade flow series is 0.40 (significant at the 1% level). On another dimension, the model matches the zeros quite well: There are 91,086 zeros in the actual data, of which 84,546 are shared with the simulated trade flows. Two caveats are nevertheless in order which indicate that there is still a lot of variation in the actual data that is not fully captured. First, the model tends to under-predict trade volumes, which stems largely from its propensity to under-predict country income levels. Second, the generated trade flows display a smaller dispersion than the actual data, as evidenced by the smaller coefficient of variation for the former (0.10 versus 0.19, calculated for the subset of non-zero trade flows).

5 Counterfactuals

With the SMM estimates, we can now fully parameterize the model and explore several counterfactual exercises. It should be emphasized that the scenarios considered here are inherently hypothetical: While it is easy to mechanically perturb the model along a dimension of interest, one cannot in reality reduce the physical distance between countries or neutralize a country's ability to leverage on its endowments or institutions. The effects computed below should instead be seen as a rough gauge of the relative importance of distance barriers and the various comparative advantage forces within the context of the stylized model that has been specified here.

I adopt a model-consistent welfare metric, namely the representative consumer's indirect utility from maximizing (1) subject to the budget constraint (2):

$$W_n = \frac{(1-\eta)^{1-\eta}\eta^{\eta}\; Y_n}{(p_n^0)^{1-\eta}\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{\frac{\eta}{1-\phi}}} \tag{24}$$

Without the term, $(1-\eta)^{1-\eta}\eta^{\eta}$, this is precisely equal to country n's real GDP. Note that the price of the domestic non-tradable, p_n^0, appears explicitly in the denominator: When solving for the implied GDP levels, Y_n, from the system of trade balance equations in (23), one can only do so relative to a base country (the US), whose income level is normalized to 1. This means that domestic factor prices, and hence the price of domestic non-tradables will be endogenous in general equilibrium, and we need to account for this in the welfare calculations. This welfare measure also focuses on the representative consumer, putting aside distributional consequences within countries.

The welfare change from policy shocks can be decomposed as the change in country nominal GDP levels, net of the weighted sum of price changes in the domestic non-tradable and in the differentiated products industries:

$$\frac{\Delta W_n}{W_n} = \frac{\Delta Y_n}{Y_n} \;-\; (1-\eta)\frac{\Delta(p_n^0)}{p_n^0} \;-\; \eta\frac{\Delta\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}}{\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}} \tag{25}$$

I assume that factors of production are fully mobile domestically, but immobile across country borders. Factors can therefore move into industries that are favored by the policy shock, with factor prices adjusting accordingly. This in principle captures an upper bound on welfare gains, since it puts aside domestic frictions that might hinder the full adjustment of the economy. On the other hand, a lower bound is provided by the $-\eta\frac{\Delta\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}}{\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}}$ term: This is the welfare gain in an extreme setting where factors are completely immobile, factor prices are pinned down by their marginal productivity in the non-tradable sector, and country GDPs are therefore fixed. In this case, welfare gains accrue solely from the decrease in the price of tradables.

For each counterfactual, I evaluate (25) by simulating a full set of country trade flows both before and after introducing the shock, to compute Y_n and $\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}$, as well as their respective percentage changes. As for the change in the price of the domestic non-tradable, this is equal to the weighted sum of domestic factor price changes, where the weights are the factor share intensities in this sector: $\frac{\Delta(p_n^0)}{p_n^0} = \sum_{f=0}^{F} s_f^0 \frac{\Delta(w_{nf})}{w_{nf}}$. I approximate $\frac{\Delta(w_{nf})}{w_{nf}}$ by the percentage change in total factor payments accruing to factor f in country n, net of any change in the endowment of that factor.[27] To fully operationalize this, I set the factor shares in the outside sector as: $s_h^0 = 0.175$, $s_l^0 = 0.325$, $s_k^0 = 0.5$, based on the average factor payment shares over the 1980s in US agriculture (the canonical non-manufacturing sector) from Mundlak (2005).[28]

5.1 Distance barriers

I first consider a reduction in distance barriers. Although physical distance and transport costs cannot be fully eliminated in practice, this exercise nevertheless allows us to assess how much

[27] More explicitly, based on the factor market clearing conditions in (15), I compute the percentage change in the factor prices as:

$$\frac{\Delta(w_{nf})}{w_{nf}} = \frac{\Delta(s_f^0(1-\eta)Y_n \;+\; \sum_{k\geq 1} s_f^k \frac{1}{J}\sum_{j=1}^{J}\sum_{h\in\Omega_n^{k(j)}}(p_{hn}^k)^{(j)}(Q_{hn}^k)^{(j)})}{(s_f^0(1-\eta)Y_n \;+\; \sum_{k\geq 1} s_f^k \frac{1}{J}\sum_{j=1}^{J}\sum_{h\in\Omega_n^{k(j)}}(p_{hn}^k)^{(j)}(Q_{hn}^k)^{(j)})} - \frac{\Delta(V_{nf})}{V_{nf}}$$

where $\Omega_n^{k(j)}$ is the set of all countries for which country n is the lowest-cost provider of variety j in industry k.

[28] I set $s_k = 0.5$ based on the total factor payment share in value-added that accrues to physical capital and land (Mundlak, 2005). In the absence of better data on the breakdown of the factor shares for skilled and unskilled labor, I split up the residual share of 0.5 according to the average factor shares for non-production (s_h) and production workers (s_l) observed in the NBER-CES dataset across manufacturing sectors.

distance and geography hold back country welfare. I consider a counterfactual where all the trade cost variables are set to minimize their adverse impact on prices, namely where log physical distance is 0 and all the five dummy variables (common language, common border, colony, GATT, RTA) are equal to 1. This is strictly speaking not a pure zero-gravity experiment: In the flexible formulation of the model, the distance markup d_{ni}^k in (17) cannot be set exactly to 1, as we cannot distinguish δ_k empirically from other industry-specific terms that affect trade flows (in particular, the μ_k's in the systematic component of productivity). What is being done here is instead to assign values to the observable distance measures so that the distance markup is as small as possible.

Table 3 reports sizeable gains in this low-gravity scenario, with an average welfare increase of 15.7% in the 83-country sample. This is comparable to the range of country welfare increases that EK (2002) reported from a zero-gravity exercise with their OECD sample (16.1%-24.1%). Decomposing this welfare change using (25), the fall in the price of tradables contributes a fair amount of this welfare gain (10.1%), but the bulk of the increase is driven by the rise in country income levels (19.1%) as the removal of distance barriers opens up more trading opportunities. This is partially offset by the rise in the price of non-tradables: The increase in foreign demand for each countries' products raises the demand for factors of production domestically. Factor prices increase as a result, bringing up the price of domestic non-tradables. The lower half of Table 3 reports the less extreme scenario where only physical distance is set to 0, namely where $\beta_{d1} = 0$. As expected, this implies more moderate welfare gains.

[TABLE 3 ABOUT HERE]

5.2 Sources of comparative advantage

How much do the different comparative advantage forces matter for country welfare within the context of this model? To address this, I perform a series of counterfactuals where I neutralize the term in the model that corresponds to each source of comparative advantage for each country in turn, while holding conditions for all other countries constant. Table 4 summarizes these results; the main column of interest is the "Mean" column, which reports the welfare loss for the country for which the comparative advantage force is shut down, averaged across the 83 country-by-country scenarios.

[TABLE 4 ABOUT HERE]

Consider first the role of Heckscher-Ohlin forces. To evaluate this for factor f, I set the industry factor intensity equal to its mean value across industries in the $-\theta\beta_f \cdot s_f^k \ln \frac{V_{if}}{V_{i0}}$ interaction term in (20). This prevents the country from exploiting production cost differences across industries that stem from its endowment of the factor, hence shutting down its ability to leverage on this endowment as a source of comparative advantage, while holding constant the average production cost across industries of the country (its absolute advantage position). The model implies fairly sizeable effects, as welfare in the affected country declines on average by -3.1% when human capital is not allowed to function as a source of comparative advantage across industries. The corresponding loss for the role of physical capital is a comparable -2.8%. The last column of Table 4 correlates the welfare change experienced in a country against its initial endowment level. Not surprisingly, more skill-abundant countries suffer a larger welfare loss when one's human-capital endowment can no longer facilitate specialization in skill-intensive industries. There is a similar negative correlation between a country's physical capital endowment and the subsequent welfare loss. When both factor endowment forces are jointly shut down, welfare in the affected country declines by a more substantial -4.5%. The decomposition of these changes confirms that the bulk of the welfare shifts are driven once again by changes in country GDP in the new trade equilibrium.[29]

I repeat this exercise in the second panel of Table 4 for the institutional determinants of trade that have received much attention in the recent literature. For each $\{l, m\}$, this is done by setting M_{km} equal to its mean value across industries in the $-\theta\beta_{lm} \cdot L_{il}M_{km}$ term in (20) for the country in question. The calculations reveal mean welfare losses of -2.1% and -0.6% respectively for the affected country when the role of country financial development and flexible labor markets as channels of comparative advantage are shut down. Of note, I find sizeable effects centered on the role of legal institutions in facilitating production in industries vulnerable to holdup or organizational frictions: Shutting down specialization on the basis of industry input-concentration (HI) or the share of relationship-specific inputs (RS) leads to an average welfare decrease of -2.8% and -1.7% respectively. The role of job complexity ($COMPL$) is particularly large, with a mean welfare loss of -4.9% when this mechanism is switched off. Once again, the correlations in the final column indicate that the stronger a country's institutions, the greater the loss suffered from switching off that channel of comparative advantage. When the roles of all institutions are shut off

[29]It should not be too surprising that the welfare change due to price movements in tradables ($k \geq 1$) is small. This is because the average level of producer prices within each country is held approximately constant in this counterfactual, and countries tend to consume a large share of their own domestic production.

simultaneously, the model delivers a mean decrease in welfare of -9.0% for the affected country, approximately twice the collective impact of the Heckscher-Ohlin forces discussed earlier.

Turning to the Ricardian forces, recall that the extent to which the productivity shocks, $(\epsilon_i^k)^{(j)}$, influence relative productivity levels is mediated by $1/\theta$. As θ increases, the stochastic component of productivity becomes less influential as a source of comparative advantage. I therefore remove this comparative advantage force by setting the $(\epsilon_i^k)^{(j)}$ draws to 0 (or equivalently, letting $\theta \to \infty$) for each country in turn. This leads to a mean welfare decline for the affected country across the 83 country-by-country scenarios of about -5.8%. There is an additional source of relative productivity differences, namely the $-\theta\beta_{lm} \cdot L_{il}M_{km}$ institutional determinants, which strictly speaking are a subset of the systematic component of productivity. As might be expected, switching off both Ricardian forces (the stochastic terms and the institutional determinants) at the same time leads to a much larger mean welfare loss of -20.0% for the affected country.

It is useful to provide some basis for comparison for the magnitudes of these welfare changes. To this end, I find that a doubling of the distance markup in an analogous country-by-country exercise leads to an average welfare decrease of -2.8% for the affected country. This lies in a similar ballpark to the losses obtained when each of the Heckscher-Ohlin or institutional sources of comparative advantage are individually shut down.

In sum, the model indicates that each of the stochastic Ricardian forces, Heckscher-Ohlin forces, and institutional determinants shares a comparable degree of importance in terms of their quantitative implications for country welfare, although the joint effect of all Ricardian forces (including the institutional determinants) generates a much larger welfare impact than the joint effect of the Heckscher-Ohlin forces. One should naturally interpret these numbers with care. In particular, Table 4 does not present a strict welfare decomposition across the different sources of comparative advantage, as the counterfactual for each set of forces has been run separately.

6 Conclusion

This paper presents an approach for quantifying the importance of different sources of comparative advantage within a common modeling framework. To understand patterns of specialization, I present an extension of the multi-country model of Eaton and Kortum (2002) to explain trade flows at the industry level. The model expresses comparative advantage as a function of country-industry matches, so that countries specialize in those industries whose production needs they can best meet with their endowment mix or institutional strengths.

I present two estimation approaches, namely an OLS baseline and a SMM procedure. Both sets of estimates confirm the relevance of traditional gravity measures for explaining bilateral trade, while jointly corroborating the role of factor endowments and country institutions – including financial development, the contracting environment, and labor market regimes – as sources of comparative advantage. The OLS estimates allow us to gauge the relative importance of the different trade determinants in explaining the intensive margin of trade. On the other hand, the SMM estimation takes into account both the intensive and extensive margins of trade. Counterfactual simulations based on the SMM estimates in turn allow us to compare the welfare consequences of shutting down each channel of comparative advantage.

A key strength of this framework lies in its flexibility, allowing the researcher to incorporate an extensive set of country-industry interaction terms identified in the recent literature as significant sources of comparative advantage. While I have tried to be comprehensive here, the model is clearly more general in that it can in principle accommodate additional relevant interaction terms or more flexible functional forms (such as non-linear effects), subject to the caveat that this will raise the computational cost for the SMM procedure. Naturally, the model I have estimated is highly stylized, and hence the exercises should not be viewed as an exact evaluation for policy purposes. That said, the paper takes useful steps in developing a quantitative procedure that ties specialization patterns to country and industry characteristics, and towards more extensive applications of simulation-based estimation methods in analyzing the determinants of trade flows.

7 References

Acemoglu, D., Antràs, P., Helpman, E., 2007. Contracts and technology adoption. American Economic Review 97, 916-943.

Alvarez, F., Lucas Jr., R.E., 2007. General equilibrium analysis of the Eaton-Kortum model of international trade. Journal of Monetary Economics 54, 1726-1768.

Anderson, J., Marcouiller, D., 2002. Insecurity and the pattern of trade: an empirical investigation. Review of Economics and Statistics 84, 342-352.

Anderson, J., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border problem. American Economic Review 93, 170-192.

Anderson, J., van Wincoop, E., 2004. Trade costs. Journal of Economic Literature 42, 691-751.

Baldwin, R., 1971. Determinants of the commodity structure of US trade. American Economic Review 61, 126-146.

Baldwin, R., 1979. Determinants of trade and foreign investment: further evidence. Review of Economics and Statistics 61, 40-48.

Bartelsman, E.J., Becker, R.A., Gray, W.B., 2000. NBER-CES manufacturing industry database.

Beck, T., Demirgüç-Kunt, A., Levine, R., 2000. A new database on financial development and structure. World Bank Economic Review 14, 597-605.

Beck, T., 2002. Financial development and international trade: is there a link? Journal of International Economics 57, 107-131.

Beck, T., 2003. Financial dependence and international trade. Review of International Economics 11, 296-316.

Becker, B., Greenberg, D., 2007. Financial development, fixed costs and international trade. Harvard Business School mimeo.

Berkowitz, D., Moenius, J., Pistor, K., 2006. Trade, law, and product complexity. Review of Economics and Statistics 88, 363-373.

Bernard, A., Eaton, J., Jensen, J.B., Kortum, S., 2003. Plants and productivity in international trade. American Economic Review 93, 1268-1290.

Botero, J., Djankov, S., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2004. The regulation of labor. Quarterly Journal of Economics 119, 1339-1382.

Burstein, A., Vogel, J., 2009. Globalization, technology, and the skill premium. UCLA mimeo.

Costinot, A., Komunjer, I., 2007. What goods do countries trade? New Ricardian predictions. NBER Working Paper 13691.

Costinot, A., 2009. On the origins of comparative advantage. Journal of International Economics 77, 255-264.

Cuñat, A., Melitz, M.J., 2009. Volatility, labor market flexibility, and the pattern of comparative advantage. Journal of the European Economic Association, forthcoming.

Davis, D., Weinstein, D., 2001. An account of global factor trade. American Economic Review 91, 1423-1453.

Deardorff, A., 1984. Testing trade theories and predicting trade flows. In: Jones, R.W., Kenen, P.B. (Eds.), Handbook of International Economics, vol. 1. Elsevier Science Publishers, Amsterdam.

Debaere, P., 2003. Relative factor abundance and trade. Journal of Political Economy 111, 589-610.

Dixit, A.K., Norman V., 1980. Theory of international trade. Cambridge University Press, Great Britain.

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

Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70, 1741-1779.

Eaton, J., Kortum, S., Kramarz, F., 2008. An anatomy of international trade: evidence from French firms. NBER Working Paper 14610.

Eaton, J., Tamura, A., 1994. Bilateralism and regionalism in Japanese and US trade and direct foreign investment patterns. Journal of the Japanese and International Economies 8, 478-510.

Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C., Mo, H., 2005. World trade flows: 1962-2000. NBER Working Paper 11040.

Feenstra, R.C., Romalis, J., Schott, P.K., 2002. U.S. imports, exports and tariff data, 1989-2001. NBER Working Paper 9387.

Golub, S.S., Hsieh, C.-T., 2000. Classical Ricardian theory of comparative advantage revisited. Review of International Economics 8, 221-234.

Gwartney, J., Lawson R., 2004. Economic freedom of the world: 2004 annual report. Vancouver: The Fraser Institute.

Hall, R., Jones, C., 1999. Why do some countries produce so much more output per worker than others? Quarterly Journal of Economics 114, 83-116.

Harrigan, J., 1997. Technology, factor supplies, and international specialization: estimating the neoclassical model. American Economic Review 87, 475-494.

Haveman, J., Hummels, D., 2004. Alternative hypotheses and the volume of trade: the gravity equation and the extent of specialization. Canadian Journal of Economics 37, 199-218.

Helpman, E., Melitz, M.J., Rubinstein, Y., 2008. Estimating trade flows: trading partners and trading volumes. Quarterly Journal of Economics 123, 441-487.

Hur, J., Raj, M., Riyanto, Y., 2006. The impact of financial development and asset tangibility on export. World Development 34, 1728-1741.

Kletzer, K., Bardhan, P., 1987. Credit markets and patterns of international trade. Journal of Development Economics 27, 57-70.

Lai, H., Trefler, D., 2002. The gains from trade with monopolistic competition: specification, estimation, and mis-specification. NBER Working Paper 9169.

Lai, H., Zhu, S.C., 2004. The determinants of bilateral trade. Canadian Journal of Economics 37, 459-483.

Levchenko, A., 2007. Institutional quality and international trade. Review of Economic Studies 74, 791-819.

Manova, K., 2008. Credit constraints, heterogeneous firms, and international trade. NBER Working Paper 14531.

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

Matsuyama, K., 2005. Credit market imperfections and patterns of international trade and capital flows. Journal of the European Economic Association 3, 714-723.

Morrow, P.M., 2008. East is East and West is West: a Ricardian-Heckscher-Ohlin model of comparative advantage. University of Toronto mimeo.

Mundlak, Y., 2005. Economic growth: lessons from two centuries of American agriculture. Journal of Economic Literature 43, 989-1024.

Nelder, J., Mead, R., 1965. A simplex method for function minimization. Computer Journal 7, 308-313.

Nunn, N., 2007. Relationship-specificity, incomplete contracts and the pattern of trade. Quarterly Journal of Economics 122, 569-600.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027-1057.

Rajan, R., Zingales, L., 1998. Financial dependence and growth. American Economic Review 88, 559-586.

Ranjan, P., Lee, J.Y., 2007. Contract enforcement and international trade. Economics and Politics 19, 191-218.

Ramondo, N., 2008. Size, geography, and multinational production. University of Texas at Austin mimeo.

Rauch, J.E., 1999. Networks versus markets in international trade. Journal of International Economics 48, 7-35.

Romalis, J., 2004. Factor proportions and the structure of commodity trade. American Economic Review 94, 67-97.

Rose, A., 2004. Do we really know that the WTO increases trade? American Economic Review 94, 98-104.

Santos-Silva, J.M.C., Tenreyro, S., 2006. The log of gravity. Review of Economics and Statistics 88, 641-658.

Shikher, S., 2010. Determinants of specialization and the role of trade costs. Suffolk University mimeo.

Svaleryd, H., Vlachos, J., 2005. Financial markets, the pattern of industrial specialization and comparative advantage. European Economic Review 49, 113-144.

Trefler, D., 1993. International factor price differences: Leontief was right! Journal of Political Economy 101, 961-987.

Trefler, D., 1995. The case of the missing trade and other mysteries. American Economic Review 85, 1029-1046.

Wynne, J., 2005. Wealth as a determinant of comparative advantage. American Economic Review 95, 226-254.

8 Data Appendix

A. BILATERAL VARIABLES

Trade volumes. From Feenstra et al. (2005), for the year 1990. A concordance from the SITC Rev 2 to SIC-87 format was constructed from Feenstra et al. (2002). Feenstra et al. (2002) records US export data at the disaggregate HS 10-digit level, where each HS-10 product is also assigned a 5-digit SITC Rev 2 and a 4-digit SIC-87 code. This is used to compute weights to map SITC Rev 2 into SIC-87 categories. Trade flows were then summed up to the 2-digit SIC level. A zero is entered for all exporter-importer-industry cells for which no trade was reported.

Distance. Physical distance is the great circle formula distance between countries' capital cities, taken from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII). A country's log distance to itself is set to zero. The "Common Language" and "Colony" dummies are also from the CEPII. The "Border" dummy is coded from the CIA World Factbook. The "RTA" and "GATT" dummies are based on Rose (2004), augmented with information from the WTO website. A value of 1 is assigned for all five dummies for a country's distance from itself.

B. INDUSTRY CHARACTERISTICS

Factor intensities ($\log(H/L)^k$, $\log(K/L)^k$). From the NBER-CES database. Skill intensity is the log of the ratio of non-production workers to total employment. Physical intensity is the log of the ratio of real capital stock to total employment. Averages over 1980-89 are used.

External capital dependence ($CAPDEP$). Constructed following Rajan and Zingales (1998) using Compustat North America data. A firm's dependence on external capital is the fraction of total capital expenditures over the period 1980-89 not financed by internal cash flow. The median value across firms in each SIC-87 2-digit category is used.

Input concentration (HI). Constructed following Levchenko (2007) using the 1987 US Input-Output (IO) Use Table. The IO-87 6-digit categories map cleanly into the SIC-87 4-digit categories; this correspondence table is available from the Bureau of Economic Analysis (BEA) website. Input use is then aggregated to the SIC 2-digit level, from which the Herfindahl index of input use is calculated.

Input Relationship-Specificity (RS). From Nunn (2007). RS is the share (by value) of inputs that are not sold on an organized exchange, based on the "liberal" classification in Rauch (1999); this corresponds to the z^{rs2} measure in Nunn (2007). The IO-87 codes are mapped to SIC-87 4-digit categories using the procedure described for the HI variable. This is aggregated to the 2-digit level by taking a weighted average, using the share of total input consumption of each 4-digit industry as weights.

Job complexity ($COMPL$). From Costinot (2009), based on a US Panel Survey of Income Dynamics (PSID) question asking how many months it would take a new employee with the requisite education to become "fully trained and qualified" in the respondents' job. Costinot (2009) reports the average for SIC-1972 3-digit industries, normalized to a maximum value of 1. I assign these values to all corresponding 4-digit sub-categories. For missing 4-digit categories, I assign the median complexity level observed at successively higher levels of aggregation (first the 3-digit, and if that is still missing, the 2-digit, and then the 1-digit level value). These are mapped from SIC-1972 to SIC-1987 using the weights developed by Bartelsman, Becker and Gray. For each SIC-1987 2-digit industry, the median over all its 4-digit sub-categories is then used. (The OLS results are robust to omitting SIC 21 and 29, for which direct information on complexity is not found in the PSID for any of the 3-digit industries.)

Sales Volatility ($SVOL$). From Cuñat and Melitz (2009). Equal to the employment-weighted standard deviation of sales growth for firms in the 1980-2004 Compustat sample.

Factor shares (s_h, s_l, s_m). From the NBER-CES database. The share of payments to skilled labor and unskilled labor are the ratios of non-production worker payroll and production worker payroll to total industry value-added respectively. The factor share of physical capital is the ratio of residual payments (total value-added minus total payroll) to total value-added. Averages over 1980-89 are used.

C. COUNTRY CHARACTERISTICS

Factor endowments ($\log(H/L)_i$, $\log(K/L)_i$). From Hall and Jones (1999), for 1988.

Financial development ($FINDEV$). From Beck et al. (2000). Credit extended by banks and non-bank financial intermediaries to the private sector divided by GDP, averaged over 1980-89.

Legal System ($LEGAL$). From Gwartney and Lawson (2004). Index measure of "Legal System and Property Rights" for 1985, rescaled between 0 and 1.

Employment Flexibility ($FLEX$). From the *Doing Business* database. Index of "Rigidity of Employment", averaged over 2003-06, rescaled to increase in flexibility between 0 and 1.

GDP. From the World Development Indicators (WDI), in current US dollars.

Country coverage. The 83 countries in the sample are: Argentina, Australia, Austria, Burundi, Belgium, Bolivia, Brazil, Central African Republic, Canada, Switzerland, Chile, China, Ivory Coast, Cameroon, Colombia, Costa Rica, Germany, Denmark, Dominican Republic, Algeria, Ecuador, Egypt, Spain, Finland, France, United Kingdom, Ghana, Greece, Guatemala, Honduras, Haiti, Hungary, Indonesia, India, Ireland, Iran, Israel, Italy, Jamaica, Jordan, Japan, Kenya, South Korea, Sri Lanka, Morocco, Madagascar, Mexico, Mali, Malawi, Malaysia, Niger, Nigeria, Nicaragua, Netherlands, Norway, New Zealand, Pakistan, Panama, Peru, Philippines, Papua New Guinea, Poland, Portugal, Paraguay, Senegal, Singapore, Sierra Leone, El Salvador, Sweden, Syria, Chad, Togo, Thailand, Tunisia, Turkey, Uganda, Uruguay, United States, Venezuela, South Africa, Zaire, Zambia, and Zimbabwe.

Table 1
OLS Regression Model of Industry Trade Flows

Dependent variable = $\ln\left(X_{ni}^{k}\right)$	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Distance and Geography:							
$\theta\beta_{d1}$: Log (Distance)	−1.152***	−1.155***	−1.153***	−1.161***	−1.162***	−1.155***	−1.155***
	(0.038)	(0.038)	(0.037)	(0.038)	(0.038)	(0.038)	(0.038)
$\theta\beta_{d2}$: Common Language	0.487***	0.495***	0.498***	0.500***	0.502***	0.492***	0.496***
	(0.068)	(0.068)	(0.068)	(0.069)	(0.069)	(0.068)	(0.068)
$\theta\beta_{d3}$: Colony	0.769***	0.770***	0.766***	0.768***	0.768***	0.771***	0.769***
	(0.108)	(0.108)	(0.107)	(0.108)	(0.108)	(0.108)	(0.108)
$\theta\beta_{d4}$: Border	0.203	0.193	0.191	0.192	0.192	0.191	0.193
	(0.149)	(0.149)	(0.148)	(0.149)	(0.149)	(0.149)	(0.149)
$\theta\beta_{d5}$: RTA	0.269***	0.289***	0.292***	0.288***	0.289***	0.291***	0.288***
	(0.073)	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)
$\theta\beta_{d6}$: GATT	0.180	0.226	0.227	0.226	0.207	0.237	0.225
	(0.237)	(0.238)	(0.241)	(0.240)	(0.237)	(0.243)	(0.238)
Heckscher-Ohlin:							
$\theta\beta_{f1}$: $\log(H/L)^k \times \log(H/L)_i$		4.148***	3.373***	2.478***	3.705***	1.646***	4.174***
		(0.158)	(0.158)	(0.168)	(0.156)	(0.243)	(0.158)
$\theta\beta_{f2}$: $\log(K/L)^k \times \log(K/L)_i$		0.056***	0.038**	0.173***	0.175***	0.041**	0.055***
		(0.018)	(0.018)	(0.019)	(0.019)	(0.018)	(0.018)
Institutional:							
$\theta\beta_{lm1}$: $CAPDEP \times FINDEV$			1.859***				
			(0.083)				
$\theta\beta_{lm2}$: $HI \times LEGAL$				35.544***			
				(1.633)			
$\theta\beta_{lm3}$: $RS \times LEGAL$					14.684***		
					(0.834)		
$\theta\beta_{lm4}$: $COMPL \times LEGAL$						7.864***	
						(0.413)	
$\theta\beta_{lm5}$: $COMPL \times \log(H/L)_i$						1.376***	
						(0.429)	
$\theta\beta_{lm6}$: $SVOL \times FLEX$							12.691***
							(2.239)
Exporter fixed effects:	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer-industry fixed effects:	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	45034	45034	45034	45034	45034	45034	45034
R^2	0.586	0.600	0.605	0.607	0.605	0.606	0.600

Notes: Robust standard errors, clustered by exporter-importer pair, are reported; ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

Table 2
Empirical Model of Industry Trade Flows (OLS, Probit, SMM)

	(1) OLS	(1a) OLS Betas	(1b) OLS Quantitative Effects	(2) Probit Marginal Effects	(2a) Probit Standardized Marg. Effects	(3) SMM
Distance and Geography:						
$\theta\beta_{d1}$: Log (Distance)	−1.161***	−0.319***		−0.172***	−0.136***	−0.919***
	(0.038)	(0.010)		(0.008)	(0.006)	(0.002)
$\theta\beta_{d2}$: Common Language	0.502***	0.062***		0.107***	0.042***	0.400***
	(0.069)	(0.008)		(0.013)	(0.005)	(0.002)
$\theta\beta_{d3}$: Colony	0.766***	0.052***		0.124***	0.018***	0.603***
	(0.107)	(0.007)		(0.027)	(0.004)	(0.003)
$\theta\beta_{d4}$: Border	0.189	0.012		−0.010	−0.002	0.130***
	(0.149)	(0.009)		(0.037)	(0.006)	(0.003)
$\theta\beta_{d5}$: RTA	0.290***	0.033***		0.044***	0.014***	0.192***
	(0.072)	(0.008)		(0.014)	(0.004)	(0.003)
$\theta\beta_{d6}$: GATT	0.217	0.025		−0.049	−0.022	0.168***
	(0.242)	(0.028)		(0.044)	(0.020)	(0.008)
Heckscher-Ohlin:						
$\theta\beta_{f1}$: $\log(H/L)^k \times \log(H/L)_i$	1.246***	0.170***	1.29	0.159***	0.074***	1.245***
	(0.250)	(0.034)		(0.029)	(0.013)	(0.037)
$\theta\beta_{f2}$: $\log(K/L)^k \times \log(K/L)_i$	0.164***	0.491***	1.56	0.016***	0.170***	0.093***
	(0.020)	(0.060)		(0.002)	(0.017)	(0.002)
Institutional:						
$\theta\beta_{lm1}$: $CAPDEP \times FINDEV$	1.279***	0.111***	1.15	0.064***	0.015***	0.883***
	(0.089)	(0.008)		(0.012)	(0.003)	(0.011)
$\theta\beta_{lm2}$: $HI \times LEGAL$	14.307***	0.654***	1.69	0.789***	0.126***	8.867***
	(1.669)	(0.076)		(0.181)	(0.029)	(0.341)
$\theta\beta_{lm3}$: $RS \times LEGAL$	9.638***	0.494***	1.59	0.678***	0.119***	7.032***
	(0.855)	(0.044)		(0.088)	(0.015)	(0.143)
$\theta\beta_{lm4}$: $COMPL \times LEGAL$	2.919***	0.145***	1.33	0.057	0.008	3.426***
	(0.448)	(0.022)		(0.048)	(0.007)	(0.154)
$\theta\beta_{lm5}$: $COMPL \times \log(H/L)_i$	1.462***	0.098***	1.20	−0.219***	−0.043***	0.611***
	(0.429)	(0.029)		(0.051)	(0.010)	(0.078)
$\theta\beta_{lm6}$: $SVOL \times FLEX$	9.043***	0.092***	1.09	−0.309	−0.010	8.831***
	(2.239)	(0.023)		(0.271)	(0.009)	(0.381)
Exporter fixed effects:	Yes	Yes	Yes	Yes	Yes	Groups
Importer-industry fixed effects:	Yes	Yes	Yes	Yes	Yes	Groups
Number of obs.	45034	45034	45034	134972	134972	–
R^2 or Pseudo-R^2	0.613	0.613	–	0.646	0.646	–

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively. For the OLS and probit regressions, standard errors are clustered by exporter-importer pair. Column 1a reports standardized beta coefficients from Column 1, while Column 1b reports the factor increase in trade in the 75th compared to the 25th percentile exporter and industry. Column 2a standardizes Column 2 to report the probability change from a one standard deviation increase in the right-hand side variable.

Table 3
Counterfactuals I: Reducing distance barriers

	% Welfare Change				Decomposition Due to change in:		
	Min.	Max.	Std. Dev.	Mean	Country GDP	Prices ($k \geq 1$)	Prices ($k = 0$)
Reducing all distance barriers	−13.1	48.5	12.3	**15.7**	19.1	10.1	−13.5
	By GDP per capita:						
	5th percentile		TCD	12.8	−53.9	13.3	53.4
	25th percentile		ZWE	13.4	−52.1	13.7	51.7
	50th percentile		SLV	16.4	0.5	14.1	1.8
	75th percentile		ESP	33.9	196.7	8.9	−171.7
	95th percentile		DNK	23.1	139.5	4.9	−121.3
Reducing physical distance alone	−37.1	49.8	12.5	**8.9**	−17.4	8.3	18.0
	By GDP per capita:						
	5th percentile		TCD	8.3	−73.1	11.3	70.1
	25th percentile		ZWE	9.1	−71.3	12.0	68.5
	50th percentile		SLV	11.8	−18.7	12.0	18.5
	75th percentile		ESP	29.9	177.2	7.5	−154.8
	95th percentile		DNK	14.0	83.7	3.1	−72.8

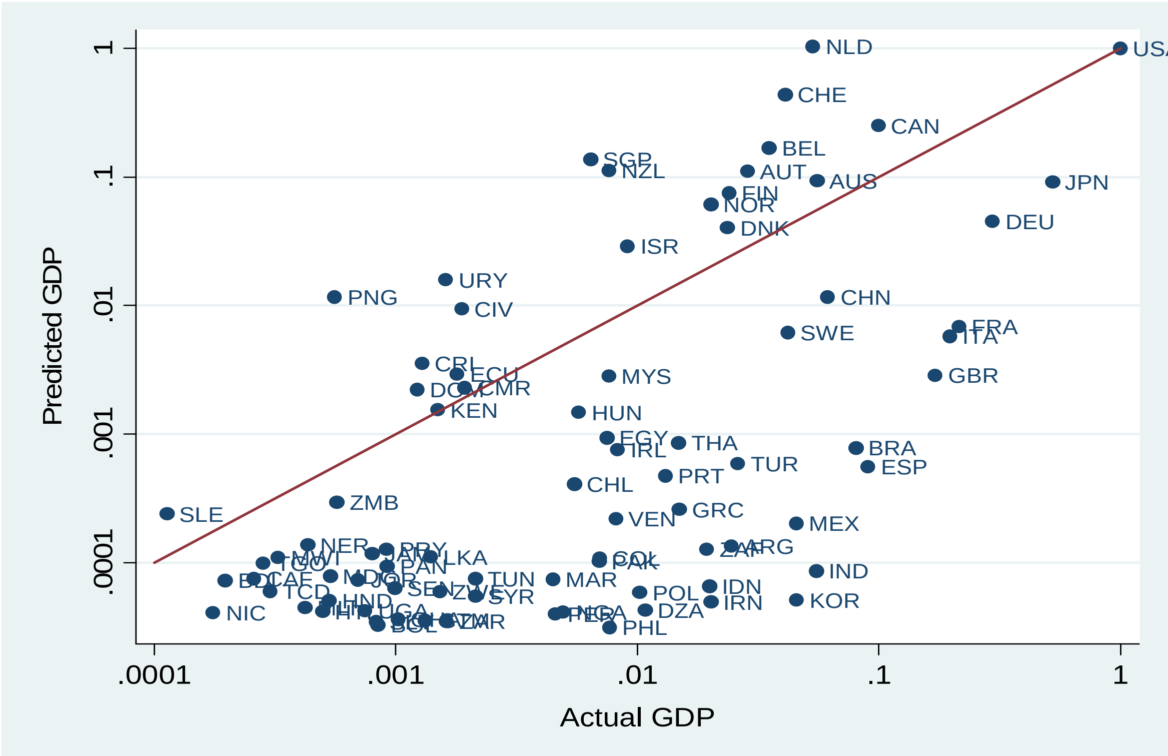
Notes: The mean percentage welfare change across countries is reported in the "Mean" column. This is broken down into the contributions from changes in country GDP, changes in the differentiated goods price index ($k \geq 1$), and changes in the price of domestic non-tradables ($k = 0$). All percentages are calculated as $100 \ln(x'/x)$, where x' and x are the final and initial values respectively.

Table 4
Counterfactuals II: Sources of Comparative Advantage

Comparative advantage force(s) switched off	% Welfare Change Min.	Max.	Std. Dev.	Mean	Decomposition Due to change in: Country GDP	Prices ($k \geq 1$)	Prices ($k = 0$)	Correlation with cty. char.
Heckscher-Ohlin forces:								
$\log(H/L)^k \times \log(H/L)_i$	−15.9	0.0	3.9	**−3.1**	−24.1	−0.0	21.1	−0.61***
$\log(K/L)^k \times \log(K/L)_i$	−14.7	4.9	3.8	**−2.8**	−22.6	−0.0	19.8	−0.35***
All Heckscher-Ohlin forces	−24.7	6.5	5.9	**−4.5**	−35.6	−0.2	31.3	
Institutional determinants:								
$CAPDEP \times FINDEV$	−19.2	0.0	3.1	**−2.1**	−16.1	−0.0	14.1	−0.60***
$HI \times LEGAL$	−20.3	0.9	3.8	**−2.8**	−22.2	−0.0	19.4	−0.64***
$RS \times LEGAL$	−16.3	8.5	2.9	**−1.7**	−13.7	−0.0	12.0	−0.41***
$COMPL \times LEGAL$	−24.7	0.0	5.8	**−4.9**	−37.9	−0.1	33.2	−0.71***
$COMPL \times \log(H/L)_i$	−6.9	0.0	1.5	**−1.1**	−8.2	−0.0	7.2	−0.56***
$SVOL \times FLEX$	−7.2	1.0	1.1	**−0.6**	−4.4	−0.0	3.8	−0.21*
All institutional determinants	−38.3	6.2	10.5	**−9.0**	−70.8	−0.2	62.1	
Ricardian forces:								
Stochastic component	−55.7	19.1	9.3	**−5.8**	−44.8	−1.0	40.1	
Stochastic component and institutional determinants	−88.1	−0.1	22.0	**−20.0**	−157.1	−1.2	138.3	
For comparison:								
Doubling distance markup	−29.7	−1.1	4.1	**−2.8**	−13.2	−1.1	11.5	

Notes: See Table 3. The final column reports the cross-country Pearson correlation between the percent welfare change and the initial level of the corresponding country characteristic; *** and * denote significance at the 1% and 10% levels respectively.

Figure 1
Assessing the Goodness of Fit: Predicted vs Actual Country GDPs



Notes: GDP levels on the horizontal axis are from the World Development Indicators (WDI). Predicted GDPs on the vertical axis are computed based on the SMM estimates. Both axes are on a log-scale; US GDP is normalized to 1. The 45-degree line is plotted for reference.