Unpacking Sources of Comparative Advantage: A Quantitative Approach*

Davin Chor†
Singapore Management University

First version: November 2006
This version: August 2007

Comments welcome

Abstract

This paper develops an approach for quantifying the relative importance of different sources of comparative advantage for country welfare in a global trade equilibrium. To explain the pattern of specialization, I present a multi-country, perfectly-competitive Ricardian model that extends Eaton and Kortum (2002) 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 specific production needs they are best able to meet with their factor endowments, institutional environment, and technological strengths. I estimate the model parameters using a large dataset of bilateral trade flows, comprising 82 countries and 20 manufacturing industries. I present results from a baseline OLS approach, and a simulated method of moments (SMM) procedure that takes into account the prevalence of zero trade flows in the data. The SMM estimates imply large average welfare gains from a hypothetical reduction in distance barriers, with developing countries benefiting substantially more than the OECD. I also examine the induced shift in industry composition when countries raise their factor endowments or improve the quality of their institutions, and quantify the welfare gains generated by such policy moves.

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

JEL Classification: C15, F11, F15, F17

*I wish to acknowledge Parag A. Pathak who started this project as a co-author and was instrumental at various stages along the way. I am also grateful to Pol Antràs, Gary Chamberlain, Elhanan Helpman, and Marc J. Melitz for their encouraging advice and guidance. I have benefited from comments from Leon Berkelmans, Filipe R. Campante, Jonathan Eaton, Doireann Fitzgerald, Kalina Manova, Brent Neiman, Nathan Nunn, Adrian Ow, Natalia Ramondo, Roberto Rigobon, Kartini Shastry, Daniel Trefler, Jeffrey G. Williamson, and seminar participants at Harvard, National University of Singapore, and Singapore Management University (SMU). Marc J. Melitz and Nathan Nunn generously shared their data, while Kalina Manova, Stephen F. Lin and Stanley Watt provided Stata code. The helpdesk staff at SMU and Peter Brown at Harvard provided vital assistance with computing issues. 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 concept of comparative advantage has been at the foundation of economists' understanding of the pattern of trade, at least since David Ricardo articulated the key intuition almost two centuries ago. The past few years have seen a much-needed resurgence in empirical work on sources of comparative advantage – those forces, such as cross-country differences in productivity or factor endowments, that determine the pattern of domestic specialization and international trade. Of note, Eaton and Kortum (2002) showed how one can parameterize a Ricardian trade model to obtain analytic expressions for trade flows in a multi-country setting, by specifying an underlying distribution that governs country productivity levels. When taken to the data, they find that their model delivers a good description of aggregate manufacturing trade within the OECD.[1] Separately, several studies have reaffirmed the role of factor endowments for explaining trade patterns within the Heckscher-Ohlin framework, showing that countries tend to be net exporters of their relatively abundant factors in North-South bilateral trade (Debeare 2003), and that countries also export more in industries that use these abundant factors more intensively (Romalis 2004).[2] Moving beyond this neo-classical focus, recent work has identified how country institutions can augment productivity, particularly in industries that are dependent on these institutional provisions to facilitate production. Such sources of comparative advantage include: financial development (Beck 2003, Manova 2006), the security of contract enforcement (Levchenko 2004, Nunn 2007, Costinot 2006), and labor market institutions (Cuñat and Melitz 2006).

The central aim of this paper is to develop a methodology for quantifying the importance of different sources of comparative advantage for country welfare in a global trade equilibrium. In order to address the cross-country pattern of industrial specialization, I adopt as a benchmark structural framework an extension of the Eaton-Kortum (EK) model that goes beyond aggregate trade volumes to explain industry trade flows. In the model, the productivity level of firms in a given country is composed of a systematic and a stochastic component, where the former (the systematic component) is driven by the interaction between country and industry characteristics. The motivation behind this is intuitive: Industries vary in the physical inputs and institutional conditions needed for production, and countries differ in their ability to provide for these industry-specific requirements. Comparative advantage therefore stems in practice from such country-industry matches. This set-up in turn enables

[1]Ricardian models of an earlier vintage, such as Dornbusch et al. (1977), were more difficult to test, largely because these were two-country models that featured complete specialization (each good exported by precisely one country). This is however inconsistent with the large volume of intra-industry trade observed in practice. Most earlier studies instead tested more general versions of the Ricardian prediction, namely whether countries tend to export relatively more in industries where domestic productivity is higher (for example, Golub and Hsieh 2000).

[2]While these studies identify a correlation between relative factor endowments and trade patterns, the literature has been much less successful in explaining the *absolute* levels of the factor content of trade. There has yet to be a full resolution to the paradox of the "missing trade" – the troubling finding that the factor content of observed trade is vastly smaller than that predicted from countries' endowments by the Vanek equations (Trefler 1995).

the researcher to evaluate the welfare effects of policy experiments that shift a country's comparative advantage by changing underlying country attributes.

At heart, this empirical specification draws on a growing body of work that identifies comparative advantage from the interaction 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.[3] The literature on institutional determinants of trade has also adopted this empirical strategy. Beck (2003) and Manova (2006) 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.[4] The importance of contract enforcement for the pattern of specialization has also gained attention, motivated by the idea that production is more likely to be held up by input suppliers where enforcement mechanisms are weak.[5] Empirical evidence has indeed confirmed that countries with better institutional rule of law export relatively more from industries that are more exposed to this hold-up problem, as measured by input concentration (Levchenko 2004), or the share of inputs that need to be customized for the final-goods producer (Nunn 2007). On a related note, Costinot (2006) showed that countries with a higher skill endowment, or where institutional transaction costs inhibiting the division of labor are lower, exhibit relatively higher export volumes in sectors where job tasks are more complex. Last but not least, Cuñat and Melitz (2006) demonstrated that countries with flexible labor markets facilitate specialization in industries that experience more volatility and that therefore benefit from being able to adjust employment margins regularly.

The generalization of the EK model presented in Section 2 provides a structural interpretation for the estimation being performed in this growing literature on sources of comparative advantage. The empirical implementation in this paper will bring together a comprehensive set of interaction terms from the papers cited above. Conveniently, the framework lends itself to a natural decomposition of the determinants of trade flows into: (i) Heckscher-Ohlin forces, as identified by interactions between country factor endowments and industry factor intensities; (ii) institutional determinants,

[3]See Baldwin (1971, 1979) for earlier work examining the correlation between industry factor intensities and industry net exports.

[4]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. See also Beck (2002) who found that financially developed countries have larger volumes of manufacturing exports relative to GDP; Wynne (2005) who advanced the idea that wealthier countries have a comparative advantage in credit-constrained industries; and Becker and Greenberg (2005) who showed that financial development facilitates exports, particularly in high fixed-cost sectors. For theoretical work on how credit constraints influence the pattern of trade, see Matsuyama (2005).

[5]See Antràs (2003) and Acemoglu et al. (2007) for theoretical work formalizing the role of contract enforcement as a source of comparative advantage within an incomplete contracting framework.

as measured by interactions between the quality of country institutions and each industry's institutional dependence; and (iii) distance barriers, as controlled for by familiar measures from the gravity literature.[6]

To estimate the underlying parameters of the model, I assemble a large dataset of bilateral trade flows, pairwise distance measures, as well as country and industry characteristics for a sample of 82 countries and 20 manufacturing industries. Section 3 estimates the closed-form trade flow expressions derived from the theory using ordinary least-squares (OLS) methods, to provide a first-pass test of how well the model explains bilateral trade patterns, as well as a basis for comparison with the results in existing work. Here, I find strong corroborating 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 attempt, to the best of my knowledge, at jointly verifying the significance of this comprehensive a list of institutional determinants of trade that have been identified in prior studies.

Although OLS provides a useful baseline, it nevertheless suffers from the drawback that zero trade observations are dropped when log trade flows are the dependent variable. These zeros make up about two-thirds of the dataset, and discarding this sizeable amount of information can systematically bias the OLS coefficients (Helpman et al. 2007, Santos-Silva and Tenreyro 2006, among others).[7] It would thus be inappropriate to use the OLS estimates for evaluating counterfactuals, without first accounting for these zeros. To this end, I modify the model in Section 4 to generate zero trade predictions. I impose a bounded support on the distribution that governs the stochastic component of firm productivity, so that a country with a low systematic productivity level may nevertheless never receive a large enough productivity shock to be able to export a good to a given market. This is a natural step in keeping with the Ricardian spirit of the model, since it attributes the zeros to large cross-country productivity gaps. It does however lead to a complication, which is that we lose closed-form expressions for trade flows. I therefore pursue a simulated method of moments (SMM) procedure to obtain an independent set of parameter estimates, by matching key statistical moments between the actual data and trade flows generated from the model (Pakes and Pollard 1989).

The structural framework adopted enables us to evaluate the relative importance of distance barriers and the various sources of comparative advantage for country welfare. Section 5 undertakes

[6]Davis and Weinstein (2001) found that incorporating productivity differences, trade costs, and non-tradable goods into the traditional Vanek equations goes some way towards reducing the extent of the "missing trade" paradox. This supports the view, implicit in this paper, that empirical work needs to take a more holistic view of the sources of comparative advantage, rather than testing for Ricardian and Heckscher-Ohlin forces in isolation.

[7]Haveman and Hummels (2004) and Anderson and van Wincoop (2004) made the point that traditional formulations of the gravity equation are inconsistent with the presence of zeros in the trade data. Eaton and Tamura (1994) presented an early effort to account for these zero observations via a tobit estimation procedure. Eaton and Kortum (2002) were not affected by this potential bias since their dataset of aggregate OECD manufacturing trade flows contains no zeros.

these counterfactual exercises, using the SMM estimates to calibrate the model. Of note, I find a large average increase in country welfare (25.6%) from a hypothetical move to a world without physical distance barriers, a figure not far from what EK find for their OECD sample (16.1%-24.1%). Beneath this overall increase, however, there is a lot of variation in how individual countries fare, with some even recording a welfare loss due to the diversion of export opportunities to other countries. In particular, developing countries tend to benefit more than the OECD, due to their greater initial distance from developed country markets.[8] I find moreover that this transition to a distance-free world is associated with a moderate increase in country levels of specialization, as measured by the country Herfindahl index of industry production shares; intuitively, countries tend to expand production in their core comparative advantage sectors for which their global market has now expanded. At the same time, from the perspective of each industry, the reduction of distance has a pro-competitive effect, since some countries start exporting from industries that they were previously excluded from due to prohibitive distance barriers. Separately, the SMM estimates imply an average country welfare gain of 6.9% from multilateral integration through the GATT, with the main beneficiaries being countries that were non-members.

I also explore a series of policy experiments associated with raising country attributes, to quantify the importance of different channels of comparative advantage. For example, several developing countries have engaged in large-scale investment in capital accumulation to spur growth, and it would be interesting to evaluate the welfare gains in the context of the global trade equilibrium in this model. I therefore consider increasing each country characteristic in turn to the world frontier level, as approximated by the maximum in my 82-country sample. I compute large gains from physical capital accumulation (an average welfare increase of 42.0%) when all countries are raised to the maximum capital endowment simultaneously. A similarly large increase (37.6%) accrues from worldwide human capital accumulation, with slightly smaller, but nevertheless substantial gains (17.7%) from a global improvement in legal institutions to first-world standards. I find a similar ordering of welfare gains when these policy shocks are applied just to a single developing country – Indonesia – that lies between the 25th and 33rd percentiles for each country attribute. Note that since country characteristics enter the model through an interaction with a corresponding industry variable, the counterfactual changes computed stem strictly from the induced shift in industry and export composition following each policy shock. As this implicitly holds any direct effects on real income constant, these exercises likely understate the magnitude of the total gains to welfare.[9]

[8] Lai and Zhu (2004) found a similar dichotomy in the relative magnitude of country welfare gains from the reduction of distance barriers, within the framework of a monopolistically-competitive bilateral trade model.

[9] This is in contrast to Anderson and Marcouiller (2002), who examined the direct effect that institutional rule of law has on trade flows. The policy counterfactuals in this paper will instead compute the gains from the underlying shift in

This paper falls within a broader research agenda seeking to understand the determinants of industry-level trade flows, often by developing variants of the traditional gravity equation. Several studies have used these models as the basis for welfare counterfactuals, to quantify the effects of moving towards a zero-gravity world (Eaton and Kortum 2002), border effects (Anderson and van Wincoop 2003), and tariff liberalization (Lai and Trefler 2002, Lai and Zhu 2004, Alvarez and Lucas 2006).[10] While this paper also performs similar distance-related counterfactuals, what I gain over and above previous studies is the ability to perform policy experiments involving country characteristics that matter for comparative advantage, as well as the ability to examine the impact on industry structure. It should moreover be stressed that the methodology presented is very general, and is clearly not limited to the specific interactions used in this paper. Any relevant country and industry variables that jointly affect the pattern of trade can in principle be included, subject to the caveat that this will also raise computational cost for the SMM estimation.

The theoretical framework in this paper is most closely related to Costinot and Komunjer (2006), who developed a similar industry-level extension of the EK model, in which firm productivity is systematically driven by the interaction of country and industry characteristics. My specific contribution here is to estimate this model in a manner consistent with the prevalence of zero observations in the trade data, in order to calibrate it for the purpose of evaluating counterfactuals. In this regard, my empirical methodology is most similar to Ramondo (2006), who also employed a simulated method of moments approach to estimate a structural model of multinational activity.[11] Separately, Shikher (2004, 2005) develops an alternative extension of the EK framework to the industry level, in which industries are connected with each other as suppliers of intermediate goods. Empirically, Shikher calibrates country technology parameters for each industry to fit the trade flow data, whereas the approach taken here will be to relate these productivity parameters to observable country and industry characteristics.

The roadmap for the paper is as follows. Section 2 presents an extension of the canonical EK model to explain industry trade flows. Section 3 discusses the baseline results from estimating the derived trade flow equations via OLS. I modify the model in Section 4 to account for the zero trade flows, and re-estimate it with the SMM procedure. I show that these estimates deliver a good fit to

comparative advantage industries that result from improvements in the rule of law.

[10]Lai and Trefler (2002) and Lai and Zhu (2004) worked with a model of monopolistic competition, in contrast to the perfectly-competitive framework in Eaton and Kortum (2002) and Alvarez and Lucas (2006). Note that Lai and Trefler (2002) expressed a healthy reservation about the counterfactuals they computed, as they documented several dimensions, including the expected behavior of price elasticities, along which the benchmark trade model under monopolistic competition appears to be misspecified. See also Hallak (2006) who explored the importance of product quality for explaining bilateral industry trade patterns.

[11]See also Bernard et al. (2003) and Eaton et al. (2005), who use a SMM approach to estimate variants of the EK model using firm-level data from the US and France respectively.

the actual data on several dimensions, including the implied country GDP levels and predicted trade flows in the global trade equilibrium. Section 5 explores various welfare counterfactuals. Section 6 concludes. Details on the data are documented in the Appendix (Section 8).

2 A Benchmark Model of Industry Trade Flows

2.1 The basic set-up

Consider a world with $n = 1, \ldots, N$ countries, in which there are a finite number of industries, indexed by $k = 0, 1, \ldots, K$. Industry 0 denotes non-tradables, which I treat as a homogenous good sector. The tradable sectors ($k \geq 1$) are differentiated products industries, where the continuum of varieties within each industry is indexed by $j^k \in [0, 1]$.[12] 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 products from different industries. Total utility is a Cobb-Douglas aggregate over the consumption of tradables and non-tradables, where the share of income spent on tradables is $\eta \in (0, 1)$.

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

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

where Y_n is total income in country n, and $p_n^k(j)$ is the price in country n of variety j from industry k. (The homogenous good is the domestic numeraire.) Solving this optimization program, it is straightforward to show that the demand for each tradable variety is:

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

[12] I normalize the measure of varieties in each industry to 1.

with $(P_n^k)^{1-\varepsilon} = \int_0^1 (p_n^k(j))^{1-\varepsilon} dj$ being the ideal price index for industry k faced by consumers in country n. The demand for the homogenous good is simply $Q_n^0 = (1-\eta)Y_n$, since consumers spend a fraction $(1-\eta)$ of income on this outside good.

Goods Prices: The market for each variety is perfectly competitive. Firms undertake production using a constant returns to scale technology, so that all firms price at average cost. (There are no fixed costs of entry or production.) Consider the market for supplying an industry-k variety ($k \geq 1$) to country n. All N countries in the world are potential providers 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 first subscript, 'n', identifies the importing country, while the second subscript, 'i', 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. This unit cost is determined by local factor prices, as well as the factor intensities dictated by the baseline technology in this industry. However, countries may possess a productivity edge in executing this baseline technology in the production of specific varieties. This feature is captured by the Ricardian productivity term, $z_i^k(j)$, formally 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.[13] Lastly, the $d_{ni}^k \geq 1$ term captures the unit price mark-up (an iceberg transport cost) caused by distance or geographic barriers that impede trade flows.

For unit production costs, I specify 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.[14] w_{if} is the local unit price of factor f, while $s_f^k \in (0,1)$ is the corresponding 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 firm takes the w_{if}'s as given, being too small to affect aggregate factor markets. Note that the model does not in general imply factor price equalization across countries because of productivity differences and transport cost barriers. These factor price terms will later allow us to assess the importance of Heckscher-Ohlin forces, specifically the role played by endowment-based production cost differences in influencing patterns of industrial specialization.

For the distance mark-up, I adopt the standard assumption that $d_{ni}^k \leq d_{nm}^k d_{mi}^k$ for any three

[13]Thus, the $z_i^k(j)$ term augments productivity in a Hicks-neutral fashion. Trefler (1995) showed that allowing for Hicks-neutral productivity differences across countries improves the ability of traditional Vanek equations to account for the observed factor content of trade.

[14]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. I will subsequently normalize by factor 0 in order to work with relative factor prices.

countries n, m and i. It is thus cheaper to transport goods directly between two countries, rather than through a third country. I allow this iceberg cost to vary by industry, since some goods may be more costly to transport, for example because of their heavier tonnage or industry-specific tariffs.

Productivity: To unpack the sources of comparative advantage, I specify the productivity of country i in producing varieties from industry k 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 productivity shock. The spread parameter β_0 therefore plays a key role in regulating the variance of these productivity shocks.

Importantly, 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. For example, countries where legal institutions securely enforce contracts will on average be more productive in industries that are more vulnerable to hold-up problems between input suppliers and producers (Levchenko 2004, Nunn 2007). Here, these institutional forces augment productivity in a Ricardian fashion, with the β_{lm} coefficients parameterizing how important each channel is for generating a productivity edge. Note that I also include exporter and industry fixed effects (λ_i and μ_k) to control for the average productivity level across all countries and industries respectively.

Finally, I specify the productivity shocks, $\epsilon_i^k(j)$, to be independent draws from the Type I extreme-value (Gumbel) distribution, with cumulative distribution function (cdf) $F(\epsilon) = \exp(-\exp(-\epsilon))$.[15] 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.[16] This specification

[15]More precisely, with a Gumbel distribution for the $\epsilon_i^k(j)$'s, $\lambda_i + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km}$ is equal to the mode of the country i productivity distribution across varieties in industry k. The mean of this distribution is $\lambda_i + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} + \beta_0 \gamma$, where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant, so that mean productivity is also increasing in this systematic component of productivity. The corresponding variance of productivity shocks is $(\beta_0^2 \pi^2)/6$, which is increasing in the spread parameter β_0.

[16]EK's justification for this distributional choice thus carries over to this extension. To recapitulate the micro-foundation EK offered, suppose that firm productivity levels within a country follow a Pareto distribution, an assumption that finds good support in the firm-level data (for example, see Helpman et al. (2004)). Then, the first order statistic for the maximum productivity level across all firms is a Fréchet random variable (see Galombos (1987), p.123 for a proof). It

facilitates 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 for variety j in industry k presented by country i to country n is therefore:

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

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 country i charges. The distribution of productivity shocks gives rise to a distribution of prices, $G_{ni}^k(p)$, presented by country i to country n for each industry-k variety. Applying the expression for the Gumbel cdf for $\epsilon_i^k(j)$ to (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 $\varphi_i^k = \exp\{\theta\lambda_i + \theta\mu_k + \theta \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km}\}$. Note that θ has the interpretation of an inverse productivity spread parameter, while φ_i^k is increasing in the systematic component of country i's productivity in industry k.[17]

2.2 Implications for trade flows

Countries procure each variety from the lowest-price provider, giving rise to the possibility of cross-border trade. Let $p_n^k(j)$ be the price actually paid by country n for variety j from industry k. Since $p_n^k(j) = \min\{p_{ni}^k(j) : i = 1, \ldots, N\}$, the industry-$k$ price distribution facing country n is given by:

$$G_n^k(p) \;=\; 1 - \prod_{i=1}^{N} [1 - G_{ni}^k(p)] \;=\; 1 - \exp\{-(\sum_{i=1}^{N} (c_i^k d_{ni}^k)^{-\theta} \varphi_i^k) p^{\theta}\} \tag{8}$$

Denote by π_{ni}^k the probability of country i being the lowest-price provider – and hence the unique exporter – of an industry-k variety to country n.[18] Following EK, we have:

$$\pi_{ni}^k \;=\; \int_0^{\infty} \prod_{s \neq i} [1 - G_{ns}^k(p)]\; dG_{ns}^k(p) \;=\; \frac{(c_i^k d_{ni}^k)^{-\theta} \varphi_i^k}{\sum_{s=1}^{N} (c_s^k d_{ns}^k)^{-\theta} \varphi_s^k} \tag{9}$$

is this most productive firm at the country's technological frontier that exports to the rest of the world, since it charges the lowest price among all domestic firms. Costinot and Komunjer (2006) showed that distributional assumptions on the productivity shocks can be relaxed to some extent, while retaining the prediction that countries that are on average more productive in an industry will export relatively more from that sector.

[17] Strictly speaking, θ is an average inverse spread parameter that applies across all industries. It is possible to allow θ to vary by industry, so that the productivity distribution in each industry has a different variance, but this will come at the cost of having additional parameters to estimate.

[18] I assume that there are no ties so that there is a unique lowest-price provider for each variety. Since the stochastic terms, $\epsilon_i^k(j)$, are independent draws across varieties, both the price distribution, $G_n^k(p)$, and this probability, π_{ni}^k, do not vary across varieties in the industry.

Notice that the numerator, $(c_i^k d_{ni}^k)^{-\theta}\varphi_i^k$, is precisely the contribution of country i to the term in the denominator.

We can now derive expressions for industry trade flows. Let X_{ni}^k be 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:

$$\frac{X_{ni}^k}{X_n^k} = \frac{\pi_{ni}^k \int_0^\infty \int_0^1 p_n^k(j) Q_n^k(j)\ dj\ dG_n^k(p_n^k)}{\sum_{i=1}^N \pi_{ni}^k \int_0^\infty \int_0^1 p_n^k(j) Q_n^k(j)\ dj\ dG_n^k(p_n^k)} = \pi_{ni}^k = \frac{(c_i^k d_{ni}^k)^{-\theta}\varphi_i^k}{\sum_{s=1}^N (c_s^k d_{ns}^k)^{-\theta}\varphi_s^k} \tag{10}$$

Observe that to evaluate the total value of industry-k consumption in country n, I integrate over varieties j and the minimum price distribution, G_n^k. It can be shown that the distribution of prices in country n conditional on country i being the minimum price provider is given once again by G_n^k. Since this price distribution does not depend on the identity of the exporting country (i), it follows that the fraction of total expenditure in industry k spent on imports from country i is precisely π_{ni}^k. This implies the closed-form (10), which expresses i's industry-k market share in country n as a function of bilateral distance, as well as underlying country and industry characteristics.

It will be useful to eliminate the denominator term by normalizing (10) by country n's expenditure share from a fixed reference country, u:

$$\frac{X_{ni}^k}{X_{nu}^k} = \frac{(c_i^k d_{ni}^k)^{-\theta}\varphi_i^k}{(c_u^k d_{nu}^k)^{-\theta}\varphi_u^k} \tag{11}$$

This last equation for (normalized) trade flows will serve as the basis for the OLS estimation in Section 3 below.[19] Equation (11) has an intuitive economic interpretation: The share of country n's market serviced by country i relative to that held by country u is decreasing in both i's relative unit cost of production (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 (captured by φ_i^k/φ_u^k).

Finally, consider the role played by the inverse spread parameter, θ. Observe that (11) 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 $\tilde{\varphi}_i^k = \exp\{\lambda_i + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km}\}$. It is convenient to interpret $(c_i^k d_{ni}^k/\tilde{\varphi}_i^k)$ as an "average" price for industry-k varieties presented by country i to country n. For the sake of illustration, 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 ($\frac{X_{ni}^k}{X_{nu}^k} < 1$). Now, a higher θ will further shrink i's relative share, so that a smaller spread in the productivity shocks biases market shares towards the *a priori* larger exporters, leading to a greater concentration of imports by source

[19] Although equation (10) also holds for $i = n$, I opt for a normalization relative to a fixed country, rather than by each country's domestic absorption (X_{nn}^k). Data on domestic production disaggregated at the industry level are available for a more limited set of countries. Also, attempts to merge UNIDO data on domestic output with the Feenstra et al. (2005) trade flows often led to implied values for X_{nn}^k that were negative, which likely reflect differences in measurement concepts for the two datasets, or imperfect concordances between the SITC and ISIC classification systems (used by Feenstra et al. (2005) and UNIDO respectively).

in country n. This feature stems from the fact that the Gumbel distribution has a thick right tail: A large spread parameter (low θ) increases the likelihood that a country with low average productivity will nevertheless get a good enough productivity shock in some varieties to emerge as the lowest-price provider. Conversely, a decrease in the variance of these productivity shocks hurts the market share of smaller exporting countries.

2.3 Comparison with Eaton and Kortum (2002)

At this juncture, it is useful to highlight the close links between the expressions for trade flows and those in Eaton and Kortum (2002), to point out the exact nature of this extension of their framework. 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 $F_i(z) = \exp(-T_i z^{\theta})$. Note that $T_i > 0$ is a country-specific location parameter (that is increasing in the technological position of the country), and $\theta > 1$ is an inverse spread parameter for the $z_i(j)$'s. (The 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:

$$\left(\frac{X_{ni}}{X_n}\right)^{EK} = \frac{(c_i d_{ni})^{-\theta} T_i}{\sum_{s=1}^{N}(c_s d_{ns})^{-\theta} T_s} \tag{12}$$

which is precisely EK's equation (10). It follows that trade flows normalized with respect to the reference country u are:

$$\left(\frac{X_{ni}}{X_{nu}}\right)^{EK} = \frac{(c_i d_{ni})^{-\theta} T_i}{(c_u d_{nu})^{-\theta} T_u} \tag{13}$$

These are clearly direct analogues of equations (10) and (11) respectively: Both sets of equations explain trade shares as a function of factor costs, distance barriers and productivity differences in a similar way, except that each term has been replaced with its industry-specific counterpart. In particular, the more general productivity term, φ_i^k, now takes the place of EK's technological parameter, T_i. This highlights the sense in which this paper directly unpacks the sources of Ricardian comparative advantage, by positing a functional form to tie down productivity to observable characteristics that reflect how well countries are able to meet the requirements of industries along various technological and institutional dimensions.[20]

I defer to Section 4 the discussion of how to close the model formally to solve for country income levels as an endogenous outcome of the global trade equilibrium. Instead, I focus first on empirically testing the trade flow equation, (11).

[20] This model nevertheless shares one potentially restrictive feature with EK, which can be seen from equations (10)-(13): The identity of the importing country n affects exporters' market shares only through n's bilateral distance from each exporter. Among other things, this rules out Armington preference biases between countries that might be relevant in practice. I leave an exploration of such avenues to future work.

3 OLS Estimation of Bilateral Industry Trade Flows

This section undertakes an OLS analysis to test how well the model explains the pattern of observed bilateral trade at the industry level. This exercise serves two purposes. First, the regression specifications follow closely those adopted in existing empirical work on sources of comparative advantage, and so the OLS results provide a basis for comparison and corroboration with the current literature. The large dataset assembled also allows me to jointly test the significance of a comprehensive list of institutional determinants of trade flows, a first to the best of my knowledge. Second, the OLS results serve as a baseline for initializing the SMM estimation in Section 4, which will account for the zero observations omitted from the OLS regressions.

3.1 Deriving the estimating equation

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

$$d_{ni}^k = \exp\{\beta_d D_{ni} + \delta_k + \zeta_{ni} + \nu_{ni}^k\} \tag{14}$$

Here, $\beta_d D_{ni}$ is a linear combination of distance variables that impose an iceberg transport cost on trade. In the regressions below, these D_{ni}'s will include measures of physical distance, shared linguistic ties, colonial links, border relationships, as well as indicator variables for trade agreements that reduce policy barriers to trade.[21] (Details on the distance measures are contained in the Data Appendix.) I allow the bilateral mark-up to vary by industry through the fixed effect, δ_k, since transport costs may be higher for particular sectors. Finally, transactions between countries may be subject to idiosyncratic shocks, $\zeta_{ni} + \nu_{ni}^k$, which includes a country-pair specific component (ζ_{ni}); I assume that these are iid draws from the following mean-zero normal distributions: $\zeta_{ni} \sim N(0, \sigma_\zeta^2)$, and $\nu_{ni}^k \sim N(0, \sigma_\nu^2)$.

Taking logarithms of (11), and substituting in the distance term (14), we obtain:

$$\begin{aligned} \ln\left(\frac{X_{ni}^k}{X_{nu}^k}\right) \;\; = \;\; & -\theta \sum_{f=0}^{F} (\ln w_{if} - \ln w_{uf})\, s_f^k \; + \; \theta \sum_{\{l,m\}} \beta_{lm}(L_{il} - L_{ul}) M_{km} \;\ldots \\ & \ldots - \; \theta\beta_d(D_{ni} - D_{nu}) + \theta(\lambda_i - \lambda_u) \; - \theta(\zeta_{ni} - \zeta_{nu}) \; - \theta(\nu_{ni}^k - \nu_{nu}^k) \end{aligned} \tag{15}$$

Throughout this exercise, the reference country 'u' will be the United States.

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

The term $\theta \sum_f (\ln w_{if} - \ln w_{uf})\, s_f^k$ on the right-hand side of (15) consists of industry factor shares (s_f^k) interacted with country characteristics (the factor prices, w_{if}). This is thus identical in spirit to the $\theta \sum_{\{l,m\}} \beta_{lm}(L_{il} - L_{ul})M_{km}$ term, in that both capture how well conditions in country i provide for the production needs of industry k. Written in this form, we have a natural decomposition: The $\theta \sum_f (\ln w_{if} - \ln w_{uf})\, s_f^k$ term picks up the role of Heckscher-Ohlin forces in determining which industries a country possesses an endowment-based cost advantage in (akin to Romalis (2004)), while the latter $\theta \sum_{\{l,m\}} \beta_{lm}(L_{il} - L_{ul})M_{km}$ term identifies sources of comparative advantage stemming from a country's ability to provide the right institutional and technological conditions for the industry.

Since $s_0^k = 1 - \sum_{f=1}^F s_f^k$ by the constant returns to scale assumption, (15) can be re-written as:

$$\begin{aligned} \ln\left(\frac{X_{ni}^k}{X_{nu}^k}\right) &= -\theta \sum_{f=1}^F \left(\ln \frac{w_{if}}{w_{i0}} - \ln \frac{w_{uf}}{w_{u0}}\right) s_f^k \; + \; \theta \sum_{\{l,m\}} \beta_{lm}(L_{il} - L_{ul})M_{km} \; \ldots \\ & \qquad \ldots - \theta\beta_d(D_{ni} - D_{nu}) \; + I_i \; + I_{nk} \; - \theta\zeta_{ni} - \theta\nu_{ni}^k \qquad (16) \end{aligned}$$

which allows us to work instead with *relative* factor prices. Note that the constants specific to exporting country i are collected into the term, I_i, while $\theta\zeta_{nu} + \theta\nu_{nu}^k$ has been re-written as an importer-industry fixed effect, I_{nk}. Conveniently, $-\theta\nu_{ni}^k \sim N(0, \theta^2\sigma_\nu^2)$ is an iid noise term. In practice, however, good data on factor prices is not readily available for a large sample of countries, so I proxy for relative factor prices, $\ln \frac{w_{if}}{w_{i0}}$, by treating them as an inverse function of relative factor endowments, $\ln \frac{V_{if}}{V_{i0}}$, where V_{if} denotes country i's endowment of factor f.[22] I set unskilled labor to be factor 0.

The above discussion implies the following estimating equation:

$$\begin{aligned} \ln\left(\frac{X_{ni}^k}{X_{nu}^k}\right) &= \sum_{f=1}^F \theta\beta_f \left(\ln \frac{V_{if}}{V_{i0}} - \ln \frac{V_{uf}}{V_{u0}}\right) s_f^k \; + \sum_{\{l,m\}} \theta\beta_{lm}(L_{il} - L_{ul})M_{km} \ldots \\ & \qquad -\theta\beta_d(D_{ni} - D_{nu}) \; + \; I_i \; + \; I_{nk} \; - \; \theta\zeta_{ni} \; - \; \theta\nu_{ni}^k \qquad (17) \end{aligned}$$

I therefore regress log trade flows from country i to n in industry k (normalized by the corresponding US export value) as a function of: (i) Heckscher-Ohlin determinants, identified by the interaction between relative factor endowments in country i, $\left(\ln \frac{V_{if}}{V_{i0}} - \ln \frac{V_{uf}}{V_{u0}}\right)$, and industry factor intensities, s_f^k; (ii) institutional determinants, picked up by the interaction between institutional characteristics, $(L_{il} - L_{ul})$, and industry measures of dependence, M_{km}; (iii) bilateral distance variables, $(D_{ni} - D_{nu})$; (iv) exporter fixed effects; and (v) importer-industry fixed effects. All relevant exporter characteristics and distance measures are taken relative to the corresponding values for the US. Standard errors are clustered by country pair, to account for the mean zero shocks, $-\theta\zeta_{ni}$.

[22] Deardorff (1982) provided a very general proof that there is a negative correlation between factor prices and the factor content of net exports. This helps justify substituting for factor prices as an inverse function of country factor endowments, insofar as the factor content of net exports is positively correlated with endowments. On this latter point, Debaere (2003) provided supporting evidence that countries are indeed net exporters of their relatively abundant factors in the case of North-South bilateral trade, and in trade between countries with very different endowment mixes.

It should be pointed out that (17) does not allow us to identify the inverse spread parameter, θ, or the industry-specific effect, $\mu_k + \delta_k$. This is a limitation of the OLS approach, since without estimates for these parameters, one cannot proceed to evaluate welfare counterfactuals. For θ in particular, EK present a wide range of estimates from 2.44 up to 12.86, depending on whether wage or goods price equations are used, as well as on the estimation procedure (OLS or instrumental variables). While OLS regressions of (17) will not help us to pin down θ, the SMM approach later in Section 4 will allow us to estimate this key parameter by utilizing information from a rich set of data moments.

3.2 Discussion of OLS regression results

I assemble a large dataset on bilateral trade flows, country and industry characteristics, as well as distance measures for the empirical exercise. For the differentiated products industries ($k \geq 1$), I work with the US 1987 Standard Industrial Classification (SIC-87) 2-digit manufacturing categories, a fairly broad level of industry aggregation. This provides 20 industry groups, listed in Table 1A, with SIC codes from 20 (food processing) to 39 (miscellaneous manufacturing).

The sample consists of 82 countries (listed in Table 1B), the largest number for which a complete dataset of all the variables could be assembled. I draw on the World Trade Flows database (Feenstra et al. 2005) for the dependent variable. The original trade data are in the Standard Industrial Trade Classification (SITC), Revision 2 format; I convert these to SIC-87 categories using detailed information on the composition of US exports to derive concordance weights.[23] The analysis focuses on one snapshot in time, 1990, which is the same year as in EK. This therefore abstracts from dynamic issues such as exchange rate fluctuations or factor accumulation over time.

The dataset contains 79.7% of all recorded manufacturing trade in 1990. While the total number of data points is $82 \times 81 \times 20 = 132,840$, only 43,404 (or 32.4%) of these observations record a positive amount of trade. This pervasiveness of zeros is a feature even of aggregate trade flows (as documented by Helpman et al. (2007)), and it presents a challenge to consistent estimation of gravity equations. Here, observations are dropped from an OLS regression of (17) when either the numerator (X_{ni}^k) or denominator (X_{nu}^k) is zero.[24] These OLS estimates can nevertheless be interpreted as effects conditional on observing positive trade flows into a country from both exporter i and the US. The empirical literature cited in the Introduction has generally taken this approach (rather than attempting to correct for the bias from discarding the zeros), and so the regressions run here provide a basis for comparison with this body of work.[25] These OLS estimates will also provide a baseline with which to

[23] This procedure follows Cuñat and Melitz (2006), with the composition of US exports calculated from Feenstra et al. (2002). Please see the Data Appendix for more details.

[24] Even with the US as the exporter (the X_{nu}^k's), 164 out of the 1,640 possible data points are zeros.

[25] One exception is Manova (2006) who decomposed the role of financial development in explaining both the extensive

initialize the SMM estimation procedure in Section 4.

Table 2 presents the results from OLS regressions of (17). As a reminder, all specifications include exporter and importer-industry fixed effects, with standard errors clustered by country pair. The discussion below focuses on the economic intuition behind the results, with the description of the variables being more brief; the Data Appendix documents in detail how the data variables were gathered and standardized to SIC-87 2-digit format. These variables have been drawn directly from or constructed following closely the methodology from existing studies, in order to facilitate comparison with the literature. As far as possible, I have used country and industry variables from the immediate years preceding 1990 to explain trade flows in that year; when multiple years of data are available, I have used averages over 1980-89 to help smooth out the effects of idiosyncratic noise in the data for any single year. Summary statistics and a correlation matrix of the industry and country characteristics can also be found in the Data Appendix.

Distance and Geography ($D_{ni} - D_{nu}$): Column (1) reports a basic specification that includes only gravity-type variables that control for distance barriers. The primary measure is a great circle distance between countries' major population centers. This is supplemented by dummy variables that proxy for linguistic, historical and socio-cultural proximity: (i) whether the countries share a common language; (ii) whether one country had ever colonized the other; (iii) whether both shared a common colonizer post-1945; and (iv) whether the countries share a land border. Two trade policy dummy variables from Rose (2004) are also included, namely whether the countries were: (i) common members in a regional trade agreement (RTA); and (ii) GATT signatories.[26]

The results in Column (1) confirm the importance of distance barriers for explaining the pattern of trade. The distance coefficients display the signs that one would expect *ex ante*, although not all are statistically significant. Of note, physical distance has a negative and highly significant effect ($\beta_{d1} = -1.12$) on (normalized) trade flows. The magnitude of this coefficient implies large effects: A hypothetical halving of physical distance between two countries would be expected to slightly more than double the amount of bilateral trade (increasing it by a factor of $(0.5)^{-1.12} = 2.17$). Similarly, sharing a common language ($\beta_{d2} = 0.71$) or a colonial relationship ($\beta_{d3} = 0.42$) both raise the propensity for trade between countries. While the border effect is positive, this is not statistically significant; likewise, the effect of a common colonizer is estimated imprecisely. Joint membership in an

and intensive margins of trade using the two-stage procedure in Helpman et al. (2007), where the first-stage consists of a selection equation for the probability of observing a positive amount of trade.

[26]These dummy variables are admittedly imperfect proxies for trade policy. One would ideally like to have an industry-specific measure of the tariffs levied between each country pair. Such data is unfortunately not available for a large set of countries in a consistently measured way. See Kee et al. (2006) for a step towards constructing indices of tariff and non-tariff barriers for a large set of countries.

RTA delivers a significant boost to bilateral trade, raising trade flows by a factor of exp(0.48) = 1.62. As in Rose (2004), I do not find a significant GATT effect with OLS over and above the trade-promoting effect of being in a common RTA. Importantly, the distance variables and fixed effects already account for a sizeable fraction (63%) of the observed variance in trade flows. These distance coefficients will remain remarkably stable even when more explanatory variables are included below.

Heckscher-Ohlin determinants ($\left(\ln \frac{V_{if}}{V_{i0}} - \ln \frac{V_{uf}}{V_{u0}}\right) s_f^k$): Column (2) adds the interaction terms pertaining to the role of factor endowments to the baseline specification. I consider $F = 3$ factors of production: human capital, physical capital, and raw materials, all measured relative to the endowment of unskilled labor. The measures of human capital per worker ($\log(H/L)$) and physical capital per worker ($\log(K/L)$) are from Hall and Jones (1999), for the year 1988. Raw materials abundance is proxied by forest land per worker ($\log(Forest/L)$) and arable land per worker ($\log(Arable/L)$). Each of these relative factor endowments is interacted with a factor intensity term, s_f^k. In keeping with the theory, these are calculated as each factor's share of total factor payments in SIC industry k, based on information on US manufacturing from the NBER-CES dataset (Bartelsman et al. 2000).[27] Skill intensity, s_h, is calculated as non-production payroll divided by the value of total shipments over the period 1980-89. Similarly, materials intensity, s_m, is equal to raw materials cost divided by total shipments. For s_k, I treat factor payments to physical capital as a residual after subtracting total payroll and raw materials cost from total shipments.

The regression results in Column (2) highlight the relevance of Heckscher-Ohlin forces for the cross-country pattern of trade. In particular, countries which are more skill abundant exhibit higher volumes of bilateral trade in industries that are skill-intensive ($\beta_{f1} = 38.04$, significant at the 1% level). Similar effects are identified for physical capital ($\beta_{f2} = 1.77$) and materials abundance ($\beta_{f4} = 1.23$), although the interaction involving forest land is not estimated precisely. Compared with Romalis' (2004) findings which were based on US import data at a finer level of industrial classification (the SIC-87 4-digit level), Column (2) corroborates his results for a large bilateral sample of countries at a broader level of industry aggregation.

Institutional determinants ($(L_{il} - L_{ul})M_{km}$): The rest of Table 2 examines five hypotheses on institutional sources of comparative advantage advanced in the recent literature, which were reviewed in the Introduction. I find here broad support in the bilateral data for the role played by institutional forces in promoting specialization in industries that are particularly dependent on these features of

[27] All the industry-specific variables used in this paper are constructed from US datasets. This is a matter of necessity, since systematic industry-level data is not available for a large sample of countries. The regression results will nevertheless remain valid insofar as industry characteristics are highly correlated across countries.

the institutional environment. I discuss each of these hypotheses in turn:

(i) External Capital Dependence: Column (3) examines the role of country financial development as a source of comparative advantage. I compute a measure of external capital dependence ($CAPDEP$) – equal to the fraction of capital expenditures not funded by internal cashflow – from Compustat for the SIC-87 2-digit industries, following the methodology of Rajan and Zingales (1998). This is interacted with the amount of private credit divided by GDP in each country ($FINDEV$), from Beck et al. (2000), to capture how well countries can provide for the financing needs of industries. I obtain a positive and highly significant coefficient ($\beta_{lm1} = 1.78$) on this interaction term, confirming the intuition that financially-developed countries export more in industries that are more reliant on external capital funding (Beck 2003, Manova 2006). At the same time, the Heckscher-Ohlin coefficients remain very similar to those in Column (2), suggesting that $CAPDEV \times FINDEV$ is identifying an independent determinant of the pattern of bilateral trade.

(ii) Input Concentration: Column (4) turns to the role played by the contracting and legal environment in facilitating production. Levchenko (2004) argued that industries heavily reliant on a few key inputs are more vulnerable to hold-up problems from suppliers, and are hence more dependent on a credible legal system to enforce contracts. Following his empirical strategy, I compute input concentration as the Herfindahl Index of intermediate input use (HI) based on the 1987 US Input-Output Use Tables. I interact this with a 1985 index measure of the strength of the legal system ($LEGAL$) developed by the Fraser Institute (Gwartney and Lawson 2004). Column (4) finds a positive and significant coefficient for this particular mechanism ($\beta_{lm2} = 3.87$), suggesting that countries with stronger legal systems are in a better position to specialize in goods with a high input concentration.[28]

(iii) Input Relationship-Specificity: Nunn (2007) expanded on this incomplete contracting logic by proposing a more refined measure of the extent to which hold-up problems with input suppliers can affect production. Using a classification from Rauch (1999), Nunn distinguished between generic inputs (for which substitutes are readily available on the open market) and those that are relationship-specific (which require suppliers to make costly investments to customize the input for the final-goods producer). Input relationship-specificity (RS) is then calculated as the fraction of intermediate inputs (by value) that cannot be procured on an organized exchange. I obtained the RS measure based on the 1987 US Input-Output Use Tables from Nunn (2007), and interacted this with $LEGAL$ in Column (5), to assess the importance of legal institutions in promoting exports in industries that require a greater share of relationship-specific inputs. I estimate this effect to be positive and significant ($\beta_{lm3} = 2.78$),

[28]Another popular index of institutional strength, from the World Bank Governance Indicators (Kaufmann et al. 2005), is available only from 1996 onwards. The results are similar if I use their "rule of law" index for 1996 instead, reflecting the high persistence in institutional conditions over time in most countries.

confirming the importance of contract enforcement institutions as a source of comparative advantage.[29]

(iv) Job Complexity: Costinot (2006) analyzed how differences across industries in the complexity of production tasks affect the cross-country pattern of specialization. Intuitively, countries with a higher skill endowment should command a productivity edge in more complex industries, as skilled workers can in principle perform more tasks in a given span of time. In addition, transactions costs can hinder the division of labor among work teams, so that countries where such frictions are exogenously lower will have an advantage in the more complex industries. Column (6) confirms that both of these mechanisms are operative in the bilateral data. I adopt Costinot's measure of job complexity ($COMPL$), which is drawn from a US Panel Survey of Income Dynamics (PSID) question that asks respondents to estimate how long a typical new employee would take to become "fully trained and qualified" in their job. As in Costinot (2006), the $COMPL$ variable yields a positive coefficient ($\beta_{lm4} = 7.16$, significant at the 1% level) when interacted with $LEGAL$. This therefore identifies yet another channel through which the legal system serves as a source of comparative advantage, by helping to enforce contracts that facilitate the division of labor.[30] I also obtain a positive significant coefficient ($\beta_{lm5} = 1.54$) when interacting $COMPL$ with countries' skill endowment ($log(H/L)$), so that countries with a skilled workforce are better-placed to specialize in industries where job tasks are more complex.

(v) Volatility: The final column in Table 2 considers how labor market flexibility can facilitate specialization in industries that experience high levels of idiosyncratic volatility. Countries where it is easier to hire workers in good times and lay them off when there is excess capacity should be better-placed to attract industries that are inherently more volatile, and hence benefit most from being able to make frequent adjustments on their employment margin. Column (7) uses the measure of annual sales volatility ($SVOL$) from Cuñat and Melitz (2006), calculated from firm data in Compustat, and interacts this with an index of country labor market flexibility ($FLEX$) from the World Bank. Consistent with Cuñat and Melitz (2006), I obtain a positive coefficient ($\beta_{lm6} = 16.05$, significant at the 1% level), confirming that countries with flexible labor institutions tend to specialize and export more in volatile industries.

[29]Nunn's (2007) results, based on the 1997 US Input-Output Use Tables, are robust to alternative ways of measuring RS. The results in Tables 2 and 3 using the earlier 1987 Tables are more sensitive to how RS is constructed: I obtain a positive and significant β_{lm3} coefficient only when the most liberal criteria for classifying input relationship-specificity are used. In particular, the results weaken if RS is constructed using the more conservative classification from Rauch (1999), or if I use a stricter criterion that treats inputs as relationship-specific only if they are both not sold on an organized exchange and are not reference-priced in trade journals.

[30]Costinot's (2006) preferred measure of the transaction costs that hinder cooperation among work teams is a measure of trust or social capital, calculated from the World Values Survey (WVS). I use $LEGAL$ as my proxy instead because the WVS measure is available for a smaller set of countries (50).

Table 3 shows that virtually all the above findings are unaffected when I run these institutional determinants jointly in a single specification (Column (1)). All but one of the institutional coefficients remain significant, with only the interaction between input concentration and $LEGAL$ (β_{lm2}) dropping out of statistical significance. This suggests that the empirical literature has successfully identified largely independent channels through which country attributes explain the pattern of specialization. Separately, the Heckscher-Ohlin coefficients on human capital (β_{f1}) and physical capital (β_{f2}) diminish in magnitude; both remain positive, although that on physical capital turns insignificant.[31] To provide a sense of the relative importance of these explanatory variables, Column (1a) reports the standardized beta coefficients based on Column (1).[32] Of note, physical distance is the single most influential determinant of trade flows with the largest beta coefficient ($\beta_{d1} = -0.35$). The Heckscher-Ohlin and institutional beta coefficients are smaller (ranging between 0.07 to 0.21, when statistically significant), although their total effect is nevertheless collectively larger than that of physical distance (the sum of the beta coefficients for the significant interactions exceeds that for physical distance).

One can further illustrate the quantitative implications of the interaction coefficients through the following thought experiment: *Ceteris paribus*, how much larger does the model predict trade flows would be for the exporter at the 75th percentile of the human capital distribution in the industry at the 75th percentile of the skill-intensity distribution, relative to the 25th percentile country in the 25th percentile industry? The interquartile gap in the human capital distribution is 0.41, while the corresponding gap for the skill-intensity distribution is 0.036. The Column (1) estimate of β_{f1} then implies that (normalized) trade flows would rise by a sizeable factor of $\exp(15.37 \times 0.41 \times 0.036) = 1.25$, or about a 25% increase, when moving from the 25th percentile country and industry to the 75th percentile. Performing similar calculations for the other interaction terms, I find the largest percent increase in normalized trade flows from this 25th-to-75th percentile exercise with the interaction between job complexity and legal institutions ($COMPL \times LEGAL$, a 50% increase), followed by the relationship-specificity term ($RS \times LEGAL$, 26%) and the job complexity-human capital interaction ($COMPL \times \log(H/L)$, 26%).

Column (2) restricts the regression to the EK sample of 19 OECD countries. The Heckscher-Ohlin interaction for human capital is now not estimated precisely, although the corresponding coefficient for physical capital becomes statistically significant.[33] Also, the effect of relationship-specificity has

[31] I have also experimented with a more conventional set of fixed effects, namely using separate importer, exporter and industry dummies. The regression results do not change much, although the R^2 drops from 0.653 to 0.575 due to the smaller number of dummy variables (results available upon request).

[32] The standardized beta coefficient is the regular OLS coefficient multiplied by the ratio of the standard deviation of normalized trade flows to the standard deviation of the relevant explanatory variable. The betas therefore 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.

[33] The common colonizer and GATT dummy variables are dropped because there is no variation in these two measures

switched signs (β_{lm3}). This suggests that several of the mechanisms generating comparative advantage are only clearly identified in the broader sample, where there is a larger variance in country attributes such as human capital and the quality of the legal system.

At this point, it should be stressed that I do *not* view the regressions in Table 3 as a canonical model to explain trade flows. While I have attempted to be comprehensive in including as many trade determinants as possible, the model from Section 2 is certainly more general in that it can accommodate an arbitrary number of economically-relevant interaction terms. The specific implementation in this paper is intended instead to illustrate the methodology for quantifying the relative importance of different sources of comparative advantage.[34]

The above regressions confirm the usefulness of the model for explaining the intensive margin of trade, namely conditional on observing positive trade flows. However, a key concern is that two-thirds of the bilateral trade observations are zeros and these have been dropped from the OLS regressions. Column (3) confirms that the same set of trade determinants has a lot of explanatory power for the extensive margin of trade as well, by running a probit regression based on (17). (The dependent variable is now an indicator equal to 1 if (X_{ni}^k/X_{nu}^k) is positive, and 0 otherwise.) I find that physical distance has a significant effect in deterring trade completely between country pairs ($\beta_{d1} = -0.64$, significant at the 1% level). Also, all but two of the Heckscher-Ohlin and institutional interactions have a positive and significant effect in predicting positive trade flows. These probit results clearly suggest that the OLS estimates do not provide the full picture. For example, the input concentration channel (β_{lm2}) is a significant predictor of the extensive margin of trade, even though the OLS coefficient would lead to the conclusion that this has a negligible effect on the magnitude of trade.

As a consequence, it would be inappropriate to use the OLS estimates as the basis for a welfare exercise, without accounting for the coefficient bias from dropping the zeros. In this regard, a simple tobit regression is unlikely to be fully satisfactory, since this assumes that the zeros arise because of censoring of the trade data.[35] One view here is that the data for trade between less-developed countries tends to be of poorer quality, with many of these observations set equal to 0 by default due

within the OECD sample.

[34]That several of the coefficients are not statistically significant should not be viewed necessarily as a refutation of prior empirical work which found significant effects for these variables. In particular, the regressions in this paper are for a relatively high level of industry aggregation compared to other studies, with just 20 industry categories.

[35]Using a tobit regression with a flexible left-censoring value, I obtain estimates that differ minimally from the OLS results in Column (1). I have also experimented with adding one US dollar to both numerator and denominator trade flows, but this is once again premised on the zeros being small volumes that have been rounded down. This dependent variable also leads surprisingly to negative coefficients on the Heckscher-Ohlin interactions involving skill and physical capital abundance. Another procedure for correcting the zeros bias is suggested by Santos-Silva and Tenreyro (2006), who pointed out that the data-generating process for the dependent variable can be viewed as a pseudo-Poisson process whose parameters can be estimated via maximum likelihood. I choose to pursue the SMM approach instead, since this estimation procedure emerges naturally from the modelling framework in Section 2. The results from these alternative procedures for dealing with the zero trade flows are available upon request.

to a lack of reporting. Nevertheless, I find that removing the countries with the lowest income per capita levels from the sample has little effect on the OLS results in Column (1) (results available upon request).[36] While it is possible that some of the zeros might be due to the rounding-down of small volumes of trade, the fact that more than half the dataset consists of zeros suggests that there are systematic forces at play inhibiting trade flows.

4 Estimation by Simulated Method of Moments (SMM)

This section modifies the model from Section 2 to enable it to deliver zero trade flow predictions. In keeping with the Ricardian nature of the model, the approach I adopt is to view the zero trade flows as arising from large productivity gaps between countries, which prevent low productivity countries from exporting to particular markets. I then re-estimate the underlying parameters by matching moments of trade flows simulated from the model with the corresponding moments from the actual data.

4.1 Modifying the theory to generate zeros

In its present form, the model from Section 2 precludes any zero trade predictions, since it establishes that each country i has a strictly positive probability of being the lowest-price supplier to any country n of a given industry-k variety (equation (9)). Suppose instead that the productivity shocks, $\epsilon_i^k(j)$, are now 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 $F(\epsilon) = \exp(-\exp(-\epsilon))$ is the standard Gumbel distribution. The bounded support now makes zero predicted trade possible: X_{ni}^k will be equal to 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:

$$\lambda_i + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} + \beta_0 \bar{x} < \lambda_{i'} + \mu_k + \sum_{\{l,m\}} \beta_{lm} L_{i'l} M_{km} + \beta_0 \underline{x}$$

In contrast, under the previous specification of a productivity distribution with infinite support, there would have been a positive probability (albeit possibly tiny) of trade between every country pair in each industry, since even countries with a poor systematic component of productivity stood a chance of obtaining a large enough productivity shock to become the lowest-price provider of at least one variety. Truncating the productivity distribution therefore represents a minimal extension of the model that generates zero trade flows, without having to introduce further features such as

[36] Specifically, I have experimented with removing the 12 poorest countries, as measured by GDP per capita in 1990, all of which are African countries. These 12 countries are a natural cut-off, since the 13th poorest country was India, a large developing country. The results are robust to removing a smaller number of countries.

fixed cost barriers.[37] I shall thus keep to the spirit of the EK framework, to explore how well this perfectly-competitive Ricardian benchmark model can go towards explaining the trade data.

4.2 SMM estimation procedure

With the bounded support on the productivity distribution, we lose closed-form expressions for trade flows. Nevertheless, the model can be estimated via a simulated method of moments (SMM) procedure, by finding parameter values that deliver predicted trade flows which match as closely as possible key statistical moments of the actual data (Pakes and Pollard 1989). This is similar to the approach taken by Ramondo (2006), who also faces complications estimating the structural parameters of her model of FDI due to the prevalence of zeros in her dataset of multinational affiliate activity. To implement this procedure, 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 in (6) and substituting in the distance and factor endowment variables, the log price of each variety in industry k is given by:

$$\ln(p_{ni}^k)^{(j)} = \frac{1}{\theta}\left(\theta\beta_d \cdot D_{ni} - \sum_{f=1}^{F} \theta\beta_f \cdot s_f^k \ln\frac{V_{if}}{V_{i0}} - \sum_{\{l,m\}} \theta\beta_{lm} \cdot L_{il}M_{km} + \tilde{I}_i + \tilde{I}_k - (\epsilon_i^k)^{(j)}\right) \tag{18}$$

Here, $(\epsilon_i^k)^{(j)}$ is a random draw from the truncated Gumbel distribution with support $[\underline{x}, \bar{x}]$, while $\tilde{I}_i = \lambda_i$ and $\tilde{I}_k = \mu_k + \delta_k$ group together the various exporter and industry fixed effects respectively. For any given realization of the parameter values, the steps for simulating a full set of bilateral industry trade flows are as follows:

1. For each variety j in industry k, compute the prices presented by all N countries to each importing country n using (18). This requires taking $N \times K \times J$ independent draws from the truncated Gumbel distribution for the productivity shocks, $(\epsilon_i^k)^{(j)}$.

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

[37] Helpman et al. (2007) view the presence of fixed costs to exporting as the key obstacle giving rise to the zero observations. Estimation then proceeds via a first-stage selection equation that determines the probability of observing positive trade between pairs of countries. I do not pursue this approach partly because data on how the fixed cost of exporting varies by country and by industry are not readily available. Note also that fixed costs alone are insufficient to generate zero trade predictions. One still needs to impose a productivity distribution with bounded support, to ensure that countries with low systematic productivity levels will never receive a large enough productivity shock to overcome the fixed cost barrier.

3. Calculate the approximate ideal price indices:

$$(P_n^k)^{1-\varepsilon} \approx \frac{1}{J} \sum_{j=1}^{J} ((p_{n,i(j)}^k)^{(j)})^{1-\varepsilon} \tag{19}$$

4. Using the ideal price indices from (19) and data on country GDP (from the World Development Indicators (WDI)), calculate the quantity demanded, $(Q_{n,i(j)}^k)^{(j)}$, for each variety in country n using the formula in (3).

5. Compute the value of exports from country i to n in industry k by summing over the relevant exporter subscripts:

$$(X_{ni}^k)^{sim} = \frac{1}{J} \sum_{\{j:\ i(j)=i\}} (p_{n,i(j)}^k)^{(j)} (Q_{n,i(j)}^k)^{(j)} \tag{20}$$

In practice, however, the number of fixed effects to be estimated in (18) is large and could potentially strain the reliability of conventional minimization algorithms.[38] To reduce the number of dimensions of the parameter vector, I proxy for the exporter fixed effects using country income per capita by setting: $\tilde{I}_i = c_1(\ln(Y_i/L_i))^{\gamma_1}$, positing that the country characteristics that determine average export levels are highly correlated with aggregate productivity as measured by income per capita. A reasonable prior is that $c_1 < 0$ and $\gamma_1 > 0$, so that the prices presented by richer countries tend to be lower, consistent with the observation that developed countries engage in more exporting. Similarly, I proxy for the industry fixed effects by $\tilde{I}_k = c_2(\ln(T_k))^{\gamma_2}$, where T_k is the total volume of world trade in industry k. Once again, I hypothesize that $c_2 < 0$ and $\gamma_2 > 0$, so that a lower price is consistent with a larger volume of trade in a given industry.

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

$$\Theta = \{\underline{x}, \bar{x}, \theta, \eta, c_1, \gamma_1, c_2, \gamma_2, \beta_{d1}, \ldots, \beta_{d7}, \beta_{f1}, \ldots, \beta_{f4}, \beta_{lm1}, \ldots, \beta_{lm6}\} \tag{21}$$

which includes the bounds of the productivity shock distribution, the inverse spread parameter, as well as the coefficients on the distance, Heckscher-Ohlin and institutional variables. I anchor the value of two parameters via calibration, namely the elasticities of substitution among goods, ε and ϕ. In particular, I take $\varepsilon = 3.8$ from Bernard et al. (2003), who estimate this from US firm-level data. I set $\phi = 2$ in order to satisfy the condition $\varepsilon > \phi > 1$.[39]

I use the Nelder-Mead (1965) simplex search algorithm to determine the parameter vector, $\hat{\Theta}$, that minimizes the distance metric between selected moments, $b(\cdot)$, of the simulated trade flows, $(X_{ni}^k)^{sim}$,

[38] The solution in standard discrete choice optimization problems is to "concentrate" out the fixed effects via a contraction mapping (Berry 1994, Berry, Levinsohn and Pakes 1995). It is however harder to apply this approach in our context since the product market shares are non-linear functions of the fixed effects of interest, and there are moreover no closed-forms for these shares.

[39] It is in principle possible to estimate ε and ϕ as part of the SMM procedure in future implementations.

and that of the actual data, X^k_{ni}:

$$\min_{\hat{\Theta}} \ (b(\hat{\Theta}) - b(\Theta))' \ \Psi \ (b(\hat{\Theta}) - b(\Theta)) \tag{22}$$

where Ψ is the optimal weight matrix. I employ a standard two-stage procedure (Hansen 1982): Using the identity matrix as the weight matrix, I perform the minimization to obtain a preliminary estimator, $\hat{\Theta}^1$.[40] The inverse squared residual matrix from this first step is a consistent estimator for the optimal weight matrix; I then use this matrix to re-run the minimization to obtain a consistent and efficient estimate of the parameter vector, $\hat{\Theta}^2$.[41] I set $J = 100$, a moderately large value; experimenting with larger values of J raises the computational burden, without changing the value of the objective function significantly.

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

1. The covariances between the trade flows and each of the distance variables and interaction terms on the right-hand side of equation (17). This gives 17 moments: $Cov(X^k_{ni}, Log(Distance)), \ldots,$ $Cov(X^k_{ni}, SVOL \times FLEX)$, which are particularly informative for estimating the $\beta_{d1}, \ldots, \beta_{d7}$, $\beta_{f1}, \ldots, \beta_{f4}, \beta_{lm1}, \ldots, \beta_{lm6}$ coefficients.

2. The mean value of trade flows by industry (20 moments). These serve to fit the remaining structural parameters of the model to match the average levels of observed trade flows.

4.3 The SMM estimates

The coefficient estimates obtained from this moment-matching procedure are reported in Table 3, Column (4). I find once again that physical distance exerts a negative and highly significant effect on trade, although the magnitude of this coefficient is smaller than found with OLS. This is consistent with the hypothesis that the elasticity of trade volumes with respect to distance is decreasing in distance; the exclusion of the zero observations thus tends to bias the magnitude of the OLS distance coefficient upwards, since the zeros correspond to high-distance country pairs where the associated distance elasticity is low (Anderson and van Wincoop 2004, p. 730).[42] I also obtain larger GATT and RTA effects than in the OLS baseline (both coefficients now significant at the 5% level), consistent with the

[40] I initialize this search algorithm with the OLS coefficient values. While the simulation procedure generates predicted values for each country's absorption of own production, X^k_{nn}, these are excluded when computing the moments, since I do not have actual data on domestic absorption.

[41] More precisely, when the vector of moments is given by: $b(\Theta) = \frac{1}{N(N-1)K} \sum_i \psi(x_i, \Theta)$ where x_i denotes the i-th trade flow observation, then the optimal weight matrix used is the inverse of:

$$\frac{1}{N(N-1)K} \sum_i (\psi(x_i, \hat{\Theta}^1) - \psi(x_i, \Theta))(\psi(x_i, \hat{\Theta}^1) - \psi(x_i, \Theta))'$$

[42] Santos-Silva and Tenreyro (2006) and Helpman et al. (2007) also obtained distance coefficients of a smaller magnitude compared to OLS when applying alternative correction procedures to address the bias from omitting the zero observations.

observation that country pairs that are not joint signatories of GATT or a common RTA are less likely to report positive trade flows.[43] The GATT and RTA coefficients were thus biased downward under OLS when the zero observations were dropped. While the common language, colony and common border coefficients remain positive, these are no longer statistically significant. Note that the common colonizer coefficient has switched signs, inheriting the negative sign of the covariance between trade flows and the common colonizer dummy in the actual data.

Turning to the factor endowment interactions, the SMM procedure results in a slightly smaller value for the effect of human capital in promoting skill-intensive exports ($\beta_{f1} = 10.13$), although this remains significant at the 1% level. The physical capital interaction is now estimated precisely ($\beta_{f2} = 0.18$, significant at the 5% level). On the other hand, the materials intensity coefficients are no longer statistically significant, which could reflect the fact that the forest and arable variables may be particularly crude proxies for materials inputs in manufacturing.

The results also confirm the important role played by the legal environment as a determinant of trade patterns. The estimation yields positive and significant effects on two of the three interaction terms involving $LEGAL$, implying that countries with strong institutional rule of law do indeed specialize in industries with a high share of relationship-specific inputs, as well as where job tasks are more complex. I also find evidence of a significant role for country financial development as a source of comparative advantage. The point estimate on the interaction between sales volatility and labor market flexibility is now negative, although this is estimated imprecisely. Column (5) of Table 3 suggests that this could be due in part to the crude nature of the proxies being used for the fixed effects in the SMM estimation: When I run OLS using the proxy, $(\ln(Y_i/L_i))^{\gamma_1}$, as a linear regressor in place of the exporter fixed effects, I obtain a negative coefficient on the $SVOL \times FLEX$ interaction. It is also useful to consider the SMM estimates obtained when matching moments calculated only from the subset of non-zero trade flows in the actual data (Column (6)). These estimates turn out to be closer to the SMM estimates in Column (4) that match moments for the full sample than they are to the OLS estimates in Column (1) that also drops the zeros, suggesting that the OLS specification implies a set of trade volumes that are inadequate for fitting the means and covariances in the observed data.

[43] Indeed, 38% of the bilateral trade observations for joint GATT signatories report positive trade flows, as compared to just 19% when the GATT dummy is equal to zero. Similarly, 80.3% of the observations for common RTA members are positive, compared to just 31.5% when the RTA dummy is equal to 0.

The additional parameters estimated from the SMM procedure are as follows:[44]

$$\{\underline{x},\ \bar{x},\ \theta,\ \eta,\ c_1,\ \gamma_1,\ c_2,\ \gamma_2\} = \{-1.61,\ 7.07,\ 12.41,\ 0.24,\ -0.0045,\ 2.34,\ -0.06,\ 2.52\}$$

In particular, the estimate of $\theta = 12.41$ (standard error $= 0.25$) lies towards the high end of the range reported by EK, being closer to the values that they obtain from instrumental variables. It seems reasonable that this point estimate for the inverse spread parameter should imply a relatively low variance for the productivity distribution within each industry, since the many industry characteristics used in the estimation already help to explain a lot of the cross-industry variation in trade flows. The support of the stochastic productivity distribution, $[\underline{x}, \bar{x}]$, covers about 99.2% of the mass of the regular Gumbel distribution.[45] I obtain an estimated share of tradables in consumption of 0.24. Finally, the point estimates of c_1 and c_2 confirm that the export price level should on average be decreasing in exporter income per capita and in total observed trade in that industry.

4.4 Assessing the goodness of fit

How sensible are these estimates in explaining the real world data? I offer evidence here that the model delivers a reasonable fit on several dimensions, including the country income levels and accompanying pattern of trade flows that it predicts for the global trade equilibrium.

Implied country GDP levels

I first show how to close the model in order to solve for the equilibrium country income levels. Intuitively, factors used more intensively in each country's comparative advantage sectors will receive higher factor prices than under autarky.[46] Total income, being the sum of all factor payments, will thus be determined by the cross-country pattern of specialization.

To solve for the implied country income levels, I appeal to a trade balance condition to close the model. Using the discrete approximation for the measure of varieties, total manufacturing exports

[44] The associated standard errors are: $\underline{x}$ (0.80)**, $\bar{x}$ (3.14)**, θ (0.25)***, c_1 (0.008), c_2 (0.01)***. (*** and ** denote significance at the 1% and 5% levels respectively.) The values of η, γ_1, and γ_2 were restricted to be positive by searching over the natural log of these variables in the minimization procedure, so I do not report standard errors for these parameters. Nevertheless, one can compute the following 95% confidence intervals: η (0.21, 0.27), γ_1 (2.12, 2.56), γ_2 (2.42, 2.62).

[45] The variance-covariance matrix implies that $(\bar{x} - \underline{x}) \sim N(8.68, 2.78^2)$, so that the probability that this interval is well-defined, namely that $\bar{x} > \underline{x}$, is virtually equal to 1.

[46] In the case of the pure two-country Heckscher-Ohlin model, the pattern of specialization would favor each country's relatively abundant factors, and raise those factor prices above their autarky levels.

from country i, EXP_i, are given by:

$$EXP_i \approx \frac{1}{J}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{\{j:\ i(j)=i\}}(p^k_{n,i(j)})^{(j)}(Q^k_{n,i(j)})^{(j)} \tag{23}$$

$$= \frac{1}{J}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{\{j:\ i(j)=i\}}\frac{\eta Y_n\ (P^k_n)^{\varepsilon-\phi}}{\sum_{k\geq 1}(P^k_n)^{1-\phi}}\ ((p^k_{n,i(j)})^{(j)})^{1-\varepsilon} \tag{24}$$

where the sum in (23) is taken over all varieties (across all industries) and over all export destinations for which country i is the lowest-price exporter.

Substituting in the formula for quantity demanded from (3) yields (24), which expresses total exports from country i as a linear combination of the income levels, Y_n, of the N countries in the world. (This property follows from the Cobb-Douglas specification for utility, since the trade quantities, Q^k_{ni}, are then linear functions of the importing country's GDP, Y_n.) I compute the coefficients of Y_n on the right-hand side of (24), namely the $\frac{\eta\ (P^k_n)^{\varepsilon-\phi}}{\sum_{k\geq 1}(P^k_n)^{1-\phi}}\ ((p^k_{n,i(j)})^{(j)})^{1-\varepsilon}$ terms, at the value of the SMM estimate $\hat{\Theta}$ (from Table 3, Column (4)) via simulation; while doing so, I collect and sum up the coefficients that correspond to each Y_n, for $n = 1, \ldots, N$. On the other hand, i's total imports, IMP_i, are equal to ηY_i, so that setting $EXP_i = IMP_i$ for each country gives a homogenous system of N linear equations in the N income levels Y_n. Inverting this system yields the implied country GDP levels, where I set the income level of the US to be 1. Note that the 82 countries in the sample had a combined output equal to 92.9% of total world nominal GDP in 1990, so that the Y_n's that I compute in this way should be a reasonable approximation for country GDP.

Figure 1 confirms that the implied values for country GDP based on the SMM estimate $\hat{\Theta}$ successfully capture the relative rank ordering of observed nominal income levels in 1990.[47] There is a tendency for the model to slightly over-predict GDP, as can be seen from the fact that most of the predicted GDP levels lie above the 45-degree line when plotted against actual GDP (taken from the WDI). Nevertheless, the Spearman rank correlation between the two variables is very high (0.69, significant at the 1% level), so that the model replicates the rank order of country income levels.[48] Note that this positive correlation is not driven purely by the use of income per capita as a fixed effect proxy in the SMM estimation: If I compute the implied income levels omitting the exporter fixed effect term when simulating the trade flow coefficients, the rank correlation with actual country GDPs falls only slightly to 0.68.[49]

Bilateral trade patterns

[47] I use a value of $J = 100$ throughout the goodness of fit and welfare counterfactual exercises.

[48] The Pearson linear correlation between the two log income series is also high (0.71, significant at the 1% level).

[49] The corresponding Pearson correlation using the implied income levels calculated by omitting the fixed effect proxy is 0.68, significant at the 1% level.

How well does the model describe the pattern and volume of bilateral trade flows? Figure 2 performs this cross-check by simulating a full set of trade volumes based on $\hat{\Theta}$ and the implied Y_n's calculated above, and plotting these against the original data.[50] Overall, the model provides a reasonable fit to the data, despite the fair amount of dispersion in both actual and predicted trade flows. Given that the model slightly over-predicts country income levels, most of the data points in Figure 2 lie above the 45-degree line, indicating that the model also tends to generate trade volumes that are larger than that observed, especially for the small trade flows in the actual data. Nevertheless, a regression of log predicted flows against log observed trade yields a positive slope coefficient of 0.77 (significant at the 1% level).[51] Note that the model also does a reasonable job in matching the zero trade flows (which are not shown in Figure 2 because of the log scale). There are 81,990 zeros in the simulated data, of which 64,506 are also common to the 89,806 zeros in the original data.

Price levels

As part of the simulation, the model also generates price levels that provide an additional dimension along which to assess goodness of fit. Figure 3 plots the ideal price indices, $\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}$, computed from (19) using the estimate $\hat{\Theta}$, against an aggregate PPP price index for consumption goods from the Penn World Tables (PWT). This is arguably not a perfect empirical counterpart to the price indices that I simulate: The ideal price index is strictly an index for manufacturing only, but such sectoral price indices are in general not available for the broad country sample in this paper. That said, Figure 3 suggests that the model does get the broad correlation with the actual data correct. The Spearman rank correlation between the generated ideal price indices and the PWT price index is a high 0.60, significant at the 1% level.[52] This is in spite of one clear outlier, Iran, for which exports of primary products (oil) likely matter more than that of manufactures.

5 Welfare Counterfactuals

What do these estimates imply from a welfare perspective about the relative importance of distance barriers and the various country characteristics that determine the pattern of trade? In this section, I calibrate the model with the SMM estimates, $\hat{\Theta}$, and then explore the welfare impact of introducing policy shocks. For example, the framework allows us to explore the effects of reducing distance barriers, to move closer towards a hypothetical zero-gravity world, an exercise which EK and Ramondo (2006)

[50]For comparability, I have scaled up the predicted Y_n's so that the the value of Y_n for the US is equal to that in the GDP variable from the WDI.

[51]I have also examined plots of predicted versus actual trade for each industry, and these tend to be very similar in nature to Figure 2 which pools all the observations.

[52]The corresponding Pearson linear correlation is also high, equal to 0.68 (significant at the 1% level).

also undertake. Additionally, by tying comparative advantage to country and industry characteristics, I can perform counterfactual experiments to examine how a country's pattern of specialization shifts with factor accumulation or an improvement in institutions.

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

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

Without the term, $(1-\eta)^{1-\eta}\eta^{\eta}$, this is precisely equal to country n's real GDP. Notice that I have introduced the price of the domestic non-tradable, p_n^0, explicitly in the denominator: When solving for the implied income levels, Y_n, from the system of trade balance equations, one can only do so relative to a base country (in our case, the US), whose income level I normalize 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.

Welfare changes 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 the ideal price index for differentiated products:

$$\frac{\Delta W_n}{W_n} = \frac{\Delta Y_n}{Y_n} \ - \ (1-\eta)\frac{\Delta(p_n^0)}{p_n^0} \ - \ \eta\frac{\Delta\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}}{\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}} \tag{26}$$

For each counterfactual, I evaluate (26) by simulating a full set of country trade flows both before and after introducing the shock, in order 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 percentage changes in domestic factor prices, 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 the percentage change in w_{nf} from the change in total factor payments accruing to factor f in country n net of any change in the endowment of that factor.[53] To fully operationalize this approach, I set the factor shares in the outside sector as follows: $s_h^0 = 0.07$, $s_l^0 = 0.13$, $s_m^0 = 0.6$, $s_k^0 = 0.2$. These are based on the average factor payment shares over the 1980s in US agriculture (the canonical non-manufacturing sector), based on Mundlak (2005).[54]

[53] More explicitly, I compute the percentage change in factor prices from:

$$\frac{\Delta(w_{nf})}{w_{nf}} = \frac{\Delta(s_f^0(1-\eta)Y_n \ + \ \sum_{k\geq 1} s_f^k \sum_{j=1}^{J}\sum_{s=1}^{N}(p_{sn}^k)^{(j)}(Q_{sn}^k)^{(j)})}{(s_f^0(1-\eta)Y_n \ + \ \sum_{k\geq 1} s_f^k \sum_{j=1}^{J}\sum_{s=1}^{N}(p_{sn}^k)^{(j)}(Q_{sn}^k)^{(j)})} - \frac{\Delta(V_{nf})}{V_{nf}}$$

where the expression for total payments to factor f is evaluated via simulation, using the prices, quantities and implied income levels from before and after the policy shock to calculate the percentage change.

[54] I set $s_m = 0.6$ as the total factor payment share to materials and land, and $s_k = 0.2$ as the factor payments to

Implicit in this counterfactual methodology, I am assuming that factors of production are fully mobile domestically, but that factor markets are segmented across countries. Factors can therefore shift into industries that are favored by the policy shock, with domestic factor prices adjusting accordingly. Note also that I focus on the impact on a representative consumer, putting aside distributional consequences within countries. The counterfactual outcomes that I report are based on the SMM point estimates from Table 3, Column (4), to illustrate the ballpark magnitudes of expected welfare changes.[55]

5.1 Distance counterfactuals

I first consider hypothetical reductions in distance barriers. Although physical distance and transportation costs can never be entirely eliminated in practice, it is nevertheless useful to examine the impact of moving towards a zero-gravity world to gauge how much distance and geography matter for country welfare.

The first panel in Table 4 considers what happens when physical distance is removed, so that it no longer enters the distance mark-up. (Equivalently, this is a counterfactual world in which $\beta_{d1} = 0$.) I find large overall gains in this scenario, with an average welfare increase of 25.6% (in population-weighted terms) in the 82-country sample.[56] This is comparable to the range of country welfare increases that EK reported from removing the distance mark-up in their OECD sample (16.1%-24.1%).[57] Table 4 also reports the decomposition of welfare changes into the three components from (26) – the changes in implied country income levels, in the price of tradables, and in the price of non-tradables respectively. The welfare gain that can be attributed to the fall in tradables prices is non-trivial (5.1%), but the bulk of the welfare increase is clearly being driven by the increase in country income levels (averaging 87.3%) as the removal of distance barriers opens opportunities for access to new markets. This is partially offset by the rise in the price of domestic non-tradables: The increase in foreign demand for each countries' products raises demand for factors of production domestically. Since these are in inelastic supply, factor prices rise as a result, raising the price of domestic non-tradables.

The large average welfare gains reported here mask a lot of heterogeneity in individual country

physical capital, from Mundlak (2005). The residual share for labor is split according to the average shares observed in the NBER-CES dataset for s_h and s_l. Note that the choice of s_h makes relatively little difference to the welfare counterfactuals, since payments to skilled labor are a small share of total factor payments.

[55] It is in principle possible to compute confidence intervals for the welfare changes by taking Monte Carlo draws from the joint distribution of the estimates, but this is computationally very burdensome.

[56] All country welfare averages reported are weighted by population (taken from the WDI). The results are very similar if weighted instead by the size of the labor force from Hall and Jones (1999).

[57] This figure is also similar to the 31% increase that Ramondo (2006) computes from moving to a world with no distance barriers to FDI.

responses to the removal of distance barriers. (The standard deviation of the welfare change is 17.2%, almost as large as the mean.) A small handful of countries in fact suffer a welfare loss, as the removal of distance barriers leads to the diversion of export opportunities to other competitors, but precisely which countries are the main beneficiaries? Figure 4 suggests one natural answer: Countries that are initially more isolated from developed country markets tend to gain more from the removal of these barriers, as illustrated by the positive correlation between the welfare change and each country's average GDP-weighted distance from the rest of the world. (The slope of the regression line is significant at the 1% level.) Along these lines, the average welfare increase in the OECD is 7.0%, whereas the rest of the world (ROW) reports a much-larger gain of 29.8% (Table 4). In short, the reduction of distance disproportionately favors less-developed countries who previously lacked access to global markets, consistent with Lai and Zhu (2004) who obtained a similar result from a model of trade under monopolistic competition.

We can further ask how country patterns of industrial specialization are affected by this transition to a world without physical distance barriers. Measuring industry concentration by the sum of squared industry shares in total country output (namely, as $\sum_{k=1}^{K}\left(\frac{\sum_n X_{ni}^k}{\sum_n \sum_k X_{ni}^k}\right)^2$ for each country i), Figure 5 suggests a trend towards increased specialization with most countries experiencing an increase in this Herfindahl index of industrial composition. The cross-country mean of industry concentration increases from 0.22 to 0.27, and a simple one-sided t-test rejects the null hypothesis that industry concentration remains unchanged following the removal of distance barriers (p-value < 0.001). Intuitively, the increase in trading opportunities prompts countries to focus in on a core set of comparative advantage industries for which their international market has now expanded. The removal of physical distance also triggers a broad diversification in each country's set of export destinations: The sum of squared export destination shares ($\sum_{n=1}^{N}\left(\frac{\sum_k X_{ni}^k}{\sum_n \sum_k X_{ni}^k}\right)^2$, for each country i) drops substantially from a cross-country average of 0.15 to 0.04 (a one-sided t-test rejects the null hypothesis of no change in this export destination Herfindahl at the 1% level). Countries that initially had a more narrow set of export partners experience larger declines in this export destination concentration index (Figure 6), confirming a cross-country convergence towards a diversification of export destination shares. The removal of distance barriers also has a pro-competitive effect that diminishes the market power of large producers in any given industry: A simple concentration index of output by country source ($\sum_{i=1}^{N}\left(\frac{\sum_n X_{ni}^k}{\sum_n \sum_i X_{ni}^k}\right)^2$, for each industry k) falls from an average (across the 20 industries) of 0.55 to a value of 0.43 (a one-sided t-test rejects the null hypothesis that this industry Herfindahl remains unchanged, significant at the 1% level).

An alternative illustration of the effect of distance barriers, which is arguably less extreme than the full removal of physical distance, is to consider what happens when the distance mark-up is exogenously halved. These results are reported in the second panel of Table 4. I find a slightly more modest increase in world welfare (an average of 8.5%), with a smaller variance in these welfare changes as well. The worst-hit country incurs a welfare loss of only -0.9%, while the largest welfare gain is 37.2%. Once again, welfare gains tend to be skewed in favor of non-OECD countries.

On a separate note, the SMM estimates also imply significant gains from multilateral integration through the GATT. I find that transitioning to a world where all countries are GATT signatories results in an average welfare gain of 6.9%. This percentage welfare change has a large negative correlation (-0.89) with the initial value of the GATT dummy, so the countries that gain the most from global integration are those that were initially not GATT members. Since most OECD countries are already GATT members, the average welfare change for this subset of countries is small (-0.1%); most of the gains accrue to non-OECD countries, who see an average welfare increase of 8.5%. Nevertheless, since the world economy as a whole experiences a net gain, it is in principle possible to devise a set of lump-sum transfers within the multilateral GATT framework that would compensate those countries that suffer a welfare loss, in order to generate a Pareto improvement relative to the initial status quo.

5.2 Country policy experiments

The approach taken in this paper of expressing productivity as a function of country and industry characteristics allows us to examine how much each country attribute and each channel of comparative advantage matters for welfare. I offer two sets of illustrative exercises along these lines. First, I consider the impact of simultaneously bringing all countries to the world frontier for each country characteristic in turn (raising this to the maximum level in the sample), to examine how a broad increase in factor endowments or an improvement in institutions affects average welfare levels. Second, I evaluate the effects of raising the country characteristics of one specific developing country.

It is worth stressing some caveats about the precise interpretation of the welfare counterfactuals that I compute. When raising a country characteristic exogenously, I do so by shocking the relevant interaction term involving that characteristic, while holding the exporter fixed effect constant. The welfare changes calculated are therefore strictly due only to the shift in the cross-country pattern of industrial specialization. Since this holds constant any direct effects from the expansion in countries' production capacities, it likely understates the magnitude of welfare changes. These exercises are moreover limited to evaluating the relative efficacy of the different policy levers in generating real income gains; this paper has less to say about the cost side of implementing these policy moves, such as the cost of foregone current consumption from physical capital accumulation or the structural adjustment costs incurred as factors are reallocated across industries.

Table 5 reports the welfare effects of moving all countries to the world frontier on each of the country characteristics that matter for comparative advantage (reported in the 'max' rows). Focusing first on the counterfactuals involving factors of production, I find that average country welfare is raised by a substantial 42.0% when physical capital stocks are exogenously raised.[58] In comparison, the gains from human capital accumulation that accrue through increased specialization in skill-intensive industries (via the interaction with s_h) are smaller (a world average of 18.5%); nevertheless, the combined effect when factoring in the increased specialization in more complex industries is a relatively high 37.6% (the "Total Effect" panel). The decomposition of these changes once again shows that most of these gains are driven by the increase in country GDP in the new trade equilibrium.

Turning to the institutional determinants of trade flows, I obtain large average gains from raising $LEGAL$ (17.7%), with most of this being attributable to the impact that an improvement in legal institutions would have in promoting specialization in industries that have a large share of relationship-specific inputs. On the other hand, the overall impact of raising financial development is rather muted, with an average welfare gain of only 0.4%. It turns out that this policy shock creates huge gains for those countries that lagged far behind in access to private credit (the maximum welfare increase is 23.1%), while severely eroding the position of countries with the best financial systems (the largest welfare loss is −25.7%), largely because the world distribution of levels of financial development is very right-skewed, so that the most financially-developed countries initially enjoyed a large advantage along this dimension over the rest of the world.

The last column of Table 5 shows that there is in general a strong negative correlation between the percentage welfare change experienced by a country and its initial rank in the sample for the country characteristic in question. This tight correlation is even clearer when graphed: The lower a country's initial relative endowment of skilled labor, the larger its subsequent gains from human capital accumulation (Figure 7). There is an even stronger negative correlation between countries' initial physical capital to labor ratio, and the gains that accrue from an exogenous increase in capital stocks (Figure 8). Finally, the poorer a country's $LEGAL$ score to begin with, the more it gains from improving these institutions to first-world standards (Figure 9).

I also examine in Table 5 the welfare consequences from the more moderate policy exercise of increasing country attributes by one standard deviation, instead of raising them to the world frontier level (results reported in the '(1+s.d.)' rows). Notice first that applying such a uniform increase to the country's institutional attributes ($FINDEV$ or $LEGAL$) has a minimal impact on welfare levels,

[58] In his counterfactual exercises, Shikher (2005) finds that equalizing the capital-labor ratio across countries has little effect on the total volume of trade. This is not necessarily inconsistent with the large welfare effects found here, since Shikher's sample consists only of OECD countries for which relative factor endowments are more similar to begin with. Reassuringly, Shikher finds that physical capital accumulation does tend to boost specialization in capital-intensive industries, consistent with what I find in a similar policy experiment with Indonesia later in this section.

with the resulting average welfare change ranging from -1.7% to 0.4%. The one standard deviation increase does not alter the relative ranking of countries along each of these institutional dimensions, and so has a minimal impact on the cross-country pattern of specialization, hence the negligible impact on country real incomes. This highlights that the large average gains previously calculated clearly stem from the dramatic shift in patterns of comparative advantage induced by raising country attributes to the maximum level. In contrast, I continue to find positive average welfare gains from raising country factor endowments by one standard deviation; in the case of the physical capital counterfactual, these are in fact fairly large (38.1%). However, a closer look at the decomposition of these changes reveals that they are due almost entirely to a fall in the price of non-tradables. These welfare gains are thus driven by an endowment effect, which reduces factor prices and hence makes domestic non-tradables cheaper to consumers; there is once again little corresponding change in country income levels, as the pattern of comparative advantage remains relatively unaltered.

I next consider raising conditions in just one country, arguably a more realistic policy exercise. Conveniently, there is one large developing country which lies between the 25th and 33rd percentiles in the sample for each of the country characteristics, Indonesia, so I examine the welfare changes from raising Indonesia to the level of the world leader along each of these dimensions (Table 6). Once again, I find the largest gains from factor accumulation, with a 43.8% welfare increase in the case of physical capital accumulation to the world frontier, and a 38.7% increase when skill endowments are similarly raised. These are about twice as large as the welfare increase should the quality of the legal system be improved (20.2%). Naturally, the magnitude of the welfare gains from raising these country attributes by one standard deviation only are smaller, though qualitatively similar. For the rest of the world, while there are some countries that do see welfare declines, the magnitudes of these losses are extremely small. Thus, any adverse beggar-thy-neighbor effects from policy shocks in Indonesia – from the diversion of export opportunities away from other countries – appear to be small.

Focusing on a policy shock to one country alone also allows us to examine in finer detail what happens to industry composition in that country. Table 7 illustrates the change to the pattern of Indonesia's exports in response to separate shocks that raise Indonesia's human capital endowment and the quality of its legal institutions to the world frontier level. In both of these scenarios, Indonesia's share of the world market increases substantially, by over 100% in some industries, due to the expansion in the country's production capabilities. At the same time, there is a fair amount of reallocation taking place between industries within Indonesia, as illustrated by the change in each industry's share of Indonesia's total exports. Furthermore, in the human capital counterfactual, I find a strong positive correlation between the change in an industry's relative size and the skill intensity of that industry: An increase in human capital endowment favors industries that use that factor of production more intensively (Figure 10). For the policy experiment involving $LEGAL$, this correlation between the change in industry share and the respective industry characteristics that capture institutional dependence (HI, RS, and $COMPL$) is weaker, because the presence of these separate mechanisms involving $LEGAL$ makes it harder to identify a clean relationship with each of the industry characteristics with a small set of just 20 industries.

In sum, the country policy counterfactuals imply average welfare gains of a reasonable magnitude, with the largest potential gains stemming from physical capital accumulation. Raising skill endowments and institutional improvements to the contracting environment also deliver significant, albeit slightly smaller, welfare gains. This holds true whether I raise all countries' characteristics to the world frontier at the same time, or just focus on improving conditions in one specific developing country.

6 Conclusion

This paper has developed a methodology for quantifying the importance of different sources of comparative advantage that jointly determine the pattern of trade, in a manner that allows the researcher to evaluate interesting welfare counterfactuals. To understand patterns of industrial specialization, I presented an extension of the Ricardian 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 provide for with their endowment mix or institutional strengths. This framework turns out to be very flexible, allowing me to incorporate the full set of country-industry interaction terms identified in the recent literature as significant sources of comparative advantage.

I estimated the underlying parameters of the model using a large bilateral trade flow dataset, that also brings together many country and industry characteristics capturing both Heckscher-Ohlin and institutional forces. I pursued two estimation approaches: (i) an OLS baseline, presented in Section 3, and (ii) a simulated method of moments (SMM) procedure that takes into account the zero trade observations, in Section 4. Both sets of estimates confirm the relevance of traditional gravity measures, particularly physical distance, for explaining the pattern of trade. I also find corroboration for the role of factor endowments and country institutions – including financial development, the contracting environment, and labor market regimes – as sources of comparative advantage, even with the relatively aggregate industry classification

The SMM estimates in turn implied large average welfare gains from reducing distance barriers and from multilateral integration, with developing countries benefiting more than the OECD as they gain better access to developed country markets. I also quantified the welfare changes from the underlying shift in patterns of comparative advantage when country attributes are raised, finding large average gains from increasing physical capital and human capital endowments, as well as from improving legal institutions to first-world standards.

The approach in this paper is certainly not without some limitations. The specification for how country and industry interactions enter into the systematic component of productivity can be generalized to make this functional form more flexible (for example, to consider non-linear effects), but I have opted not to do so in this implementation because of the computational cost of increasing the number of parameters. Also, the counterfactual exercises treat the policy changes as exogenous shocks, in order to gauge the magnitude of consequent welfare gains in the hypothetical new equilibrium. This puts aside dynamic issues, such as the time frame required for the world economy to adjust to the new trade equilibrium, as well as potential endogenous policy responses on the part of other countries. Nevertheless, the contribution of this paper lies in the steps taken towards establishing a quantitative methodology for tying specialization patterns to country and industry characteristics, and towards more extensive applications of structural estimation methods to analyze the determinants of trade flows. My priorities for future work include exploring the use of additional data moments to obtain more efficient parameter estimates, as well as alternative ways to account for the direct country and industry fixed effects which were proxied for in the SMM procedure.

7 References

Acemoglu, Daron, Pol Antràs, and Elhanan Helpman, (2007), "Contracts and Technology Adoption," forthcoming in the *American Economic Review*.

Alvarez, Fernando, and Robert E. Lucas, Jr., (2005), "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade," *NBER Working Paper* 11764.

Anderson, James, and Douglas Marcouiller, (2002), "Insecurity and the Pattern of Trade: An Empirical Investigation," *Review of Economics and Statistics* 84: 342-352.

Anderson, James, and Eric van Wincoop, (2003), "Gravity with Gravitas: A Solution to the Border Problem," *American Economic Review* 93: 170-192.

Anderson, James, and Eric van Wincoop, (2004), "Trade Costs," *Journal of Economic Literature* 42: 691-751.

Antràs, Pol, (2003), "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics* 118: 1375-1418.

Baldwin, Robert, (1971), "Determinants of the Commodity Structure of US Trade," *American Economic Review* 61: 126-146.

Baldwin, Robert, (1979), "Determinants of Trade and Foreign Investment: Further Evidence," *Review of Economics and Statistics* 61: 40-48.

Bartelsman, Eric J., Randy A. Becker, and Wayne B. Gray, (2000), NBER-CES Manufacturing Industry Database, http://www.nber.org/nberces/

Barro, Robert, and Lee Jong-Wha, (2000), "International Data on Educational Attainment, Updates and Implications," *NBER Working Paper* No. 7911.

Beck, Thorsten, Aslý Demirgüç-Kunt, and Ross Levine, (2000), "A New Database on Financial Development and Structure," *World Bank Economic Review* 14: 597-605.

Beck, Thorsten, (2002), "Financial Development and International Trade: Is There a Link?" *Journal of International Economics* 57: 107-131.

Beck, Thorsten, (2003), "Financial Dependence and International Trade," *Review of International Economics* 11: 296-316.

Becker, Bo, and David Greenberg, (2005), "Financial Development and International Trade," University of Illinois at Urbana-Champaign manuscript.

Bernard, Andrew, Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, (2003), "Plants and Productivity in International Trade," *American Economic Review* 93: 1268-1290.

Berry Steven, (1994), "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics* 25: 242262.

Berry, Steven, James Levinsohn, and Ariel Pakes, (1995), "Automobile Prices in Market Equilibrium," *Econometrica* 63: 841-890.

Botero, Juan, Simeon Djankov, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, (2004), "The Regulation of Labor," *Quarterly Journal of Economics* 119: 1339-1382.

Costinot, Arnaud, (2006), "On the Origins of Comparative Advantage," University of California at San Diego manuscript.

Costinot, Arnaud, and Ivana Komunjer, (2006), "What Goods do Countries Trade? New Ricardian Predictions," University of California at San Diego manuscript.

Cuñat, Alejandro, and Marc J. Melitz, (2006), "Volatility, Labor Market Flexibility, and Comparative Advantage," Harvard University manuscript.

Davis, Donald, and David Weinstein, (2001), "An Account of Global Factor Trade," *American Economic Review* 91: 1423-1453.

Deardorff, Alan, (1982), "The General Validity of the Heckscher-Ohlin Theorem," *American Economic Review* 72: 683-694.

Debaere, Peter, (2003), "Relative Factor Abundance and Trade," *Journal of Political Economy* 111: 589-610.

Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson, (1977), "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods," *American Economic Review* 67: 823-839.

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

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, (2005), "An Anatomy of International Trade: Evidence from French Firms," New York University mimeo.

Eaton, Jonathan, and Akiko Tamura, (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, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, Hengyong Mo, (2005), "World Trade Flows: 1962-2000," *NBER Working Paper* 11040.

Feenstra, Robert C., John Romalis, and Peter K. Schott, (2002), "U.S. Imports, Exports and Tariff Data," http://www.internationaldata.org/

Galombos, Janos, (1978), *The Asymptotic Theory of Extreme Order Statistics*, New York: John Wiley.

Golub, Stephen S., and Chang-Tai Hsieh, (2000), "Classical Ricardian Theory of Comparative Advantage Revisited," *Review of International Economics* 8: 221-234.

Gwartney, James, and Robert Lawson (2004), *Economic Freedom of the World: 2004 Annual Report*, Vancouver: The Fraser Institute. Data from www.freetheworld.com.

Hall, Robert, and Charles Jones, (1999), "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics* 114: 83-116.

Hallak, Juan Carlos, (2006), "Product Quality and the Direction of Trade," *Journal of International Economics* 68: 238-265.

Hansen, Lars Peter, (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50: 1029-1054.

Haveman, Jon, (—), Industry Concordances, http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/tradeconcordances.html.

Haveman, Jon, and David Hummels, (2004), "Alternative Hypotheses and the Volume of Trade: The Gravity Equation and the Extent of Specialization," *Canadian Journal of Economics* 37: 199-218.

Helpman, Elhanan, Marc J. Melitz, and Yona Rubinstein, (2007), "Estimating Trade Flows: Trading Partners and Trading Volumes," NBER Working Paper #12927.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi (2005), "Governance Matters IV: Governance Indicators for 1996-2004," *World Bank Policy Research Department Working Paper* 3630.

Kee, Hiau Looi, Alessandro Nicita, and Marcelo Olarreaga, (2006), "Estimating Trade Restrictiveness Indices," *World Bank Policy Research Working Paper* 3840.

Lai, Huiwen, and Daniel Trefler, (2002), "The Gains from Trade with Monopolistic Competition: Specification, Estimation, and Mis-Specification," *NBER Working Paper* No. 9169.

Lai, Huiwen, and Susan Chun Zhu, (2004), "The Determinants of Bilateral Trade," *Canadian Journal of Economics* 37: 459-483.

Levchenko, Andrei, (2004), "Institutional Quality and International Trade," *IMF Working Paper* 04/231.

Manova, Kalina, (2006), "Credit Constraints, Heterogenous Firms, and International Trade," Harvard University manuscript.

Matsuyama, Kiminori, (2005), "Credit Market Imperfections and Patterns of International Trade and Capital Flows," *Journal of the European Economic Association* 3: 714-723.

Mundlak, Yair, (2005), "Economic Growth: Lessons from Two Centuries of American Agriculture," *Journal of Economic Literature* 43: 989-1024

Nelder John, and Roger Mead, (1965), "A Simplex Method for Function Minimization," *Computer Journal* 7: 308-313.

Nunn, Nathan, (2007), "Relationship-Specificity, Incomplete Contracts and the Pattern of Trade," forthcoming in the *Quarterly Journal of Economics*.

Pakes, Ariel, and David Pollard, (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica* 57: 1027-1057.

Rajan, Raghuram, and Luigi Zingales, (1998), "Financial Dependence and Growth," *American Economic Review* 88: 559-586.

Ramondo, Natalia, (2005), "Size, Geography, and Multinational Production," University of Chicago manuscript.

Rauch, James E. (1999), "Networks versus Markets in International Trade," *Journal of International Economics* 48: 7-35.

Romalis, John, (2004), "Factor Proportions and the Structure of Commodity Trade," *American Economic Review* 94: 67-97.

Rose, Andrew, (2004), "Do We Really Know that the WTO Increases Trade?" *American Economic Review* 94: 98-104.

Santos-Silva, J.M.C., and Silvana Tenreyro, (2006), "The Log of Gravity," *Review of Economics and Statistics* 88: 641-658.

Shikher, Serge, (2004), "Putting Industries into the Eaton-Kortum Model," Suffolk University manuscript.

Shikher, Serge, (2005), "Accounting for International Trade," Suffolk University manuscript.

Trefler, Daniel, (1995), "The Case of the Missing Trade and Other Mysteries," *American Economic Review* 85: 1029-1046.

Wynne, José, (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, in thousands of current US dollars. This was converted from the SITC Rev 2 classification into US 1987 SIC format using the detailed information on US exports in Feenstra, Romalis and Schott (2002), henceforth FRS. FRS record US export data at the highly disaggregate Harmonized System (HS) 10-digit level, where each HS-10 product is also assigned a 5-digit SITC Rev 2 and a 4-digit SIC-87 category. I use the FRS 1990 data to derive concordance weights to map SITC Rev 2 categories into SIC-87 format. This follows the procedure adopted in Cuñat and Melitz (2006).

Two complications arise. First, the SIC-87 categories in FRS are export-based, in that classification is based on observed finished products. However, the distinction between SIC industries is often defined according to the production process. To cite an example used by FRS, SIC 2011 and SIC 2013 are both for processed meats, with the difference being that 2011 conducts its own slaughtering while 2013 uses purchased carcasses. When products are observed at the dock, it is not possible to distinguish between the two, and so trade flows for both are merged under SIC 2011, with 2013 omitted from the dataset. Table 1.3 in FRS lists the affected industries, detailing which categories have been excluded and which codes the export value has been merged under. For our year of interest (1990), I break up these merged trade flows for the affected categories in proportion to the value of US total shipments as reported in the NBER-CES database.[59] Having done this, the SITC codes associated with the included SIC industry are also assigned to the previously excluded SIC industries. A second complication relates to Feenstra et al.'s (2005) use of SITC categories with the letters 'A' and 'X', for trade flows not observed at a more disaggregate level. I apportioned trade in these 'A' and 'X' categories on the basis of the truncated (more aggregate) SITC code. In other words, I treat 111A and 111X as coming from the 3-digit SITC category 111, and use FRS to construct weights to map SITC 111 into SIC categories.

Finally, trade flows were summed up to the 2-digit SIC level, yielding 20 industry groups. A zero is imputed for all exporter-importer-industry cells for which no trade was reported.

Distance: The primary physical distance measure is the great circle formula distance between countries' major population centers, made available by the Centre d'Etudes Prospectives et d'Informations

[59] One exception: SIC 2092 is excluded from FRS, with the associated trade flows being merged under SIC 0912 and 0913, which are primary fishing industries. Since shipment data for the 09XX categories is not available in the NBER-CES database, I imputed all of 0912 and 0913 to 2092.

Internationales (CEPII).[60] I set a country's distance from itself to zero, so that distance does not impose an iceberg mark-up for internal country trade.

Several measures of linguistic and historical proximity were also taken from the CEPII: (i) a "Common Language" dummy equal to 1 if at least 9% of each country's population speaks a shared language; (ii) a "Colony" dummy equal to 1 if one of the countries had ever colonized the other; and (iii) a "Common colonizer" dummy equal to 1 if the two countries were ruled by the same colonial power post-1945. These measures are supplemented by three binary variables coded by Rose (2004), which I augmented with reference to the CIA World Factbook[61] and the WTO website[62] to cover all country pairs in my sample: (i) "Border", equal to 1 if the countries have a common land border; (ii) "RTA", equal to 1 if the countries are signatories in one of the following regional trade agreements: EC, US-Israel, Canada-US (the precursor of NAFTA), CARICOM, PATCRA, ANZ-CERTA, CACM, SPARTECA, EFTA; and (iii) "GATT", equal to 1 if both countries were GATT/WTO members by the end of 1990. I code a value of 1 for all six dummies for a country's distance from itself.

B. INDUSTRY CHARACTERISTICS

Factor intensities: From the NBER-CES Manufacturing Industry Database (Bartelsman et al. 2000). Skill intensity (s_h) is calculated as the ratio of non-production worker payroll to the value of total shipments. Materials intensity (s_m) equals material costs (raw materials, parts and supplies, and energy) divided by total shipments. Payments to capital are a residual (total shipments minus total payroll and materials cost), with physical capital intensity (s_k) being the ratio of these residual payments to total shipments. All factor intensities are calculated for the period 1980-89.

External capital dependence ($CAPDEP$): Constructed following the methodology in Rajan and Zingales (1998), defined for a given firm to be the fraction of total capital expenditures over 1980-89 not financed by internal cash flow. Computed from the Compustat dataset, which contains the universe of publicly-traded firms in the US. The median value across firms in each SIC-87 2-digit category is used as the measure of external capital dependence in that industry. (The measure in Rajan and Zingales (1998) is constructed for a different classification system, namely ISIC 3- and 4-digit industries.)

Input concentration (HI): Constructed following Levchenko (2004). Equal to the Herfindahl index of intermediate input use, as detailed in the 1987 US Input-Output (IO) Use Table at the IO 6-digit

[60] http://www.cepii.fr/anglaisgraph/bdd/distances.htm
[61] http://www.odci.gov/cia/publications/ factbook/index.html
[62] http://www.wto.org/english/thewto_e/ gattmem_e.htm

level. The IO-87 categories map cleanly into the SIC-87 4-digit categories based on the correspondence table provided by the Bureau of Economic Analysis (BEA).[63] When an IO-87 category maps into more than one SIC category, I split the inputs in proportion to US domestic shipments in the SIC destination categories, using the value of total shipments by US firms from the NBER-CES database as weights. Input use is then aggregated to the SIC 2-digit level, from which the input Herfindahl is calculated.

Input Relationship-Specificity (RS): From Nunn (2007), by e-mail communication. RS is equal to the share (by value) of inputs that are not sold on an organized exchange. This corresponds to the measure z^{rs2} in Nunn (2007). Data on input use is from the 1987 US Input-Output Use Table. Rauch (1999) provides the classification of goods into: (i) those sold on an organized exchange; (ii) those reference-priced in commercial publications; and (iii) goods that fall in neither of the above categories.[64] Moving from (i) to (iii), I have successively more differentiated and hence more relationship-specific inputs. Rauch provides two codings, one "conservative" and one "liberal"; I use the "liberal" classification. I map the IO-87 codes to SIC-87 4-digit categories using the correspondence table provided by the BEA. I aggregate the measure up 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 (2006). The 1985 and 1993 instalments of the US Panel Survey of Income Dynamics (PSID) contained a question asking respondents to gauge how many months it would take a typical new employee with the requisite education background to become "fully trained and qualified" in the respondents' job. Costinot (2006) calculates the average response for SIC-1972 3-digit industries, normalized to a maximum value of 1. I assign these values to the corresponding 4-digit sub-categories, and matched these to SIC-1987 codes using the correspondence table developed by Bartelsman, Becker and Gray.[65] For missing 4-digit level observations, I assign the median complexity level observed at successively higher levels of industry aggregation (first at the 3-digit level, and if that is still missing, at the 2-digit level, and then at the 1-digit level). The value of $COMPL$ for each 2-digit industry was then taken to be the median over all its 4-digit sub-categories. There are two industry groups for which this imputation process may seem overly liberal, namely SIC 21 and 29, for which direct source information on complexity was not available in the PSID for any of the corresponding 3-digit industries. The OLS results are similar if I omit these two industry groups from the analysis.

[63] Available at: http://www.bea.gov/bea/pn/ndn0016.zip. All SIC 4-digit industries are associated with a unique IO-87 6-digit category, except for SIC 3999 which is matched with two IO-87 6-digit categories.

[64] The classification is available at: http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/TradeData.html#Rauch

[65] Available at: http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/Concordances/FromusSIC/sic7287.txt

Sales Volatility ($SVOL$): From Cuñat and Melitz (2006), by e-mail communication. Equal to the employment-weighted standard deviation of sales growth for firms in the 1980-2004 Compustat sample. Only firms that provided at least 5 years of data were included in this calculation, while observations where the absolute sales growth rate exceeded 300% were omitted as outliers.

C. COUNTRY CHARACTERISTICS

Factor endowments: Physical capital per worker ($\log(K/L)$) and human capital per worker ($\log(H/L)$) are taken from Hall and Jones (1999). The physical capital data is based on the Penn World Tables (PWT) for 1988, while the human capital variable is constructed using data on education attainment in 1985 from the Barro-Lee (2000) dataset. Specifically, the stock of human capital is constructed as the Mincerian return-weighted average of the stock of workers with given levels of completed education.

Raw materials abundance is proxied by hectares of forest land per worker ($\log(Forest/L)$), and hectares of arable land per worker ($\log(Arable/L)$). Both land measures are from the World Development Indicators (WDI). The measure of forest land is available only for 1990, while that for arable land is an average over 1980-89.

Financial development ($FINDEV$): From Beck et al.'s (2000) Financial Structure and Economic Development Database, March 14 2005 update.[66] Equal to the amount of credit extended by banks and other financial intermediaries to the private sector divided by GDP, averaged over 1980-89.

Legal System ($LEGAL$): From Gwartney and Lawson (2004).[67] Index measure of "Legal System and Property Rights" for 1985, rescaled between 0 and 1. Based on a composite of five sub-indices on: judicial independence; impartiality of courts; protection of intellectual property; military interference in the rule of law and the political process; and integrity of the legal system. These component indices are drawn from the International Country Risk Guide (ICRG) and the Global Competitiveness Report (GCR), the former being a private institutional assessment, while the latter is a survey of business executives based in each country of interest.

Employment Flexibility ($FLEX$): From the World Bank's *Doing Business* database. Index of "Rigidity of Employment", averaged over 2004-05, rescaled to be increasing in labor market flexibility and to lie between 0 and 1. The Doing Business index is itself an average of three sub-indices on: the difficulty of hiring a new worker; restrictions on expanding or contracting the number of working hours; and the difficulty and expense of dismissing a redundant worker. The indices are coded based

[66] http://www.worldbank.org/research/projects/finstructure/FinStructure_Database_60_03.xls

[67] http://www.freetheworld.com/2004/2004dataset.xls

on the methodology in Botero et al. (2004).[68]

GDP: Both GDP and GDP per capita are taken from the WDI, measured in current US dollars.

Population: From the WDI.

Price Indices: From the PWT, measured in purchasing power parity units.

[68] For more details, see: http://www.doingbusiness.org/ExploreTopics/HiringFiringWorkers/

Table 1A

List of SIC-87 2-digit Industries (20)

SIC Major groups: (2-digit level)

20: Food and Kindred Products
21: Tobacco Products
22: Textile Mill Products
23: Apparel and other Finished Products made from Fabrics and similar materials
24: Lumber and Wood Products, except Furniture
25: Furniture and Fixtures
26: Paper and Allied Products
27: Printing, Publishing, and Allied Industries
28: Chemicals and Allied Products
29: Petroleum Refining and Related Industries
30: Rubber and Miscellaneous Plastics Products
31: Leather and Leather Products
32: Stone, Clay, Glass, and Concrete Products
33: Primary Metal Industries
34: Fabricated Metal Products, except Machinery and Transportation Equipment
35: Industrial and Commercial Machinery, and Computer Equipment
36: Electronic and other Electrical Equipment, except Computer Equipment
37: Transportation Equipment
38: Measuring, Analyzing, and Controlling Instruments
(Photographic, Medical and Optical Goods; Watches and Clocks)
39: Miscellaneous Manufacturing Industries

Table 1B
List of Countries in Sample (82)

Countries: (ISO codes in parentheses)

Argentina (ARG), Australia (AUS), Austria (AUT), Burundi (BDI), Belgium (BEL), Bolivia (BOL), Brazil (BRA), Central African Republic (CAF), Canada (CAN), Switzerland (CHE), Chile (CHL), China (CHN), Ivory Coast (CIV), Cameroon (CMR), Colombia (COL), Costa Rica (CRI), Germany (DEU), Denmark (DNK), Dominican Republic (DOM), Algeria (DZA), Ecuador (ECU), Egypt (EGY), Spain (ESP), Finland (FIN), France (FRA), United Kingdom (GBR), Ghana (GHA), Greece (GRC), Guatemala (GTM), Honduras (HND), Haiti (HTI), Hungary (HUN), Indonesia (IDN), India (IND), Ireland (IRL), Iran (IRN), Israel (ISR), Italy (ITA), Jamaica (JAM), Jordan (JOR), Japan (JPN), Kenya (KEN), South Korea (KOR), Sri Lanka (LKA), Morocco (MAR), Madagascar (MDG), Mexico (MEX), Mali (MLI), Malawi (MWI), Malaysia (MYS), Niger (NER), Nigeria (NGA), Nicaragua (NIC), Netherlands (NLD), Norway (NOR), New Zealand (NZL), Pakistan (PAK), Panama (PAN), Peru (PER), Philippines (PHL), Papua New Guinea (PNG), Poland (POL), Portugal (PRT), Paraguay (PRY), Senegal (SEN), Singapore (SGP), Sierra Leone (SLE), El Salvador (SLV), Sweden (SWE), Syria (SYR), Chad (TCD), Togo (TGO), Thailand (THA), Tunisia (TUN), Turkey (TUR), Uganda (UGA), Uruguay (URY), United States (USA), South Africa (ZAF), Zaire (ZAR), Zambia (ZMB), Zimbabwe (ZWE)

Table 2
Empirical Model of Bilateral Industry Trade Flows (OLS)

Dependent variable = $\ln\left(\frac{X_{ni}^k}{X_{nu}^k}\right)$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance and Geography: $D_{ni} - D_{nu}$							
β_{d1}: Log (Distance)	-1.12***	-1.12***	-1.12***	-1.12***	-1.12***	-1.13***	-1.12***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
β_{d2}: Common Language	0.71***	0.68***	0.68***	0.68***	0.67***	0.68***	0.68***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
β_{d3}: Colony	0.42***	0.48***	0.48***	0.49***	0.49***	0.48***	0.49***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
β_{d4}: Common colonizer	0.20	0.22	0.22	0.22	0.23	0.23	0.22
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
β_{d5}: Border	0.04	0.06	0.07	0.07	0.07	0.07	0.07
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
β_{d6}: RTA	0.48***	0.46***	0.46***	0.45***	0.45***	0.46***	0.45***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
β_{d7}: GATT	0.09	0.11	0.12	0.11	0.11	0.14	0.11
	(0.25)	(0.24)	(0.25)	(0.24)	(0.25)	(0.25)	(0.24)
Heckscher-Ohlin determinants: $\left(\ln\frac{V_{if}}{V_{i0}} - \ln\frac{V_{uf}}{V_{u0}}\right) s_f^k$							
β_{f1}: $s_h \times \log(H/L)$		38.04***	32.09***	37.47***	34.87***	24.42***	37.07***
		(1.66)	(1.63)	(1.69)	(1.73)	(1.86)	(1.65)
β_{f2}: $s_k \times \log(K/L)$		1.77***	1.30***	1.77***	1.59***	0.59***	1.88***
		(0.19)	(0.19)	(0.19)	(0.19)	(0.20)	(0.19)
β_{f3}: $s_m \times \log(Forest/L)$		0.13	0.17*	0.11	0.11	0.08	0.12
		(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
β_{f4}: $s_m \times \log(Arable/L)$		1.23***	0.94***	1.25***	1.22***	1.19***	1.25***
		(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
Institutional determinants: $(L_{il} - L_{ul})M_{km}$							
β_{lm1}: $CAPDEP \times FINDEV$			1.78***				
			(0.09)				
β_{lm2}: $HI \times LEGAL$				3.87**			
				(1.94)			
β_{lm3}: $RS \times LEGAL$					2.78***		
					(0.74)		
β_{lm4}: $COMPL \times LEGAL$						7.16***	
						(0.40)	
β_{lm5}: $COMPL \times \log(H/L)$						1.54***	
						(0.32)	
β_{lm6}: $SVOL \times FLEX$							16.05***
							(2.10)
Number of obs.	40501	40501	40501	40501	40501	40501	40501
R^2	0.629	0.644	0.649	0.644	0.644	0.651	0.645

Notes: Robust standard errors, clustered by exporter-importer pair, are reported; ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively. All specifications include exporter and importer-industry fixed effects. All distance and exporter variables are taken relative to the US ('u'), by subtracting the corresponding US value of that variable.

Table 3
Empirical Model of Bilateral Industry Trade Flows (OLS, Probit, SMM)

In Columns (1), (1a), (2), (5), Dependent variable = $\ln\left(\frac{X_{ni}^k}{X_{nu}^k}\right)$

	(1) OLS All	(1a) OLS Betas	(2) OLS EK	(3) Probit All	(4) SMM All	(5) OLS FE proxy	(6) SMM Non-zeros
Distance and Geography:							
β_{d1}: Log (Distance)	−1.13***	−0.35***	−1.25***	−0.64***	−0.36***	−0.76***	−0.36***
	(0.03)	(0.01)	(0.12)	(0.03)	(0.13)	(0.05)	(0.07)
β_{d2}: Common Language	0.66***	0.10***	0.49**	0.50***	0.15	0.60***	0.15
	(0.06)	(0.01)	(0.20)	(0.04)	(0.80)	(0.09)	(0.21)
β_{d3}: Colony	0.49***	0.05***	0.57***	−0.05	0.45	0.80***	0.44
	(0.10)	(0.01)	(0.19)	(0.08)	(8.20)	(0.16)	(0.75)
β_{d4}: Common colonizer	0.24	0.01	—	0.00	−1.44	−0.34*	-1.48
	(0.15)	(0.01)		(0.11)	(9.05)	(0.20)	(70.71)
β_{d5}: Border	0.09	0.01	−0.06	−0.16	0.68	0.77***	0.69*
	(0.15)	(0.01)	(0.25)	(0.15)	(4.08)	(0.18)	(0.36)
β_{d6}: RTA	0.45***	0.04***	0.16	−0.40***	0.78**	1.22***	0.70**
	(0.10)	(0.01)	(0.17)	(0.15)	(0.39)	(0.17)	(0.31)
β_{d7}: GATT	0.14	0.01	—	−0.30*	0.73**	0.17	0.51
	(0.25)	(0.02)		(0.15)	(0.35)	(0.15)	(0.37)
Heckscher-Ohlin determinants:							
β_{f1}: $s_h \times \log(H/L)$	15.37***	0.10***	1.68	3.54***	10.13***	5.42***	10.15***
	(2.00)	(0.01)	(5.20)	(0.98)	(0.60)	(2.03)	(0.59)
β_{f2}: $s_k \times \log(K/L)$	0.28	0.03	2.90**	−0.09	0.18**	−0.03	−0.20
	(0.20)	(0.02)	(1.18)	(0.08)	(0.08)	(0.21)	(0.17)
β_{f3}: $s_m \times \log(Forest/L)$	0.09	0.03	0.30	0.08**	0.04	0.11***	0.05
	(0.10)	(0.03)	(0.21)	(0.04)	(0.68)	(0.04)	(0.14)
β_{f4}: $s_m \times \log(Arable/L)$	1.00***	0.21***	0.99**	0.26***	−0.29	−0.47***	−0.26**
	(0.15)	(0.03)	(0.40)	(0.06)	(0.70)	(0.06)	(0.11)
Institutional determinants:							
β_{lm1}: $CAPDEP \times FINDEV$	1.17***	0.07***	1.21***	0.25***	1.81***	1.90***	2.03**
	(0.09)	(0.01)	(0.17)	(0.06)	(0.59)	(0.10)	(0.81)
β_{lm2}: $HI \times LEGAL$	−0.55	−0.004	−22.36***	8.57***	1.77	−6.36***	1.17
	(1.95)	(0.01)	(7.92)	(0.78)	(3.86)	(1.95)	(1.08)
β_{lm3}: $RS \times LEGAL$	4.70***	0.20***	−14.85***	4.26***	1.43**	2.45***	1.41***
	(0.76)	(0.03)	(3.37)	(0.32)	(0.57)	(0.50)	(0.51)
β_{lm4}: $COMPL \times LEGAL$	4.87***	0.14***	5.59***	0.51**	0.36**	5.44***	0.71
	(0.45)	(0.01)	(1.85)	(0.22)	(0.16)	(0.49)	(0.57)
β_{lm5}: $COMPL \times \log(H/L)$	2.16***	0.10***	6.68***	0.36**	1.81*	1.07***	1.96***
	(0.33)	(0.01)	(0.96)	(0.15)	(1.00)	(0.34)	(0.36)
β_{lm6}: $SVOL \times FLEX$	10.57***	0.10***	6.52**	1.26	−2.42	−7.06***	−2.35
	(2.07)	(0.02)	(3.18)	(1.18)	(2.23)	(1.11)	(2.60)
Number of obs.	40501	40501	6030	116080	–	40501	–
R^2 or Pseudo-R^2	0.653	0.653	0.747	0.612	–	0.425	–

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively. For Columns (1), (1a), (2), and (3), exporter and importer-industry fixed effects are included; robust standard errors, clustered by exporter-importer pair, are reported. Column (1a) contains standardized beta coefficients from the Column (1) specification. Column (2) is restricted to the 19 OECD countries in EK. Column (3) performs a probit regression on the probability of observing positive normalized trade flows. Column (5) performs OLS using the exporter fixed effects proxies from the SMM estimation; the usual importer-industry dummies are included, and standard errors are clustered by country pair.

Table 4
Welfare Counterfactuals I
Quantifying the effect of distance and policy barriers

	% Welfare Change				Decomposition Due to change in:			Correlation
	Min.	Max.	Std. Dev.	Wtd. Avg.	Country GDP	Prices ($k \geq 1$)	Prices ($k = 0$)	
Reducing Specific Barriers:								
Log (Distance)	−15.3	67.1	17.2	25.6	87.3	5.1	−66.8	0.54***
			OECD:	7.0	12.7	4.0	−9.7	
			ROW:	29.8	104.3	5.3	−79.8	
Halving distance mark-up	−0.9	37.2	6.3	8.5	32.3	1.0	−24.7	0.17
			OECD:	3.5	11.9	0.7	−9.1	
			ROW:	9.6	36.9	1.0	−28.3	
Global GATT	−2.5	54.6	12.6	6.9	27.7	0.3	−21.1	−0.89***
			OECD:	−0.1	−0.4	0.0	0.3	
			ROW:	8.5	34.1	0.3	−25.9	

Notes: Based on the SMM estimates from the full sample. 'Wtd. Avg.' reports the population-weighted mean welfare change. The decomposition breaks down this mean 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 ($k = 0$). For the first two panels, the final column reports the cross-country Pearson correlation between the percent welfare change and an average distance measure (the log GDP-weighted average distance from the 82 countries in the sample). For the third panel, the final column reports the correlation between percent welfare change and the initial state of the GATT dummy. *** denotes significance at the 1% level. The "OECD" average comprises 23 high-income OECD countries (Australia, Austria, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, Great Britain, Greece, Ireland, Iceland, Italy, Japan, Luxembourg, the Netherlands, Norway, New Zealand, Portugal, Sweden, and the USA). The "ROW" (rest of the world) average is for the remaining 59 countries in the sample.

Table 5
Welfare Counterfactuals II
Quantifying the effect of raising country characteristics

	% Welfare Change				Decomposition Due to change in:			Correlation with cty. char.
	Min.	Max.	Std. Dev.	Wtd. Avg.	Country GDP	Prices ($k \geq 1$)	Prices ($k = 0$)	
Raising Factor endowments:								
$s_h \times \max(\log(H/L))$	−0.7	45.8	11.1	18.5	64.4	0.5	−46.5	−0.89***
$s_h \times (1+\text{s.d.})\log(H/L)$	0.4	6.4	1.5	3.3	−0.8	0.4	3.7	−1.00***
$s_k \times \max(\log(K/L))$	−0.4	105.2	27.9	42.0	16.7	0.1	25.2	−1.00***
$s_k \times (1+\text{s.d.})\log(K/L)$	0.2	89.2	24.3	38.1	−0.1	0.2	38.1	−1.00***
Raising Institutional attributes:								
$CAPDEP \times \max(FINDEV)$	−25.7	23.1	7.5	0.4	0.7	0.5	−0.8	−0.13
$CAPDEP \times (1+\text{s.d.})FINDEV$	−1.1	0.4	0.3	−1.7	0.2	1.3	0.1	0.68***
$HI \times \max(LEGAL)$	−0.1	17.1	3.2	2.3	9.7	0.1	−7.5	−0.73***
$HI \times (1+\text{s.d.})LEGAL$	0.1	0.2	0.0	0.1	−0.0	0.1	0.0	−0.20*
$RS \times \max(LEGAL)$	−0.3	47.5	12.9	13.2	55.5	0.4	−42.7	−0.89***
$RS \times (1+\text{s.d.})LEGAL$	0.3	0.5	0.0	0.4	−0.2	0.5	0.2	0.41***
$COMPL \times \max(LEGAL)$	−0.0	8.0	1.8	1.7	7.2	0.1	−5.5	−0.71***
$COMPL \times (1+\text{s.d.})LEGAL$	0.0	0.1	0.0	0.1	−0.1	0.1	0.1	0.77***
$COMPL \times \max(\log(H/L))$	−0.1	55.1	13.4	23.3	84.9	0.1	−62.3	−0.89***
$COMPL \times (1+\text{s.d.})\log(H/L)$	0.5	6.4	1.5	3.5	−0.6	0.6	3.2	−1.00***
Total effect:								
$\max(\log(H/L))$	0.5	97.0	22.1	37.6	157.2	1.2	−120.8	−0.90***
$(1+\text{s.d.})\log(H/L)$	0.2	1.1	0.2	0.7	−1.4	1.0	1.0	0.77***
$\max(LEGAL)$	−0.1	62.9	16.7	17.7	74.4	0.1	−57.3	−0.90***
$(1+\text{s.d.})LEGAL$	0.4	0.6	0.1	0.5	−0.3	0.6	0.2	0.51***

Notes: Based on the SMM estimates from the full sample. 'Wtd. Avg.' reports the population-weighted mean welfare change. The decomposition breaks down this mean 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 ($k = 0$). The 'max' rows refer to policy experiments that raise all countries to the world frontier (the maximum value in the sample) for that country characteristic. The '$(1 + \text{s.d.})$' rows refer to policy shocks that raise all countries by 1 standard deviation for that characteristic. 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%, 5% and 10% levels respectively.

Table 6
Welfare Counterfactuals IIIA
Policy experiments: Raising one country's (IDN) attributes

	IDN rank (out of 82)	% Welfare Change for IDN Due to change in: Total	Country GDP	Prices ($k \geq 1$)	Prices ($k = 0$)	% Welfare Change for sample Min.	Max.	Std. Dev.	Wtd. Avg.	Correlation with cty. char.
Raising:										
$\max(\log(H/L))$	30	38.7	147.9	0.7	-109.8	-0.1	38.7	4.3	1.9	-0.07
$(1+\text{s.d.})\log(H/L)$	30	17.6	58.9	0.2	-41.4	-0.4	17.6	1.9	0.8	-0.06
$\max(\log(K/L))$	33	43.8	17.7	0.1	26.1	-0.0	43.8	0.5	1.9	-0.03
$(1+\text{s.d.})\log(K/L)$	33	42.5	12.0	0.0	30.5	0.0	42.5	4.7	1.8	-0.03
$\max(FINDEV)$	23	8.8	40.0	0.1	-31.4	-0.2	8.8	1.0	0.4	-0.12
$(1+\text{s.d.})FINDEV$	23	0.5	2.9	0.0	-2.4	-0.03	0.5	0.1	0.0	-0.12
$\max(LEGAL)$	36	20.2	85.8	0.0	-65.9	-0.3	20.2	2.2	1.0	-0.05
$(1+\text{s.d.})LEGAL$	36	11.1	47.8	0.1	-36.9	-0.4	11.1	1.2	0.5	-0.06

Notes: Based on the SMM estimates from the full sample. The decomposition breaks down the total welfare change for IDN into that due to the change in country GDP, the change in the differentiated goods price index ($k \geq 1$), and the change in the price of domestic non-tradables ($k = 0$). 'Wtd. Avg.' reports the population-weighted mean welfare change. 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%, 5% and 10% levels respectively.

Table 7
Welfare Counterfactuals IIIB
Policy experiments: Impact on IDN's domestic industry structure

SIC	Industry description	Raising $\log(H/L)$ % change WLD share	Raising $\log(H/L)$ % change IDN share	Raising $LEGAL$ % change WLD share	Raising $LEGAL$ % change IDN share
20	Food products	10.5	4.6	92.6	7.1
21	Tobacco products	44.5	−24.5	97.4	11.9
22	Textile mills products	15.7	−9.5	91.9	6.4
23	Apparel	7.7	−104.7	30.8	−54.9
24	Wood products	28.1	66.5	122.2	36.7
25	Furniture	31.2	44.4	117.9	32.3
26	Paper products	36.9	−22.5	83.9	−1.5
27	Printing	54.1	178.4	157.3	71.8
28	Chemical products	30.6	130.4	162.2	76.6
29	Petroleum refining	17.3	23.5	123.1	37.6
30	Rubber and misc plastics	34.8	53.7	130.0	44.5
31	Leather products	19.4	43.0	163.9	78.4
32	Stone, clay, glass, concrete	28.3	−10.2	94.3	8.8
33	Primary metal industries	19.3	70.5	112.3	26.8
34	Fabricated metal products	18.3	−51.9	56.4	−29.0
35	Machinery and computers	51.2	139.7	173.2	87.7
36	Electronic products	5.9	−3.2	80.8	−4.7
37	Transportation equipment	18.1	−1.6	90.2	4.7
38	Instruments	21.3	199.8	101.3	15.7
39	Misc manufacturing	39.8	110.1	141.6	56.1
	Correlation:	With s_h: 0.59***		With HI: −0.08	
		With $COMPL$: 0.72***		With RS: 0.04	
				With $COMPL$: 0.30	

Notes: Based on the SMM estimates from the full sample. The two policy experiments illustrated are: (i) Raising IDN's human capital-labor ratio to the highest in the sample; and (ii) Raising IDN's $LEGAL$ score to the highest in the sample. '% change in WLD share' reports the percent change in IDN's share of total production in the industry for the 82-country sample. '% change in IDN share' reports the percent change in the industry output as a share of total IDN production. The (Pearson) correlations reported are between the percent changes and the corresponding industry characteristic; note that the correlation between either percent change and the industry characteristic is the same, since the percent changes in WLD share and IDN share are equal up to a constant. *** denotes significance at the 1% level.

Appendix Table 1A
Summary of Manufacturing Industry Characteristics
(20 industries, SIC-87 2-digit level)

	Min.	10th	25th	Med.	75th	90th	Max.	Std. Dev.
Skill intensity (s_h)	0.010	0.028	0.047	0.061	0.083	0.121	0.158	0.036
Capital intensity (s_k)	0.102	0.192	0.228	0.278	0.315	0.386	0.549	0.093
Materials intensity (s_m)	0.347	0.365	0.470	0.504	0.603	0.654	0.874	0.121
Ext. Capital Dep. ($CAPDEP$)	−1.206	−0.751	−0.148	−0.028	0.165	0.587	0.941	0.498
Input Concentration (HI)	0.064	0.100	0.122	0.140	0.155	0.187	0.324	0.051
Input Relationship-Spec. (RS)	0.530	0.590	0.808	0.926	0.966	0.979	0.988	0.146
Job Complexity ($COMPL$)	0.148	0.154	0.329	0.402	0.588	0.728	1	0.221
Sales Volatility ($SVOL$)	0.124	0.130	0.144	0.152	0.179	0.198	0.219	0.026

Appendix Table 1B
Pairwise Correlation of Manufacturing Industry Characteristics
(20 industries, SIC-87 2-digit level)

	s_h	s_k	s_m	$CAPDEP$	HI	RS	$COMPL$
s_k	0.34						
s_m	−0.70***	−0.84***					
$CAPDEP$	0.53**	−0.14	−0.06				
HI	−0.08	0.01	0.17	−0.10			
RS	0.70***	0.29	−0.64***	0.18	−0.15		
$COMPL$	0.58***	0.30	−0.31	0.65***	0.01	0.22	
$SVOL$	0.02	−0.26	0.17	0.38*	−0.11	−0.16	0.09

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

Appendix Table 2A
Summary of Country Characteristics

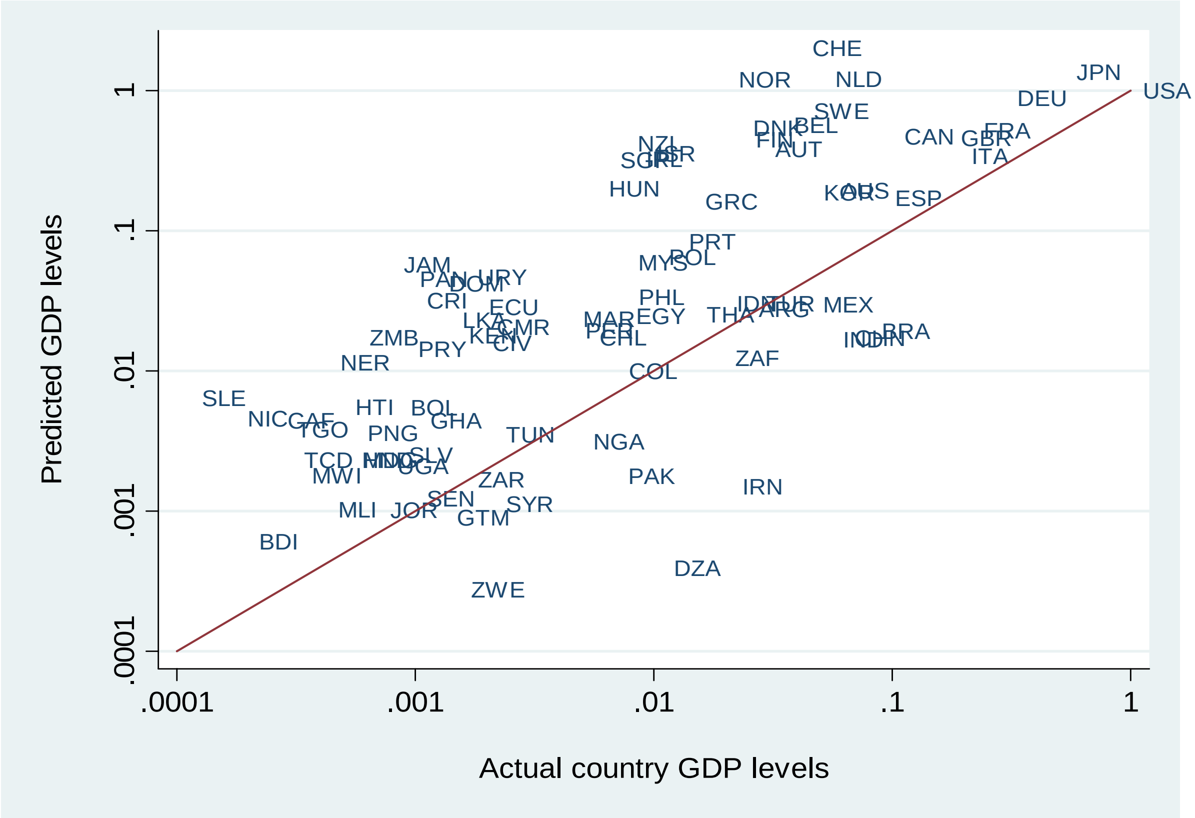
	Min.	10th	25th	Med.	75th	90th	Max.	Std. Dev.
$\log(H/L)$	0.07	0.26	0.39	0.59	0.81	1.04	1.21	0.29
$\log(K/L)$	5.76	7.05	8.33	9.71	10.83	11.32	11.59	1.59
$\log(Forest/L)$	−11.04	−6.70	−5.67	−4.88	−3.10	−2.25	−1.22	1.88
$\log(Arable/L)$	−11.09	−6.40	−5.68	−5.10	−4.58	−4.17	−2.83	1.08
Financial Devt. ($FINDEV$)	0.01	0.10	0.16	0.28	0.51	0.79	1.38	0.30
Legal Quality ($LEGAL$)	0.17	0.26	0.35	0.495	0.67	0.79	0.83	0.19
Labor Mkt. Flexibility ($FLEX$)	0.1	0.34	0.43	0.57	0.7	0.83	1	0.20

Appendix Table 2B
Pairwise Correlation of Country Characteristics

	$\log(H/L)$	$\log(K/L)$	$\log(Forest/L)$	$\log(Arable/L)$	$FINDEV$	$LEGAL$
$\log(K/L)$	0.81***					
$\log(Forest/L)$	−0.03	−0.09				
$\log(Arable/L)$	−0.04	−0.07	0.55***			
$FINDEV$	0.58***	0.65***	−0.23**	−0.34***		
$LEGAL$	0.69***	0.63***	−0.12	−0.13	0.68***	
$FLEX$	0.45***	0.37***	−0.15	−0.31***	0.30***	0.29***

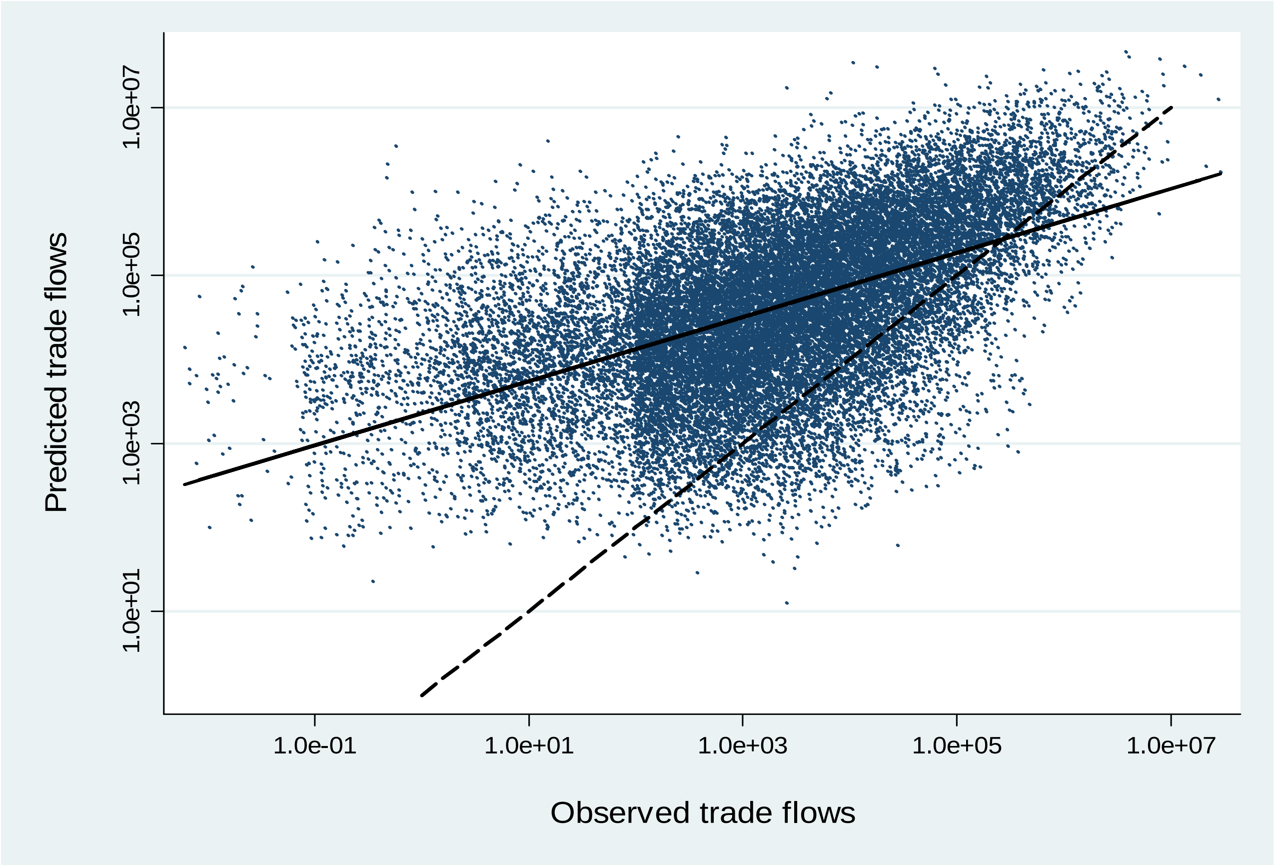
Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

Figure 1
Assessing the Goodness of Fit: Predicted Income Levels vs Observed Country GDP Levels
(normalized, US=1)



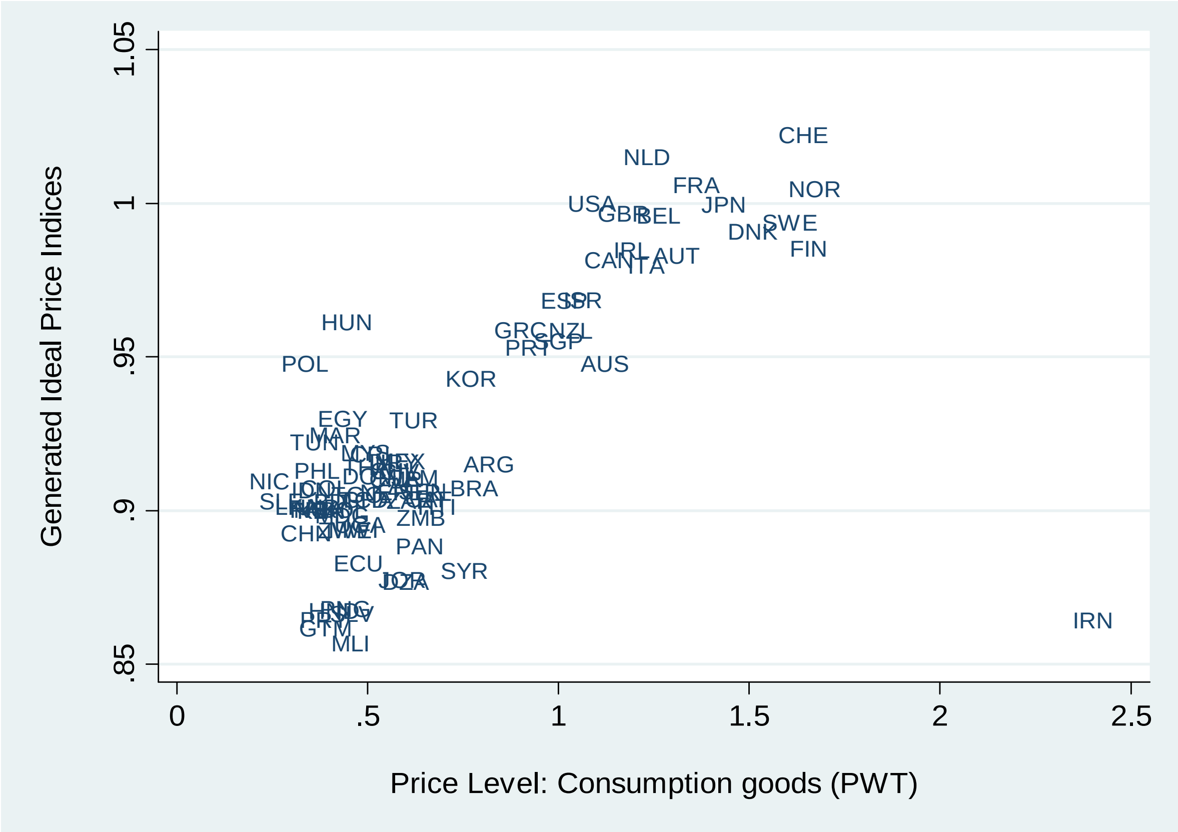
Notes: Observed country GDP levels plotted on the horizontal axis are from the World Development Indicators (WDI). The predicted country GDPs on the vertical axis are computed using the SMM estimates from the full sample. The US income level is normalized to 1. Both axes employ a log-scale; the original units are in thousands of current (1990) US dollars. The 45-degree line is plotted for reference. The Pearson correlation between the two log-income variables is 0.71, while the Spearman rank correlation is 0.69, both significant at the 1% level.

Figure 2
Assessing the Goodness of Fit: Predicted Trade Flows vs Observed Trade Flows



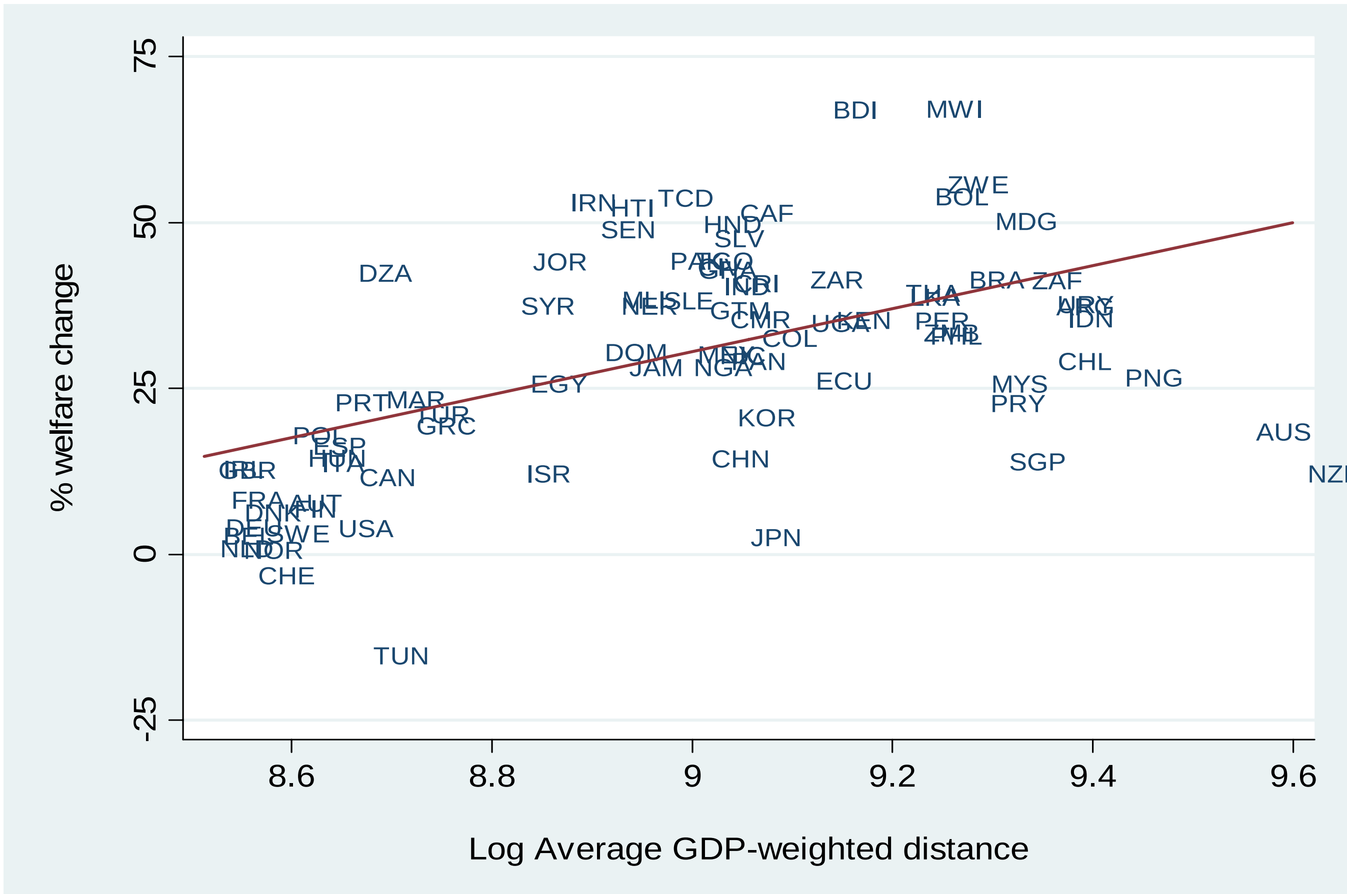
Notes: Observed trade flows plotted on the horizontal axis are from Feenstra et al. (2005), concorded to 2-digit SIC-87 industrial groups. Predicted trade flows on the vertical axis are generated from the model using the SMM estimates from the full sample. Both axes employ a log-scale; the original units are in thousands of current (1990) US dollars. The dashed line is the 45-degree line, while the solid line is the log-linear regression line (slope = 0.77, significant at the 1% level).

Figure 3
Assessing the Goodness of Fit: Generated Ideal Price Indices vs Observed Price Indices
(normalized, US=1)



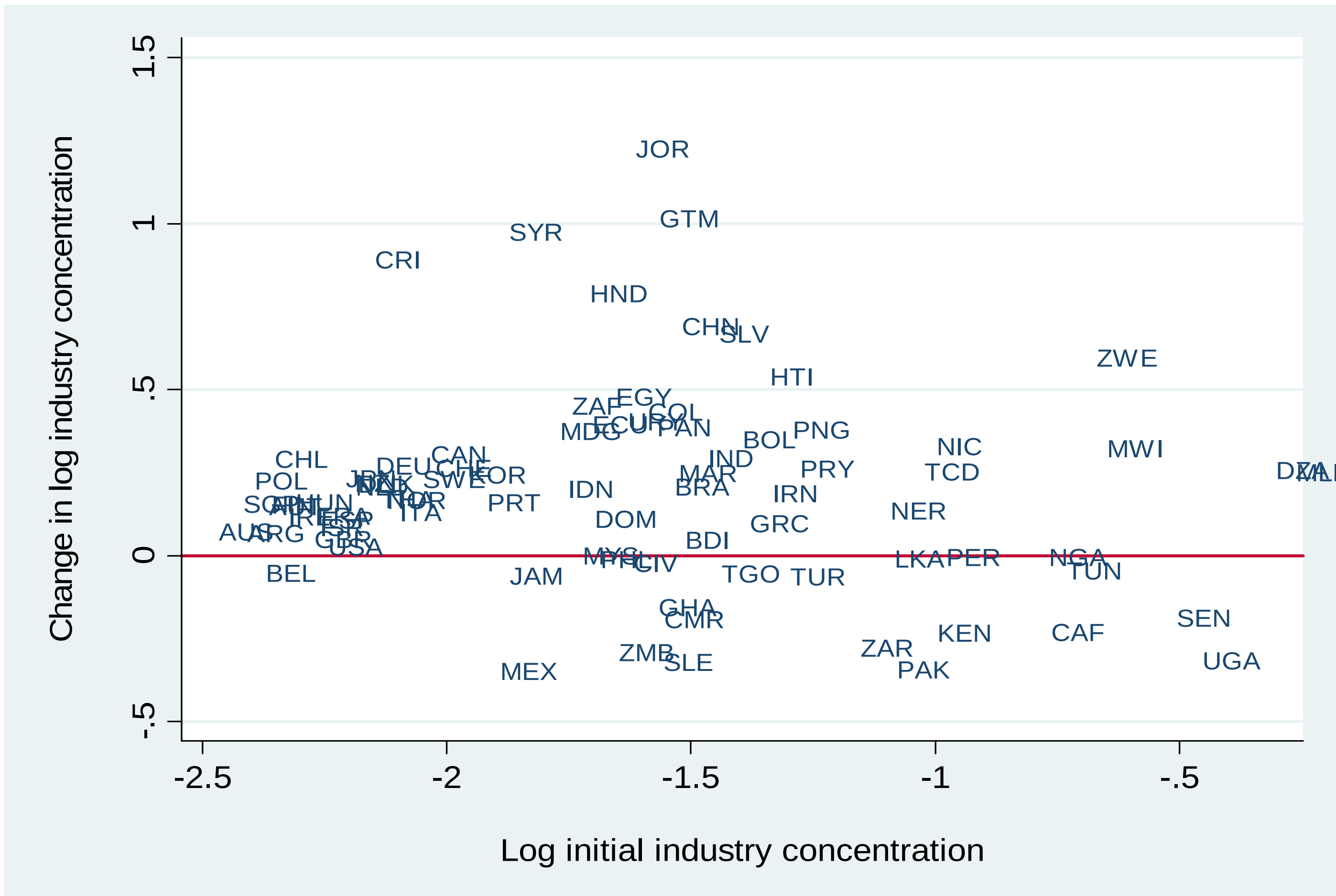
Notes: The horizontal axis plots country price indices for consumption goods from the Penn World Tables (PWT), adjusted for purchasing power parity. The vertical axis plots the country ideal price indices for manufactures generated by the model, using the SMM estimates from the full sample. Both axes are normalized with the US value equal to 1. The Pearson correlation between the two variables is 0.68, while the Spearman rank correlation is 0.60, both significant at the 1% level.

Figure 4
Welfare Effect of a Counterfactual Removal of Physical Distance Barriers



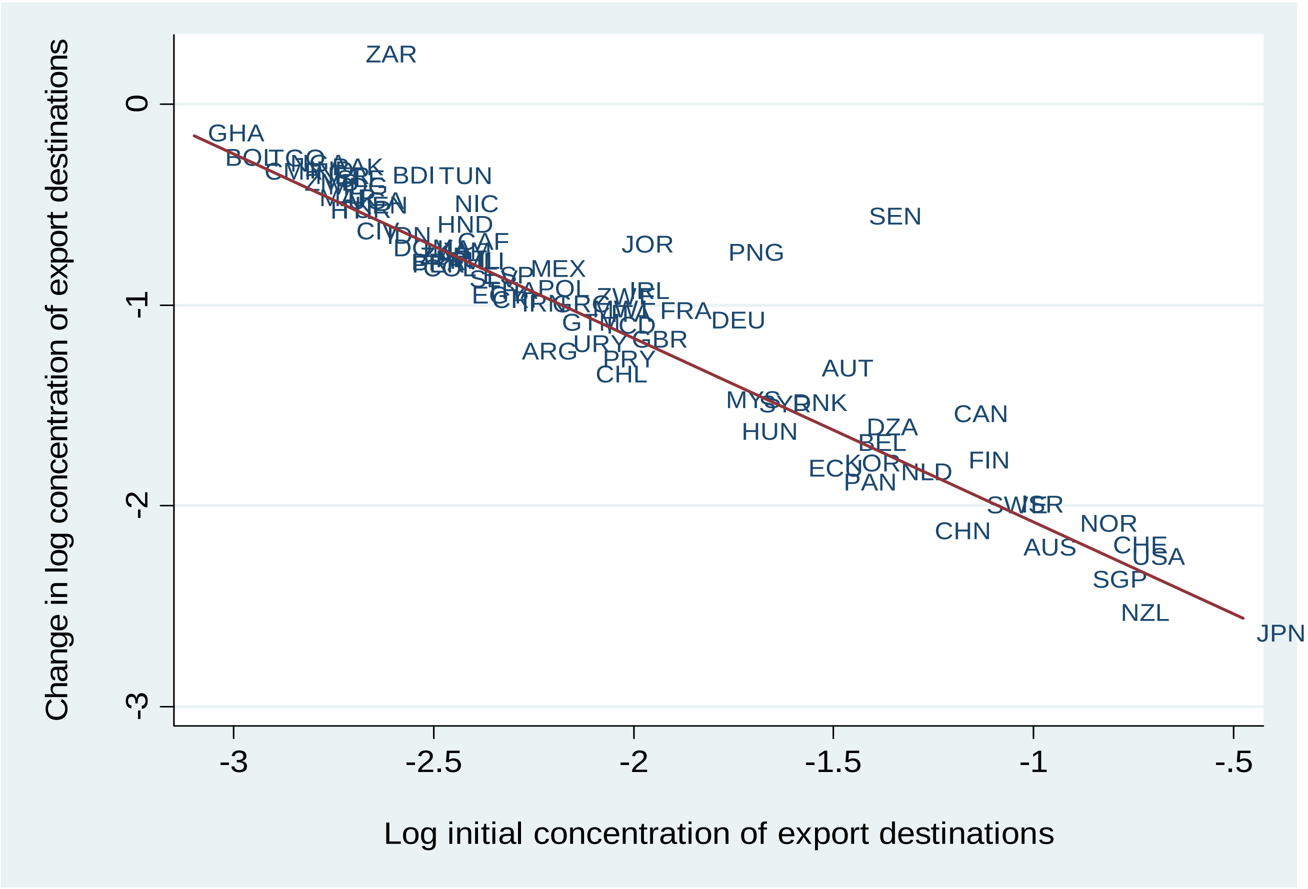
Notes: The vertical axis plots the percent welfare change from a hypothetical removal of physical distance, as described in Section 5.1. This is plotted against the log mean GDP-weighted distance of each country from all countries in the sample. The linear best fit line is illustrated (slope=32.39, significant at the 1% level).

Figure 5
Increase in Industrial Specialization from a Counterfactual Removal of Physical Distance Barriers



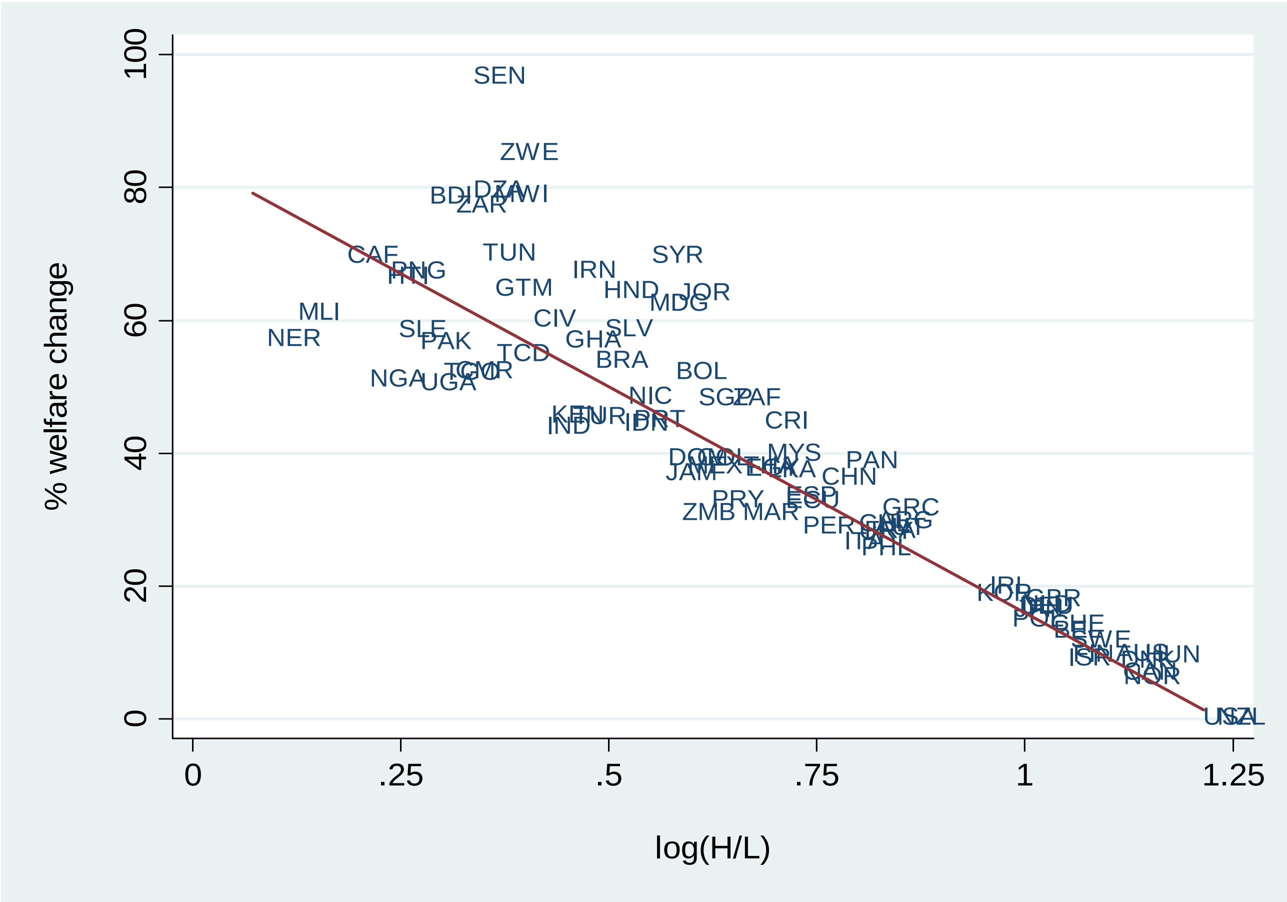
Notes: For each country, industry concentration is calculated as the Herfindahl index of industry output. The log initial value is plotted on the horizontal axis, while the change in log concentration is on the vertical axis. The zero-change line is illustrated.

Figure 6
Diversification of Export Destinations from a Counterfactual Removal of Physical Distance Barriers



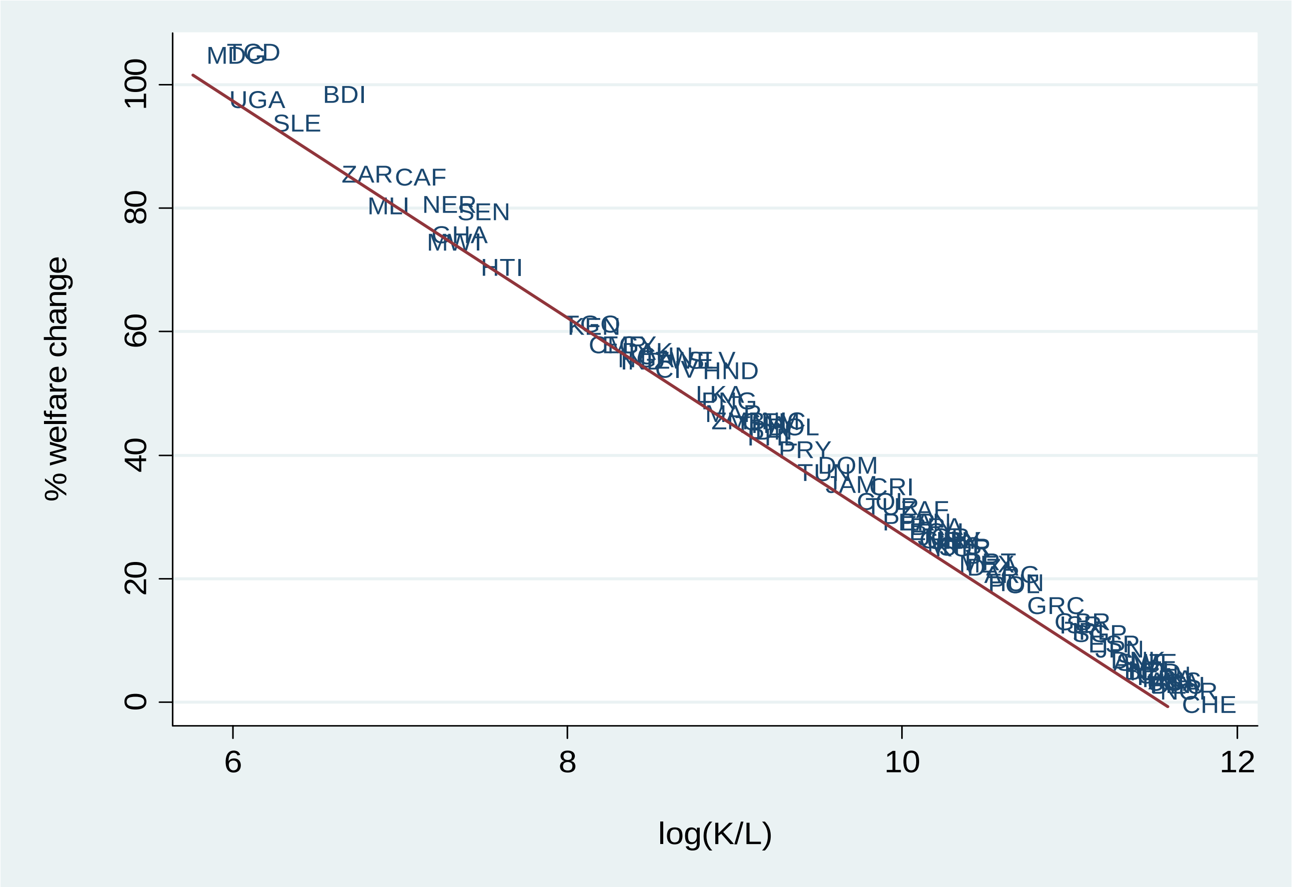
Notes: For each country, the concentration of export destinations is equal to the sum of squared export destination shares in total domestic exports. The log initial value is plotted on the horizontal axis, while the change in log concentration is on the vertical axis. The linear best fit line is illustrated (slope = -0.92, significant at the 1% level).

Figure 7
Welfare Effect of a Counterfactual Increase in Human Capital Endowment



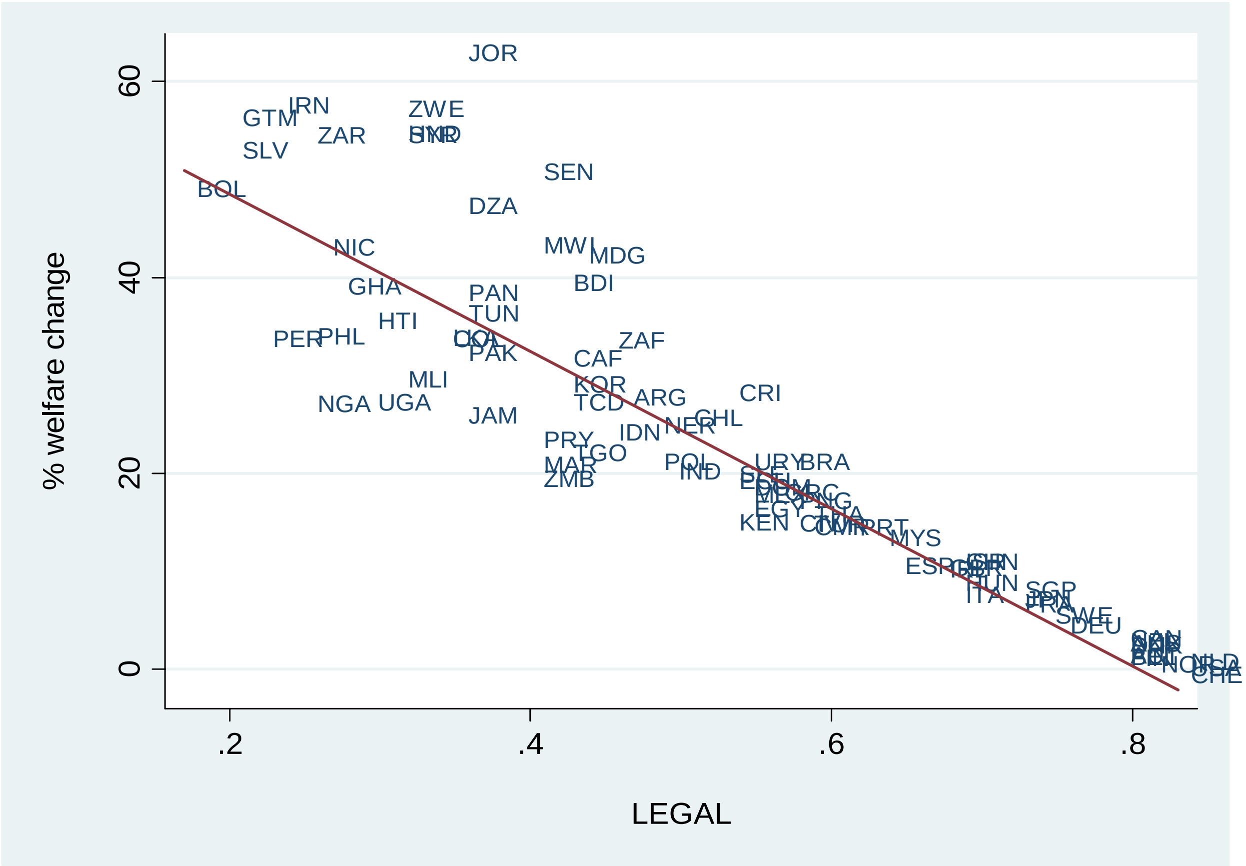
Notes: The vertical axis plots the percent welfare change from increasing all countries' human capital endowment to the world frontier (the highest value observed in the sample), as described in Section 5.2. This is plotted against the initial log human capital-labor ratio of each country. The linear best fit line is illustrated; the Pearson linear correlation is -0.90, significant at the 1% level.

Figure 8
Welfare Effect of a Counterfactual Increase in Physical Capital Endowment



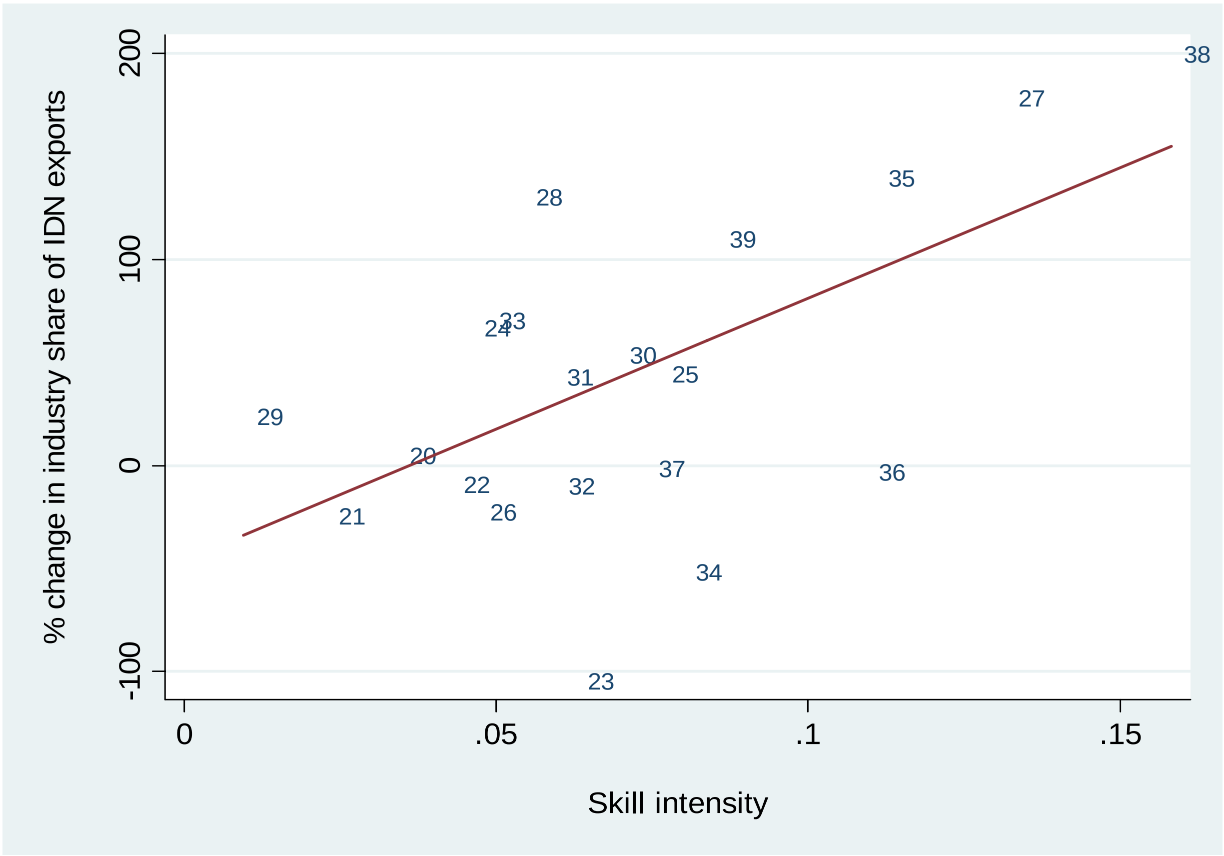
Notes: The vertical axis plots the percent welfare change from increasing all countries' physical capital endowment to the world frontier (the highest value observed in the sample), as described in Section 5.2. This is plotted against the initial log capital-labor ratio of each country. The linear best fit line is illustrated; the Pearson linear correlation is -1.00, significant at the 1% level.

Figure 9
Welfare Effect of a Counterfactual Improvement in Legal Institutions



Notes: The vertical axis plots the percent welfare change from improving all countries' legal institutions to the world frontier (the highest index value observed in the sample), as described in Section 5.2. This is plotted against the initial index value of LEGAL for each country. The linear best fit line is illustrated; the Pearson linear correlation is -0.90, significant at the 1% level.

Figure 10
Effect of Raising Indonesia's Human Capital Endowment on Domestic Industrial Composition



Notes: The vertical axis plots the percent change in each industry's share of Indonesia's total exports (including domestic absorption) following an increase in Indonesia's human capital endowment to the world frontier (the highest value observed in the sample), as described in Section 5.2. This is plotted against the skill intensity of each industry. The linear best fit line is illustrated; the Pearson linear correlation is 0.59, significant at the 1% level.