Unpacking Sources of Comparative Advantage: A Quantitative Approach*

Davin Chor†
Singapore Management University

First version: November 2006
This version: October 2008

Comments welcome

Abstract

This paper develops an approach for quantifying the importance of different sources of comparative advantage for country welfare. To explain patterns of specialization, I present a multi-country trade 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 and institutional strengths. I estimate the model parameters on a large dataset of bilateral trade flows, presenting results from both a baseline OLS approach, as well as a simulated method of moments (SMM) procedure to account for the prevalence of zero trade flows in the data. I apply the model to explore various quantitative questions, in particular how much distance, Ricardian productivity, factor endowments, and institutional conditions each matter for country welfare in the global trade equilibrium. I also illustrate the shift in industry composition and the accompanying welfare gains in policy experiments where a country raises its factor endowments or improves the quality of its institutions.

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. I thank Pol Antràs, Gary Chamberlain, Elhanan Helpman, and Marc J. Melitz for their patient advice and guidance. I have also benefited from comments from Leon Berkelmans, Filipe R. Campante, Arnaud Costinot, Jonathan Eaton, Doireann Fitzgerald, Hiroyuki Kasahara, 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, Singapore Management University, the Midwest International Economics Group Meeting (Michigan), the North American Winter Meetings of the Econometric Society (New Orleans), the Empirical Investigations in Trade and Investment meeting (Keio University), and the Far East Meetings of the Econometric Society (Singapore). Marc J. Melitz and Nathan Nunn generously shared their data, while Kalina Manova, Stephen F. Lin and Stanley Watt provided Stata code. Gurmeet Singh Ghumann provided excellent research assistance. All errors are my own.

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

1 Introduction

The concept of comparative advantage has been the basis of our understanding of the pattern of international trade 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 country differences in productivity or factor endowments, that determine patterns of specialization and trade.

On the Ricardian model, it is only recently that Eaton and Kortum (2002) showed how one can derive analytic expressions for trade flows in a general multi-country setting, by providing a parametrization of the underlying distribution that governs country productivity levels. The good fit of their model to the manufacturing trade data delivered an important piece of evidence on the role of productivity differences in determining comparative advantage (Eaton and Kortum 2002, Costinot and Komunjer 2007).[1] Separately, several studies have reaffirmed the role of factor endowments for explaining trade patterns within the Heckscher-Ohlin framework. These have shown that countries tend to be net exporters of their relatively abundant factors in North-South bilateral trade (Debaere 2003), and that countries also export more in industries that use these abundant factors more intensively (Romalis 2004).[2] Moving beyond this neoclassical focus, a recent cluster of work has identified how country institutions can augment productivity, particularly in industries that are dependent on these institutional provisions to facilitate production. Such institutional sources of comparative advantage include: financial development (Beck 2003, Manova 2006), the security of contract enforcement (Levchenko 2007, Nunn 2007, Costinot 2007), and labor market flexibility (Cuñat and Melitz 2007).[3]

This paper develops a methodology for quantifying the importance of these various sources of comparative advantage for country welfare in a global trade equilibrium. As a benchmark structural framework, I present an extension of the Eaton-Kortum (EK) model that goes beyond aggregate trade volumes to explain the cross-country pattern of specialization and industry trade flows. In the model, the productivity level of firms is composed of a systematic and a stochastic component, where the systematic component is driven by the interaction between country and industry characteristics. The motivation for this is intuitive: Industries vary in the technological and institutional conditions needed for production,

[1]Ricardian models of an earlier vintage, such as Dornbusch et al. (1977), could not easily be taken to empirical work, largely because these were two-country models featuring complete specialization (each good exported by precisely one country), which is clearly inconsistent with the large volume of intra-industry trade observed in practice. Most earlier studies instead tested more reduced-form Ricardian implications, such as whether countries tend to export more in industries where domestic productivity is higher, without a full theoretical model in mind. For a review of this earlier empirical work, see Section 3 of Deardorff's (1984) Handbook chapter; for a more recent example, see Golub and Hsieh (2000).

[2]While these studies focus on the correlation between relative factor endowments and trade patterns, there is a related vast literature seeking to explain the *absolute* levels of the factor content of trade. A key puzzle here was 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).

[3]Strictly speaking, these institutional explanations of trade flows can be viewed as a subset of the Ricardian model, insofar as the mechanism through which institutions operate is to boost domestic productivity in specific industries.

and countries differ in their ability to provide for these industry-specific requirements. Comparative advantage therefore stems in practice from such country-industry matches.

At heart, this empirical specification draws on existing 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 import market share in industries that use their abundant factors more intensively.[4] The literature on institutional determinants of trade has also adopted this empirical strategy, by applying or constructing novel measures of an industry's dependence on particular institutional conditions. 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.[5] Similarly, several studies have shown that countries with better institutional rule of law export relatively more in industries that are more exposed to holdup problems or other institutional frictions, as measured by input concentration (Levchenko 2007), the share of customized inputs (Nunn 2007), or job task complexity (Costinot 2007).[6] Cuñat and Melitz (2007) have also demonstrated that countries with flexible labor markets facilitate specialization in more volatile industries, as these are the industries that benefit most from being able to adjust employment margins regularly.

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

Section 3 estimates the closed-form trade flow expressions using ordinary least-squares (OLS) methods, to provide a first-pass test of the model. For the empirical implementation, I assembled a large dataset of bilateral industry trade flows, pairwise distance measures, as well as country and industry characteristics for a sample of 83 countries and 20 manufacturing industries. This includes a comprehensive set of all the country-industry interaction terms from the papers cited above, which facilitates a comparison with the results in the existing literature. Here, I find strong corroborating evidence for the

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

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

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

importance of factor endowments, financial development, legal institutions, and labor market regimes as sources of comparative advantage, even when all interaction terms are run in one regression. This represents a first exercise (to the best of my knowledge) at jointly verifying the significance of this extensive a list of institutional determinants of trade from prior studies.

While OLS provides a useful baseline, it suffers from the drawback that zero trade observations are dropped when log trade flows are the dependent variable. These zeros constitute about two-thirds of the dataset, and discarding this sizeable amount of information can systematically bias the OLS coefficients (see Santos-Silva and Tenreyro 2006 and Helpman et al. 2007, among others).[7] It would thus be important to first account 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 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 no longer have explicit 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 in the actual data with those from trade flows simulated from the model (Pakes and Pollard 1989).

Using these SMM estimates, I explore the quantitative implications of the model for the global trade equilibrium in Section 5. A first set of counterfactual exercises relates to distance and geography. The model implies a sizeable average increase in country welfare (15.7%) from a hypothetical move to a world where measures of distance barriers are minimized, comparable to what EK find for their OECD sample (16.1%-24.1%). The transition path itself reveals interesting patterns: An initial reduction in distance favors existing producer countries in each industry, raising the concentration of the location of production. Only as barriers are reduced more substantially do new countries emerge as viable exporters, eventually lowering the concentration of production location below the status quo level.

Second, the structural framework allows us to assess the relative importance of the various sources of comparative advantage for country welfare. I do so by shutting down country-by-country the relevant terms in the empirical model that capture each of these comparative advantage forces. The calculations indicate that within the context of this model, Ricardian forces, Heckscher-Ohlin forces, and distance barriers are almost equally important from a welfare perspective: The average country welfare loss from shutting off either the Ricardian or the Heckscher-Ohlin terms lies in the same ballpark (−36.6% and −34.5% respectively), and this is similar to the welfare loss (−37.8%) when raising the distance markup

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

country-by-country to the maximum level in the sample. The institutional determinants of trade, which are strictly speaking a subset of the Ricardian forces, also play a substantial role. Legal institutions feature prominently here, with welfare suffering the most when the channels from Levchenko (2007) and Nunn (2007) related to how the contracting environment alleviates holdup problems are shut down.

Last but not least, the model can be readily applied to consider country policy experiments. For example, several developing countries have pursued concerted policies of capital accumulation to expand their exports and thereby promote growth, and it would be interesting to evaluate the gains from such policy moves. To this end, I examine policy shocks to an illustrative large developing country (Indonesia). I demonstrate the underlying shift in industry structure and the ensuing welfare gains from such changes as an increase in factor endowments or an improvement in the quality of domestic institutions.

This paper falls within a broader research agenda seeking to understand the importance of different determinants of trade flows, often by developing variants of the traditional gravity equation. Such models have been used to evaluate various 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 2007).[8] While the methodology developed in this paper facilitates similar distance-related exercises, it further enables the researcher to explore policy experiments involving country characteristics that matter for comparative advantage, to examine the impact on industry structure, trade patterns and country welfare.

Several recent papers have sought to take a more holistic view on sources of comparative advantage by incorporating both Ricardian and Hescksher-Ohlin forces in empirical work (as does this paper), in order to better account for the determinants of trade flows at the industry level (Harrigan 1997, Morrow 2008).[9] A closely-related paper along these lines is Shikher (2007), who explores an alternative extension of the EK model to explain industry trade flows. Empirically, Shikher calibrates country-industry technology parameters in order to fit the output and trade data, whereas the approach that I take here will instead be to relate these productivity parameters to observable country and industry characteristics.

In terms of empirical machinery, SMM methods have previously been applied to estimate the structural parameters of various versions of the EK model (Bernard et al. 2003, Eaton et al. 2005). The specific approach in this paper is most similar to Ramondo (2008), who employed an SMM procedure to estimate a structural model of multinational activity in a manner consistent with the prevalence of zeros. At this juncture, it is useful to note that the methodology developed in this paper is in fact very

[8] 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 (2007). Of note, Lai and Trefler (2002) expressed a healthy reservation about the counterfactuals they computed, as they documented several dimensions, such as the implied price elasticities, along which their model appears misspecified.

[9] In the same spirit, Davis and Weinstein (2001) found that allowing for productivity differences across countries in the traditional Vanek equations was one way to reduce the extent of the "missing trade" paradox.

general and is clearly not limited to the specific set of interactions considered here. 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 raise computational cost for the SMM estimation.

The roadmap for the paper is as follows. Section 2 extends the canonical EK model to explain industry trade flows. Section 3 presents 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 assess the fit of the model to 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.

2 A Benchmark Model of Industry Trade Flows

2.1 The basic setup

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

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

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

where $Q_n^k(j^k)$ denotes the quantity of variety j^k from industry k consumed in country n. (In what follows, I suppress the superscript k for varieties unless there is cause for confusion.) Utility from tradables is aggregated via a nested constant elasticity of substitution (CES) function. Define $\varepsilon = 1/(1-\alpha) > 1$ to be the elasticity of substitution between any two varieties from the same industry, and $\phi = 1/(1-\beta) > 1$ to be the corresponding elasticity between varieties drawn from different industries. I assume that $\varepsilon > \phi$, so that varieties from the same industry are closer substitutes than varieties from different industries. Total utility is a Cobb-Douglas aggregate over the consumption of tradables and non-tradables, with the share of income spent on tradables equal to $\eta \in (0, 1)$.

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 non-tradable good is the domestic numeraire.) Solving this optimization program, it is straightforward to show that the demand for each tradable variety is:

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

where $(P_n^k)^{1-\varepsilon} = \int_0^1 (p_n^k(j))^{1-\varepsilon} dj$ is the ideal price index for industry k faced by consumers in country n. 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 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 producers of this variety. Following EK's notation, let $p_{ni}^k(j)$ denote the price that country i would charge for exporting variety j to country n (the 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, while $d_{ni}^k \geq 1$ denotes the iceberg transport cost incurred due to distance or geographic barriers. The $z_i^k(j)$ terms capture the Ricardian productivity of country i in the manufacture of variety j; formally, $z_i^k(j)$ is equal to the number of units of variety j that country i can produce using the same bundle of factors that would produce one unit under the baseline technology.

It is convenient to specify the unit production cost, c_i^k, to be a Cobb-Douglas aggregate over factor prices in country i, namely: $c_i^k = \prod_{f=0}^F (w_{if})^{s_f^k}$, where $f = 0, 1, \ldots, F$ indexes factors of production.[10] w_{if} is the local unit price of factor f, while $s_f^k \in (0,1)$ is the share of total factor payments in industry k that accrues to this factor. Under constant returns to scale, we have: $\sum_{f=0}^F s_f^k = 1$. Each firm takes the w_{if}'s as given, being too small to affect aggregate factor markets. These factor price terms capture the role of Heckscher-Ohlin forces, namely endowment-based production cost differences, in influencing trade patterns. Note that the model does not in general imply factor price equalization across countries due to the presence of productivity differences and transport cost barriers.

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

Productivity: In order to relate productivity to observable characteristics, I specify the log produc-

[10] 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.

tivity of country i in industry-k varieties to be:

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

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

The systematic component of productivity is driven by a linear combination of country characteristics (L_{il}, indexed by l) and industry characteristics (M_{km}, indexed by m). This embeds the idea that it is precisely the interaction between pairs $\{l, m\}$ of country and industry attributes that determines a country's productivity position in that industry. For example, countries where legal institutions securely enforce contracts will on average be more productive in industries that are more vulnerable to holdup problems between input suppliers and producers (Levchenko 2007, Nunn 2007). These $L_{il}M_{km}$ interaction terms thus serve primarily to capture the role of institutional determinants of the pattern of trade, with the β_{lm} coefficients parameterizing how important each institutional channel is for generating a productivity edge. Note that exporter and industry fixed effects (λ_i and μ_k) are also included to control for the average productivity level across all countries and industries respectively.

As for the stochastic component of productivity, the $\epsilon_i^k(j)$'s are independent draws from the Type I extreme-value (Gumbel) distribution, with cumulative distribution function (cdf) $F(\epsilon) = \exp(-\exp(-\epsilon))$. This is the natural counterpart to EK's specification of a Fréchet distribution for productivity levels, since the natural log of a Fréchet random variable inherits a Gumbel distribution.[11] This probability specification 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 presented by country i to country n for variety j in industry k is:

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

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

[11] The micro-foundation that EK offer for this distributional choice thus carries over to this extension of their model: Suppose that firm productivity levels within a country follow a Pareto distribution, an assumption that finds good support in the data (for example, see Helpman et al. 2004). Then, the order statistic for the maximum productivity level across all firms (hence the technological frontier of the country) is a Fréchet random variable. Costinot and Komunjer (2007) show that this distributional choice for the productivity shocks can be relaxed to some extent. In an earlier draft, Costinot (2007) also independently recognized this way of extending the EK model to the industry level.

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

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

where $\theta = \frac{1}{\beta_0}$ and $\varphi_i^k = \exp\{\theta\lambda_i + \theta\mu_k + \theta\sum_{\{l,m\}}\beta_{lm}L_{il}M_{km}\}$. Note that θ has the interpretation of an inverse productivity spread parameter. Also, φ_i^k is increasing in the systematic component of country i's productivity in industry k.[12]

2.2 Implications for trade flows

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

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

Also, let π_{ni}^k be the probability of country i being the lowest-price provider – and hence the unique exporter – of an industry-k variety to country n.[13] We have:

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

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

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

Observe that to evaluate the total value of industry-k consumption in country n in the denominator, I integrate over varieties j and the minimum price distribution, G_n^k. As pointed out by EK (2002), 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 conditional 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

[12] 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 additional parameters to estimate.

[13] 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.

industry-k market share in country n as a function of underlying country and industry characteristics, as well as bilateral distance.

It is instructive to re-express (10) by normalizing it with respect to country n's expenditure share from a fixed reference country, u:

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

This last equation has an intuitive interpretation: Country i's market share in country n (normalized by the market share of the reference 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 (φ_i^k/φ_u^k). As for 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 exported by country i to country n. Now, suppose for example 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 lower θ will raise $\frac{X_{ni}^k}{X_{nu}^k}$, so a larger spread in the productivity shocks shifts market shares in favor of the initially smaller exporting countries. 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 systematic productivity will nevertheless get a good enough productivity shock in some varieties to emerge as the lowest-price provider.

At this juncture, it is useful to highlight the close links between the expressions for trade flows and those in EK (2002). To recapitulate, EK develop a model of aggregate trade flows in which the exporter i productivity terms, $z_i(j)$, are independent draws from a Fréchet distribution, with cdf $F_i(z) = \exp(-T_i z^{\theta})$. Note that $T_i > 0$ is a country-specific location parameter that reflects 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, which is precisely equation (10) in EK:

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

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) in this paper: Both pairs of equations explain trade shares as a function of factor costs, distance barriers and productivity in a similar way, except that each term has now been replaced with its industry-specific counterpart. In particular, the more general productivity term, φ_i^k, takes the place of EK's technological parameter, T_i. This highlights the sense in which this paper unpacks sources of Ricardian comparative advantage, by positing a functional form for φ_i^k that relates productivity to observable characteristics to reflect how well countries are able to meet the requirements of industries along various technological and institutional dimensions.[14]

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 estimating the parameters of the model.

3 OLS Estimation of Model of Bilateral Industry Trade Flows

I present the OLS baseline estimates in this section. It turns out that the regression specifications for trade flows implied by (10) resemble closely those adopted in existing empirical work, and so the OLS results provide a basis for comparison and corroboration with the current literature on sources of comparative advantage. In addition, the regressions I run represent a first attempt (to the best of my knowledge) at jointly testing the significance of such a comprehensive list of institutional determinants of trade flows. These OLS coefficients provide a baseline against which to compare the SMM estimates later in Section 4, where we will deal with the potential bias from the omission of the zero trade flows.

3.1 Deriving the estimating equation

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

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

Here, $\beta_d D_{ni}$ is a linear combination of distance variables that impose an iceberg transport cost on trade. In the empirical implementation below, these D_{ni}'s will include physical distance, and indicator variables for shared linguistic ties, colonial links, border relationships, as well as trade agreements that reduce policy barriers to trade.[15] The distance markup is allowed to vary by industry through the fixed effect, δ_k, since transport costs may vary with the nature of the goods being shipped. Finally, trade between countries may be subject to idiosyncratic shocks, $\zeta_{ni} + \nu_{ni}^k$, which include a country-pair specific component (ζ_{ni}); I assume that these are iid draws from mean-zero normal distributions: $\zeta_{ni} \sim N(0, \sigma_\zeta^2)$, and $\nu_{ni}^k \sim N(0, \sigma_\nu^2)$.

[14]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.

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

Substituting this distance term (14) into (10), and making use of the fact that $s_0^k = 1 - \sum_{f=1}^F s_f^k$, it is straightforward to derive the following:

$$\ln(X_{ni}^k) = -\theta \sum_{f=1}^{F} \left(\ln \frac{w_{if}}{w_{i0}} \right) s_f^k \; + \; \theta \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} - \theta \beta_d D_{ni} + I_i + I_{nk} - \theta \zeta_{ni} - \theta \nu_{ni}^k \qquad (15)$$

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

In practice, since good data on factor prices is not readily available for a large sample of countries, 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.[16] This leads to the estimating equation:

$$\ln(X_{ni}^k) = \sum_{f=1}^{F} \theta \beta_f \left(\ln \frac{V_{if}}{V_{i0}} \right) s_f^k \; + \; \theta \sum_{\{l,m\}} \beta_{lm} L_{il} M_{km} - \theta \beta_d D_{ni} + I_i + I_{nk} - \theta \zeta_{ni} - \theta \nu_{ni}^k \qquad (16)$$

I therefore regress log bilateral industry trade flows, $\ln(X_{ni}^k)$, as a function of: (i) Heckscher-Ohlin forces, as picked up by the interaction between country factor endowments, $\left(\ln \frac{V_{if}}{V_{i0}}\right)$, and industry factor intensities, s_f^k; (ii) institutional forces, through the interaction between country institutional characteristics, L_{il}, and industry measures of dependence, M_{km}; (iii) bilateral distance variables, D_{ni}; (iv) exporter fixed effects, I_i; and (v) importer-industry fixed effects, I_{nk}. Standard errors are clustered by country pair, to allow for correlated shocks among observations from the same country pair ($-\theta\zeta_{ni}$). Equation (16) thus provides a neat decomposition of the determinants of trade flows, which embeds the empirical specifications in Romalis (2004) and the recent literature on institutional determinants of trade flows. (The $\theta\beta_f \left(\ln \frac{V_{if}}{V_{i0}}\right) s_f^k$ and $\theta L_{il} M_{km}$ terms on the right-hand side are essentially similar in spirit, in that both capture how well conditions in country i provide for the production needs of industry k.)

3.2 Discussion of OLS regression results

The empirical implementation uses a large dataset of bilateral trade flows, distance measures, and country and industry characteristics. The sample consists of 83 countries, listed in Table 1A, the largest number for which I could assemble a balanced dataset of all the country variables. For the differentiated products industries ($k \geq 1$), I work with the US 1987 Standard Industrial Classification (SIC-87) 2-digit manufacturing categories, a fairly broad level of industry aggregation. This provides 20 industry groups,

[16] 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.

listed in Table 1B, with SIC codes from 20 (food processing) to 39 (miscellaneous manufacturing). The analysis focuses on one year, 1990, the same year as in EK. This therefore abstracts from dynamic issues such as factor accumulation over time.

The trade data are from Feenstra et al. (2005)'s World Trade Flows database. The original data are in the Standard Industrial Trade Classification (SITC), Revision 2 format. I convert this to SIC-87 format using detailed information on the composition of US exports to derive concordance weights.[17] As for the country and industry characteristics, these were drawn directly from or constructed following closely the methodology of existing studies, in order to facilitate comparison with the literature. The Data Appendix documents the details on the data sources, as well as how I standardized the relevant variables into the common SIC-87 2-digit format. As far as possible, I use data from the immediate years preceding 1990. When multiple years are available, I use averages over 1980-89 to help smooth out the data from any single year. (Summary statistics of the industry and country variables, including means and pairwise correlations, are in Appendix Tables 1 and 2.)

The sample of 83 countries comprises 82.4% of all recorded manufacturing trade in 1990. While the total number of data points is $83 \times 82 \times 20 = 136,120$, only 45,034 (or 33.1%) of these contain a positive amount of trade. This pervasiveness of zeros is a major feature the bilateral trade data and it presents a challenge to consistent estimation of gravity equations, since the zeros are dropped from the regression sample when the dependent variable is log trade flows. The OLS results thus have to be interpreted strictly as estimates conditional on observing a positive trade flow. I return to this issue of correcting for the bias from discarding the zeros in the SMM estimation in Section 4.

Table 2 presents the baseline OLS regressions of (16), where the explanatory variables are introduced in turn. The recent literature, including Romalis (2004) and the papers on institutions and trade, each work with trade data that is concorded in different formats, at varying levels of aggregation. Table 2 thus verifies that the patterns identified in this preceding literature are also present in the dataset used here, which works with bilateral trade flows at the relatively coarse 2-digit level of industry aggregation.

Distance: Column (1) reports a basic specification that includes standard measures of trade barriers (D_{ni}) from the gravity equation literature. The OLS coefficients generally confirm the importance of distance in explaining trade patterns, although not all are statistically significantly. Of note, physical distance has a negative and significant effect ($\beta_{d1} = -1.152$) in impeding trade flows, the magnitude of which implies large effects: A hypothetical halving of physical distance would slightly more than double the volume of bilateral trade, raising it by a factor of $(0.5)^{-1.152} = 2.22$. Sharing a common language (β_{d2}) or colonial ties (β_{d3}) both raise the propensity for trade between countries. While the border effect (β_{d4}) is positive, this is not statistically significant. I also include two commonly-used dummy variables

[17]This procedure follows Cuñat and Melitz (2007), with the composition of US exports calculated from Feenstra et al. (2002). Please see the Data Appendix for details.

to capture aspects of trade policy. Joint membership in an RTA (β_{d5}) delivers a statistically significant boost to bilateral trade. However, I do not find a significant GATT effect (β_{d6}) with OLS (Rose 2004). These distance coefficients will remain remarkably stable even as more explanatory variables are included in subsequent specifications.

Heckscher-Ohlin: Column (2) demonstrates the relevance of Heckscher-Ohlin forces for the cross-country pattern of trade. Here, country measures of factor endowments per worker ($\log(H/L)_i$ and $\log(K/L)_i$ for human and physical capital respectively) are interacted with industry measures of factor intensity, where the latter are captured by the log factor usage per worker in the industry in question ($\log(H/L)^k$ and $\log(K/L)^k$).[18] I find that countries which are more skill abundant do indeed exhibit higher volumes of bilateral exports in more skill-intensive industries. Similarly, countries which have more physical capital per worker tend to export more in capital-intensive industries ($\beta_{f1} = 4.148$ and $\beta_{f2} = 0.056$, both significant at the 1% level). These echo the findings in Romalis (2004).

Institutional determinants: The rest of Table 2 finds broad support for the hypotheses on institutional sources of comparative advantage advanced recently. Column (3) examines the role of country financial development, captured by the ratio of private credit to GDP in the economy ($FINDEV$). This is interacted against a measure of industry dependence on external finance ($CAPDEP$), calculated following the methodology of Rajan and Zingales (1998). I obtain a positive and highly significant coefficient on this interaction term (β_{lm1}), confirming that financially-developed countries are more successful exporters in industries that depend more on external capital funding (Beck 2003, Manova 2006).

The next few columns turn to the role of the contracting and legal environment in facilitating production. Levchenko (2007) argued that industries that rely heavily on a few key inputs are more vulnerable to holdup problems from suppliers, and are hence more dependent on the legal system to enforce contracts. Column (4) examines this mechanism by interacting a Herfindahl index of input-use concentration in each industry (HI) against a measure of the strength of legal systems in each country ($LEGAL$). The positive and significant coefficient obtained (β_{lm2}) suggests that countries with stronger legal systems are in a better position to specialize in goods with a high input concentration.[19] Expanding on this incomplete contracting logic, Nunn (2007) developed a more refined measure of the extent to which holdup problems affect production, calculated as the share of inputs that are classified as relationship-specific (RS). The interaction between this industry measure with the country variable $LEGAL$ yields a positive and significant effect (β_{lm3}), providing further confirmation of the importance of contracting institutions

[18] I use this measure of industry factor intensities, instead of factor payment shares, as the former measure accounts for more of the variance in the trade data (in terms of regression R^2).

[19] The $LEGAL$ measure used here is for the year 1985, taken from the Economic Freedom around the World reports (Gwartney and Lawson 2004). 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, which likely reflects the high persistence in institutional conditions over time in most countries.

in facilitating specialization and exports in contract-dependent industries.[20] On a related note, Costinot (2007) proposed a different channel through which legal institutions can matter, by providing a contracting environment that facilitates the division of labor among work teams. It is argued in particular that frictions impeding the division of labor have more adverse productivity effects on industries where job tasks are more complex ($COMPL$). Column (5) demonstrates that this logic appears relevant for explaining trade patterns, as countries with stronger institutions do indeed export more in complex industries (β_{lm4}). As in Costinot (2007), I also find that countries with a higher skill endowment are better-placed to export in more complex industries (β_{lm5}), suggesting that skilled workers are indeed able to perform complex tasks more efficiently.

The final column in Table 2 considers the effect of labor market institutions. Consistent with Cuñat and Melitz (2007), I obtain a positive coefficient (β_{lm6} significant at the 1% level) indicating that countries with flexible labor institutions ($FLEX$) do export more in industries that experience greater sales volatility ($SVOL$); these are precisely the industries that rely most on being able to adjust employment to respond to changing market conditions.

Full model: All the above conclusions remain intact when I run these institutional determinants jointly in a single specification, as evidenced by Table 3, Column (1). In particular, all of interaction terms capturing Heckscher-Ohlin and institutional determinants are statistically significant, suggesting that the empirical literature has to date successfully identified largely independent channels through which country attributes influence the pattern of trade. To provide a gauge of the relative importance of these explanatory variables, Column (1a) reports the standardized beta coefficients based on the specification in Column (1).[21] Physical distance is not surprisingly the most influential distance variable (beta coefficient, $\beta_{d1} = -0.31$). That said, the Heckscher-Ohlin and the institutional terms collectively have a larger role in explaining trade flows than physical distance, with the sum of the beta coefficients for all eight interaction terms exceeding that for physical distance. Of note, the physical capital endowment and legal institutions appear to have the largest influence on trade flows (as suggested by the large beta coefficients: $\beta_{f2} = 0.491$, $\beta_{lm2} = 0.654$, $\beta_{lm3} = 0.494$). Column (1b) provides an alternative summary of the quantitative effects. For each interaction, this reports *ceteris paribus* how much larger the model predicts export volumes would be for the exporting country at the 75th versus 25th percentile, in the 75th versus 25th percentile industry. To illustrate, consider the β_{f1} coefficient: The interquartile gap in the human capital distribution in this sample of 83 countries is 0.415, while the corresponding gap in the industry skill-intensity distribution is 0.494. The Column (1) estimate of β_{f1} then implies

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

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

that trade flows would rise by a sizeable factor of $\exp(1.246 \times 0.415 \times 0.494) = 1.29$, namely a 29% increase, when moving from the 25th percentile country and industry to the 75th percentile. Repeating this for the other interactions confirms that the country attributes with the largest role as sources of comparative advantage are the physical capital endowment ($\log(H/L)^k \times \log(H/L)_i$, a 56% increase) and legal institutions ($HI \times LEGAL$, a 69% increase; $RS \times LEGAL$, 59%; $COMPL \times LEGAL$, 33%).

In sum, the above regressions confirm that the model provides a useful benchmark for explaining the intensive margin of trade, namely conditional on observing positive trade flows. However, a key concern with OLS is that two-thirds of the bilateral trade observations in the dataset are in fact zeros, and these are dropped from the regression sample. Columns (2) and (2a) clearly suggest that OLS does not provide the full picture: A probit regression based on equation (16) reveals that the same set of trade determinants also has a lot of explanatory power for the extensive margin of trade. (Column (2) reports marginal effects, while Column (2a) standardizes these to report the approximate increase in the probability of observing a positive trade flow when the covariate is raised by one standard deviation.) For example, physical distance has a significant effect in deterring trade completely, while several of the sources of comparative advantage are also significant determinants of whether trade is non-zero.

As a consequence, it would be inappropriate to use the OLS estimates for a welfare exercise, without first accounting for the potential coefficient bias from dropping the zeros. One view here is that the zeros arise due to measurement issues, either because small volumes are rounded down to zero, or because of the lack of reporting from less-developed countries (with zero assumed as a default). While this may explain some of the zeros, the fact that the probit regression does a good job of predicting the zeros suggests that there are more systematic economic forces at play inhibiting trade flows.[22] Moreover, removing the countries with the lowest per capita income levels (and hence presumably the poorest quality data) has little effect on the OLS results (available upon request).[23] In keeping with the Ricardian spirit of the model, the approach I pursue is instead to view the zero trade flows as arising from large productivity gaps between countries, which prevent low productivity countries from exporting to particular markets. This requires a minimal modification of the EK framework to generate zero trade predictions. The underlying parameters can then be re-estimated via simulated method of moments (SMM), by matching moments of trade flows simulated from the model with the corresponding moments from the actual data.[24]

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

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

[24] Other approaches for dealing with the zeros-bias require more extreme departures from the EK model. Santos-Silva and Tenreyro (2006) propose a Pseudo-Poisson maximum likelihood procedure for estimating gravity equations, but this entails assuming a non-standard distribution for the error terms in (16). Helpman et al. (2007) implement a two-stage estimation method, with a first-stage selection equation that determines the probability of observing positive trade. Their approach views the presence of fixed costs to exporting as a key obstacle giving rise to the zeros, whereas such fixed costs are absent in the baseline EK model here.

4 Estimation by Simulated Method of Moments (SMM)

4.1 Modifying the theory to generate zeros

In its present form, the model from Section 2 precludes any zeros, since equation (9) establishes that each country i has a strictly positive probability (albeit possibly tiny) of being the lowest-price supplier to any country n for a given industry-k variety. Therefore, 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 $F(\epsilon) = \exp(-\exp(-\epsilon))$. The bounded support now makes zero predicted trade possible: X_{ni}^k will equal zero if there exists another country, i', which is systematically more productive than i in this industry, to the extent that i cannot possibly become the lowest-price exporter even with the best productivity shock, $\bar{x}$. Formally, $X_{ni}^k = 0$ if and only if there exists a country $i' \neq i$ such that:

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

In contrast, under the previous specification of a Gumbel distribution with unbounded support, there would have been a positive probability of trade between every country pair in each industry, since even countries with a poor systematic component of productivity stood a chance of obtaining a large enough productivity shock to become the lowest-price exporter of at least one variety. Truncating the productivity distribution therefore represents a minimal extension of the model to generate zero trade flows, without having to introduce further features such as fixed cost barriers.[25]

4.2 SMM estimation procedure

With the bounded support assumption, we unfortunately lose closed-form expressions for trade flows. Nevertheless, given a set of parameter values, a complete set of trade flows can readily be simulated based on the underlying model. I therefore pursue estimation via a SMM procedure that searches for parameter values that deliver predicted trade flows which match key statistical moments of the actual data as closely as possible (Pakes and Pollard 1989). To implement this in practice, I take a discrete approximation of the measure of varieties; with a slight abuse of notation, I index the varieties in each industry by $j = 1, 2, \ldots, J$. Using the price equation (6), and substituting in the distance and factor endowment terms following the steps in (16), the log price of each variety in industry k is given by:

$$\ln(p_{ni}^k)^{(j)} = \frac{1}{\theta}\left(\theta\beta_d \cdot D_{ni} - \sum_{f=1}^{F} \theta\beta_f \cdot s_f^k \ln\frac{V_{if}}{V_{i0}} - \sum_{\{l,m\}} \theta\beta_{lm} \cdot L_{il} M_{km} + \tilde{I}_i + \tilde{I}_k - (\epsilon_i^k)^{(j)}\right) \tag{17}$$

[25]Note that introducing fixed costs alone would be 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.

Here, $(\epsilon_i^k)^{(j)}$ is a random draw from the truncated Gumbel distribution with support $[\underline{x}, \bar{x}]$, while $\tilde{I}_i$ captures all exporter fixed effects (such as δ_i) and $\tilde{I}_k$ groups together all industry-specific terms (such as μ_k and δ_k). 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 (17). This requires $N \times K \times J$ independent draws from the truncated Gumbel distribution for the productivity shocks, $(\epsilon_i^k)^{(j)}$. (Once drawn, these shocks are fixed throughout the rest of the estimation procedure.)

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

$$(P_n^k)^{1-\varepsilon} \approx \frac{1}{J}\sum_{j=1}^{J}((p_{n,i(j)}^k)^{(j)})^{1-\varepsilon} \tag{18}$$

3. Using the ideal price indices from (18), calculate the quantity demanded, $(Q_{n,i(j)}^k)^{(j)}$, for each variety in country n using the expression from (3). Here, the country GDP data for Y_n are taken from the World Development Indicators (WDI).

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

$$(X_{ni}^k)^{sim} = \frac{1}{J}\sum_{\{j:\ i(j)=i\}}(p_{n,i(j)}^k)^{(j)}(Q_{n,i(j)}^k)^{(j)} \tag{19}$$

In practice, however, the number of fixed effects to be estimated in (17) is large and could strain the reliability of conventional minimization algorithms. To reduce the number of parameters to be estimated, I arrange the countries into five groups in ascending order of their aggregate export volumes in 1990, and assign the same exporter fixed effect to each group of countries (which I denote by $\tilde{I}_{i1}, \ldots \tilde{I}_{i5}$, in increasing order of observed trade). Similarly, I sort the SIC industries into three groups according to the magnitude of total trade in each industry, and assign the same industry fixed effect $(\tilde{I}_{k1}, \ldots \tilde{I}_{k3})$ to each industry group. (See Appendix Table 3A for the list of groups; the cutoffs between groups were selected at natural breakpoints in the pecking order of trade volumes by exporter or industry.) Substituting the expression for quantity demanded (3) into (19), it is straightforward to verify that $(X_{ni}^k)^{sim}$ is invariant to any constant additive term that shifts all the log prices in (17) by the same amount. I therefore set $\tilde{I}_{i1} = 0$ and $\tilde{I}_{k1} = 0$ as a normalization, since one of the country fixed effects and one of the industry fixed effects cannot be identified.

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

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

This comprises the distance, Heckscher-Ohlin and institutional coefficients, as well as the group fixed effects. Before estimation, I set the remaining model parameters as follows. For the inverse spread parameter θ, EK (2002) present a range of values from 2.44 up to 12.86, depending on the estimation method. I set $\theta = 8.28$, a central value in EK (2002). The good fit of the baseline OLS regressions indicates that the closed-form expressions from the theory yield reasonable approximations for actual trade, which in turn suggests that the productivity shock distribution should include most of the relevant mass of the Gumbel distribution. The support of the truncated distribution is thus set to cover the central 99% of the mass of the (unbounded) Gumbel distribution, which implies $\underline{x} = -1.667$ and $\bar{x} = 5.296$.[26] For the elasticities of substitution, I take $\varepsilon = 3.8$ from Bernard et al. (2003), who estimate this from US firm-level data. I set $\phi = 2$ to satisfy the condition $\varepsilon > \phi > 1$. I follow EK in setting $\eta = 0.13$ for the consumption share of the manufacturing sector in total GDP. While it is possible to estimate some of these parameters, such as ε, ϕ and η, by introducing additional moments to match, I opt not to do so to focus on estimating the β coefficients in Θ, which are our primary interest.

The estimation problem is then to determine the parameter vector, $\hat{\Theta}$, that minimizes the distance metric between selected moments, $b(\cdot)$, of the simulated trade flows, $(X_{ni}^k)^{sim}$, and that of the actual data, X_{ni}^k, in the spirit of Hansen (1982):

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

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

1. The OLS regression coefficients from (16). This gives 14 moments, which are particularly informative for estimating $\beta_{d1}, \ldots, \beta_{d6}, \beta_{f1}, \beta_{f2}, \beta_{lm1}, \ldots, \beta_{lm6}$.

2. The share of total trade flows in each group for Exporter Groups 2-5 and SIC Industry Groups 2-3. These should be informative for estimating $\tilde{I}_{i2}, \tilde{I}_{i3}, \tilde{I}_{i4}, \tilde{I}_{i5}$, $\tilde{I}_{k2}$ and $\tilde{I}_{k3}$, the country and industry fixed effects respectively.

Since the problem is exactly identified (as many moments as there are parameters), I set the optimal weight matrix Ψ to the identity matrix. In practice, I first use a Newtonian search algorithm to first

[26] The log price expression in (17) indicates that θ will be difficult to identify if one were to attempt to estimate it, since $1/\theta$ enters multiplicatively with the coefficients and fixed effects parameters. In particular, when the productivity shock distribution is exactly Gumbel, it is well-known from the theory of discrete choice models that θ cannot be identified. Given the prior that the truncated productivity shock distribution covers most of the support of the Gumbel distribution, it is likely not feasible to estimate θ, hence the decision to calibrate it. Likewise, it would not be easy to estimate either $\underline{x}$ or $\bar{x}$, since the Gumbel distribution is very flat in its upper and lower tails.

determine the relevant parameter subspace in which the minima lies, before using the Nelder-Mead (1965) simplex search to obtain the final estimates.[27] I set $J = 500$; experimenting with larger J raises the computational burden, without changing the value of the objective function substantially.

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

4.3 The SMM estimates

The $\hat{\Theta}^{SMM}$ estimates are reported in Table 3, Column (3). Physical distance retains a negative and highly significant effect on trade, although the magnitude of this coefficient is smaller than found with OLS (β_{d1} is now -0.919). This is a feature found with other bias correction methods related to the omission of zeros in the gravity equation literature, such as Santos-Silva and Tenreyro (2006) and Helpman et al. (2007). One explanation that has been offered is that the elasticity of trade volumes with respect to distance declines over longer distances; the exclusion of the zeros thus biases the magnitude of the OLS distance coefficient upwards, since the zeros correspond to high-distance country pairs where the associated distance elasticity is low (Anderson and van Wincoop 2004, p.730).

Turning to the Heckscher-Ohlin and the institutional determinants, I find positive and significant effects of these interaction terms that echo the baseline OLS results. (The small standard errors here are a reflection of the good fit of the model moments to the data moments, as documented in Appendix Table 3B.) It does appear that accounting for the zeros tends to reduce the SMM coefficients slightly compared to the corresponding OLS coefficients, suggesting that these sources of comparative advantage have quantitatively more explanatory power for the intensive margin of trade (how much countries trade) than the extensive margin (whether countries trade). This is consistent for example with Manova (2006) who finds using the Helpman et al. (2007) bias-correction method that about two-thirds of the effect of financial development ($CAPDEP \times FINDEV$) operates through the intensive margin and about one-thirds through selection into exporting.

How sensible are these estimates for explaining the actual trade data? I offer evidence here that the model delivers a reasonable fit on several dimensions, namely the country income levels and accompanying

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

pattern of trade flows that it predicts in the global trade equilibrium.

Implied country GDP levels: I first close the model in order to solve for the equilibrium country income levels, by appealing to a trade balance condition. With the discrete approximation for the measure of varieties, total manufacturing exports from country i, EXP_i, are given by:

$$\begin{aligned} EXP_i &\approx \frac{1}{J}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{\{j:\ i(j)=i\}}(p^k_{n,i(j)})^{(j)}(Q^k_{n,i(j)})^{(j)} \\ &= \frac{1}{J}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{\{j:\ i(j)=i\}}\frac{\eta Y_n\ (P^k_n)^{\varepsilon-\phi}}{\sum_{\kappa\geq 1}(P^\kappa_n)^{1-\phi}}\ ((p^k_{n,i(j)})^{(j)})^{1-\varepsilon} \end{aligned} \tag{20}$$

where the sum is taken over all varieties (across all industries) and over all export destinations for which country i is the lowest-price exporter. Substituting in for $(Q^k_{n,i(j)})^{(j)}$ from (3) yields (20), which expresses total exports from country i as a linear combination of the country income levels, Y_n. This property follows from the Cobb-Douglas specification for utility, since the trade quantities are then linear functions of the importing country's GDP. I calculate the GDP coefficients, namely the $\frac{\eta\ (P^k_n)^{\varepsilon-\phi}}{\sum_{\kappa\geq 1}(P^\kappa_n)^{1-\phi}}\ ((p^k_{n,i(j)})^{(j)})^{1-\varepsilon}$ terms, at the the SMM estimate $\hat{\Theta}^{SMM}$ via simulation (with $J = 500$). On the other hand, i's total imports, IMP_i, are equal to ηY_i. The balanced trade condition, $EXP_i = IMP_i$ for each country, therefore gives a homogenous system of N linear equations in the N income levels Y_n. Setting income for the US to 1, and inverting this system yields the implied equilibrium country GDPs.[28] Note that the 83 countries in the sample had a combined output equal to 92.7% of world nominal GDP in 1990, so the country income levels computed should be reasonable approximations for actual GDP.

Figure 1 confirms that the implied values for country GDP based on $\hat{\Theta}^{SMM}$ successfully capture the relative rank ordering of observed nominal income levels in 1990 (taken from the WDI). There is a tendency for the model to slightly under-predict GDP, particularly for low-income countries which tend to cluster under the 45-degree line. Nevertheless, the Spearman rank correlation between the two variables is very high (0.54, significant at the 1% level), so that the model reproduces the rank order of country income levels. (Likewise, the Pearson linear correlation between the two log income series is high, equal to 0.58 and significant at the 1% level.)

Bilateral trade patterns: Figure 2 compares the actual data with a set of simulated trade flows based on $\hat{\Theta}^{SMM}$ and the implied Y_n's calculated above. For comparability, I scale up the predicted Y_n's, so that the value of US GDP is equal to that in the WDI. Overall, the model provides a reasonable fit to the data, with the Pearson linear correlation between the two log trade flow series being 0.40 (significant at the 1% level). The model also matches the zeros quite well (not shown due to the log scale): There

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

are 91,086 zeros in the actual data, of which 84,546 are shared with the simulated trade flows. Two brief caveats are nevertheless in order. First, the model tends to under-predict trade volumes, which stems largely from the propensity of the model to under-predict country income levels. Second, the generated trade flows display a smaller amount of dispersion than the actual data, as evidenced by the smaller coefficient of variation for the former (0.10 versus 0.19, calculated for the subset of non-zero trade flows). The plots of predicted versus actual trade for each industry are similar in nature to Figure 2 (available on request).

5 Welfare Counterfactuals

The structural approach adopted now allows us to explore various interesting counterfactual exercises on the relative importance of distance and the various sources of comparative advantage from a welfare perspective. I use the SMM estimates $\hat{\Theta}^{SMM}$ to explore these counterfactual implications of the model in this section. For a start, the framework allows us to examine the welfare effects of reducing distance barriers along the lines studied in EK (2002). Furthermore, since the model ties comparative advantage to observable country and industry characteristics, this facilitates quantifying the importance of different sources of comparative advantage.

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):

$$W_n = \frac{(1-\eta)^{1-\eta}\eta^{\eta}\; Y_n}{(p_n^0)^{1-\eta}\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{\frac{\eta}{1-\phi}}} \tag{21}$$

Without the term, $(1-\eta)^{1-\eta}\eta^{\eta}$, this is precisely equal to country n's real GDP. Note that the price of the domestic non-tradable, p_n^0, has been introduced explicitly in the denominator: When solving for the implied GDP 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 is normalized to 1. This means that domestic factor prices, and hence the price of domestic non-tradables will be endogenous in general equilibrium, and we need to account for this in the welfare calculations. Note also that this welfare measure focuses on the impact on a representative consumer, putting aside distributional consequences within countries.

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

$$\frac{\Delta W_n}{W_n} = \frac{\Delta Y_n}{Y_n} \; - \; (1-\eta)\frac{\Delta(p_n^0)}{p_n^0} \; - \; \eta\frac{\Delta\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}}{\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}} \tag{22}$$

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

For each counterfactual, I evaluate (22) by simulating a full set of country trade flows both before and after introducing the shock, to compute Y_n and $\left(\sum_{k\geq 1}(P_n^k)^{1-\phi}\right)^{1/(1-\phi)}$, as well as their respective percentage changes. As for the change in the price of the domestic non-tradable, this is equal to the weighted sum of domestic factor price changes, where the weights are the factor share intensities in this sector: $\frac{\Delta(p_n^0)}{p_n^0} = \sum_{f=0}^{F} s_f^0 \frac{\Delta(w_{nf})}{w_{nf}}$. I approximate the percentage change in w_{nf} as the change in total factor payments accruing to factor f in country n, net of any change in the endowment of that factor.[29] To fully operationalize this, I set the factor shares in the outside sector as follows: $s_h^0 = 0.175$, $s_l^0 = 0.325$, $s_k^0 = 0.5$, based on the average factor payment shares over the 1980s in US agriculture (the canonical non-manufacturing sector) from Mundlak (2005).[30]

5.1 Reducing distance barriers

I first consider a transition towards a hypothetical zero-gravity world. Although physical distance and transport costs can never be completely eliminated in practice, this nevertheless provides a useful gauge of how much distance and geography hold back country welfare. To this end, I consider a counterfactual where all distance variables are set to minimize their adverse impact on prices, namely where log physical distance is 0 and all the five dummy variables (common language, common border, colony, GATT, RTA) are equal to 1. Note that this is strictly speaking not a pure zero-gravity experiment: In the flexible formulation of the model, the distance markup d_{ni}^k in (14) cannot be set exactly to 1, as we cannot distinguish δ_k empirically from other industry-specific terms that affect trade flows (in particular, the μ_k's in the systematic component of productivity). What is being done here is instead to set the observable

[29]More explicitly, I compute the percentage change in factor prices as:

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

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

[30]I set $s_k = 0.5$ based on the total factor payment share in value-added to physical capital and land (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.

distance measures such that the distance markup is as small as possible.

Table 4 reports sizeable overall gains in this low-gravity scenario, with an average welfare increase of 15.7% in the 83-country sample. This is comparable to the range of country welfare increases that EK (2002) reported from a zero-gravity exercise with their OECD sample (16.1%-24.1%). Decomposing this welfare change using (22), the fall in the price of tradables contributes a fair amount of this welfare gain (10.1%), but the bulk of the increase is driven by the increase in country income levels (19.1%) as the removal of distance barriers opens more trading opportunities. This is partially offset by the rise in the price of non-tradables: The increase in foreign demand for each countries' products raises demand for factors of production domestically. Factor prices rise as a result, bringing up the price of domestic non-tradables. (The lower half of Table 4 reports the less extreme scenario where only physical distance is set to 0, or equivalently where $\beta_{d1} = 0$. As expected, this implies more moderate welfare gains.)

The move towards a zero-gravity world exhibits some interesting features. Figure 3 graphs this transition as barriers are successively reduced from the status quo ($x = 0$), to the case where all observable barriers are set to minimize the distance markup ($x = 1$), and beyond. The solid line indicates an initial rise in the concentration of production by country source within each industry, as illustrated by a Herfindahl index of producer shares in each industry, $\sum_{i=1}^{N}\left(\sum_k X_{ni}^k / \sum_i \sum_k X_{ni}^k\right)^2$, averaged across the 20 industries (plots for each of the 20 industries are similar). Evidently, the initial removal of distance barriers favors existing producer countries, which can be described as an expansion of exports on the intensive margin. This increase in concentration is accompanied by a fall in welfare in the global equilibrium (illustrated for Zimbabwe and Spain, at the 25th and 75th per capita income percentiles respectively; trends are similar for other countries). Only as distance is reduced more substantially do exports expand also on the extensive margin as new countries start exporting, so that this concentration index eventually falls below the status quo level and country welfare also turns around to register positive gains.[31] Throughout this transition, the prices of tradables are declining uniformly, so what is driving this non-monotonicity in welfare is the response of country incomes as the trade equilibrium shifts.

5.2 Sources of comparative advantage

To assess how much the various comparative advantage forces matter for country welfare, I neutralize the terms in the empirical model that correspond to each source of comparative advantage country-by-country. These results are summarized in Table 5, the main column of interest here being the "Mean" column, which reports the welfare loss for the country for which the comparative advantage force is shut down, averaged across the 83 country-by-country scenarios.

[31] The mean of the 20 industry Herfindahl indices of producer concentration drops from a status quo value of 0.16 to 0.12 (at $x = 1$). A simple one-sided t-test rejects the null hypothesis of no change in this producer concentration Herfindahl across industries at the 1% level.

Consider first the role of Ricardian forces. Removing the stochastic component of productivity by setting the $(\epsilon_i^k)^{(j)}$ draws to 0 for each country in turn leads to a mean welfare loss of about -5.8%. Since the stochastic shocks inherit a thick right-tail from the Gumbel distribution, removing these shocks tends to worsen country productivity on average, leading to a drop in welfare. This welfare loss is much smaller than the -36.6% decline when the systematic component of productivity is shut down, which is done by setting the $-\sum_{\{l,m\}} \theta\beta_{lm} \cdot L_{il}M_{km} + \tilde{I}_i + \tilde{I}_k$ term in (17) equal to the maximum value in the sample, so that the systematic component of productivity does not reduce log prices differentially across countries and industries.[32] The decomposition of these changes once again shows that most of the welfare shifts are being driven by changes in country GDP in the new trade equilibrium. Moreover, once the systematic component has been neutralized, removing the stochastic shocks has little further impact on country welfare (the differences only show up in the third decimal place). This is reassuring, as it indicates that the key productivity term is the systematic component which depends on observable fundamentals, namely the actual characteristics of the countries and industries being studied.

Turning to the Heckscher-Ohlin forces, the second panel in Table 5 tabulates the effect of neutralizing the interaction terms that capture the role of factor endowments. (This is done for each f by setting $-\theta\beta_f \cdot s_f^k \ln \frac{V_{if}}{V_{i0}}$ in (17) to the maximum value in the sample.) The model implies large effects here, as country welfare declines on average by -20.8% when human capital is not allowed to be a force in reducing prices across countries or industries; the corresponding loss in the case of physical capital is -28.7%. The last column of Table 5 correlates the percentage welfare change experienced in a country against its initial factor endowment level. Not surprisingly, more skill-abundant countries suffer a larger welfare loss when this source of comparative advantage is shut down (see Figure 4, Panel A). There is a similar negative correlation between a country's physical capital endowment and the subsequent welfare losses (Panel B). When both factor endowment motives for trade are shut down, welfare declines by a substantial -34.5%.

I perform a similar exercise to assess the quantitative impact of the institutional determinants that have received attention in the recent literature. (For each $\{l, m\}$, this is done by setting $-\theta\beta_{lm} \cdot L_{il}M_{km}$ in (17) to the maximum value in the sample; this is the sense in which the institutional determinants are a subset of the Ricardian forces, specifically the systematic component of productivity.) The calculations point to fairly sizeable roles for financial development and flexible labor market institutions, with mean welfare losses of -19.2% and -9.4% respectively when these channels of comparative advantage are shut down. The most substantial welfare effects though center on legal institutions, particularly on their role in establishing a contracting environment that facilitates production in industries vulnerable to holdup problems: Shutting down specialization on the basis of industry input-concentration (HI) or the share of

[32] For simplicity, this assumes that all of the fixed effects, $\tilde{I}_i + \tilde{I}_k$, can be attributed to the systematic component of productivity.

relationship-specific inputs (RS) implies average welfare losses of -33.5% and -31.8% respectively. The role of job complexity ($COMPL$) here is also non-trivial, with a mean welfare loss of -15.7% when this mechanism is switched off. Once again, the correlations in the final column indicate that the stronger a country's institutions, the more each institutional determinant will matter in terms of the welfare loss from shutting down that source of comparative advantage. (This is illustrated in Figure 4, Panel C for $FINDEV$ and Panel D for $LEGAL$.)

How do these effects contrast with the impact of distance? To provide some basis for comparison, the last part of Table 5 considers the analogous exercise of raising distance barriers (the $\theta\beta_d \cdot D_{ni}$ terms in (17)) country-by-country to the maximum observed in the dataset. The model yields average welfare losses here that are of a similar order of magnitude (-37.8%).

In short, the model points to the conclusion that Ricardian forces, Heckscher-Ohlin forces, and distance variables all share a similar degree of importance in terms of their quantitative implications for country welfare. (That said, it should be stressed that the numbers in Table 5 should not be interpreted as a strict decomposition, as the counterfactuals for each set of forces have been run separately.)

5.3 Country policy experiments

The model also allows us to examine more closely the effects of raising the characteristics of specific countries, to evaluate the consequences of such country policy experiments. I illustrate this using the example of a large developing country, Indonesia, which lies between the 25th and 33rd percentiles on each of the country characteristics studied. In particular, I consider what happens when raising Indonesia to the world frontier level (approximated by the maximum level in the sample) along each of these country dimensions, to provide some benchmark numbers for the potential gain from a broad increase in factor endowments or an improvement in the quality of institutions.

Table 6 reports the results from these policy experiments. I first raise each country characteristic for Indonesia as it appears in each interaction term. The welfare gains from expanding Indonesia's factor endowments are fairly substantial, equal to 9.8% and 18.5% respectively from raising the per worker human capital and physical capital ratios to the world frontier level. A sizeable part of this increase is actually driven by an endowment effect, which decreases domestic factor prices and makes the non-tradable good ($k = 0$) relatively cheaper to consumers. This is particularly so in the scenario where Indonesia's human capital is raised; the change in non-tradables price would also have been even more negative in the physical capital scenario had this endowment effect not been present. For the institutional determinants, the figures suggest especially large benefits from raising legal institutions to first-world standards, particularly when all three interactions involving $LEGAL$ are taken into account (under "Joint effects"). While Indonesia clearly gains under these policy experiments, the repercussions on the rest of the world (ROW) tend to be tiny. Any adverse beggar-thy-neighbor effects from policy shocks in Indonesia – from the diversion of export opportunities away from other countries – thus appear to be small.

Focusing on a policy shock to one country alone also allows us to explore in finer detail what happens to industry composition in that country. Table 7 confirms that there is indeed a substantial amount of reallocation taking place between industries in Indonesia in response to each policy change. These production patterns indeed shift towards industries that are more dependent on the country attribute that has been enhanced. For example, when Indonesia's skill endowment is raised, more skill-intensive and complex industries tend to expand at the expense of less skill-intensive and complex ones, as evidenced by the positive correlation reported between the change in industry relative size and $\log(H/L)^k$ and $COMPL$ respectively (see also, Figure 5, Panel A). Figure 5 provides several further illustrations of these systematic shifts in industry structure: Physical capital accumulation favors capital-intensive industries (Panel B), an expansion of private credit favors industries that are dependent on external finance (Panel C), and an improvement in the legal environment shifts resources towards industries that are more vulnerable to holdup problems (Panel D).

It is worth highlighting some caveats about the precise interpretation of these welfare counterfactuals. When raising a country characteristic exogenously, I shock the relevant interaction term involving that characteristic, while holding the exporter fixed effect constant. The welfare changes calculated are therefore strictly due to the induced shift in the pattern of industrial specialization. Since this holds constant any direct level effects from the expansion in countries' production capacities, it likely understates the magnitude of welfare changes. These exercises also focus solely on the gains from these policy moves. There are certainly costs to implementing these policies, such as the foregone current consumption from physical capital accumulation or the structural adjustment costs as factors move across industries, but these lie outside the scope of the model.

6 Conclusion

This paper develops a methodology for estimating and 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 pertinent counterfactual scenarios. To understand patterns of specialization, I present an extension of the multi-country 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.

I pursue two estimation approaches: (i) an OLS baseline in Section 3, and (ii) a simulated method of moments (SMM) procedure in Section 4 that takes into account the prevalence of zero trade observations. Both sets of estimates confirm the relevance of traditional gravity measures, particularly physical distance, for explaining bilateral trade flows. I also jointly corroborate the role of factor endowments and country institutions – including financial development, the contracting environment, and labor market regimes – as sources of comparative advantage. The SMM estimates in turn imply welfare effects of a reasonable magnitude in various counterfactual scenarios. In particular, these suggest that Ricardian forces, Heckscher-Ohlin forces, and distance barriers are approximately equally important in terms of their influence on country welfare. Among the institutional determinants of trade, the model also points to legal institutions as being particularly influential, with the largest gains stemming from improvements to the contracting environment that alleviate holdup problems in production.

A key strength of this modeling framework is that it is in fact very flexible, allowing the researcher to incorporate a full set of country-industry interaction terms identified in the recent literature as significant sources of comparative advantage. While I have attempted to be comprehensive here, the model is certainly more general in that it can in principle accommodate additional relevant interaction terms or more flexible functional forms (such as non-linear effects), subject to the caveat that this will raise the computational cost for the SMM procedure. There are of course some limitations to bear in mind when interpreting the counterfactual results. These exercises treat the policy changes as exogenous shocks for simplicity, putting aside such dynamic issues as the process of adjustment to the new trade equilibrium, or potential policy responses by other countries. That said, the paper takes useful steps 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.

7 References

Acemoglu, Daron, Pol Antràs, and Elhanan Helpman, (2007), "Contracts and Technology Adoption," *American Economic Review* 97: 916-943.

Alvarez, Fernando, and Robert E. Lucas, Jr., (2007), "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade," *Journal of Monetary Economics* 54: 1726-1768.

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.

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/

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, (2007), "Financial Development and International Trade," University of Illinois at Urbana-Champaign mimeo.

Berkowitz, Daniel, Johannes Moenius, and Katharina Pistor, (2006), "Trade, Law, and Product Complexity," *Review of Economics and Statistics* 88: 363-373.

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

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, and Ivana Komunjer, (2007), "What Goods do Countries Trade? New Ricardian Predictions," University of California at San Diego mimeo.

Costinot, Arnaud, (2007), "On the Origins of Comparative Advantage," University of California at San Diego mimeo.

Cuñat, Alejandro, and Marc J. Melitz, (2007), "Volatility, Labor Market Flexibility, and the Pattern of Comparative Advantage," *NBER Working Paper* 13062.

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.

Deardorff, Alan, (1984), "Testing Trade Theories and Predicting Trade Flows," in *Handbook of International Economics, vol. 1*, eds. Ronald Jones and Peter Kenen, Amsterdam: Elsevier BV.

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/

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.

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

Harrigan, James, (1997), "Technology, Factor Supplies, and International Specialization: Estimating the Neoclassical Model," *American Economic Review* 87: 475-494.

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.

Hur, Jung, Manoj Raj, and Yohanes Riyanto, (2006), "The Impact of Financial Development and Asset Tangibility on Export," *World Development* 34: 1728-1741.

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

Kletzer, Kenneth, and Pranab Bardhan, (1987), "Credit Markets and Patterns of International Trade," *Journal of Development Economics* 27: 57-70.

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, (2007), "Institutional Quality and International Trade," *Review of Economic Studies* 74: 791-819.

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

Manova, Kalina, (2008), "Credit Constraints, Equity Market Liberalizations and International Trade," *Journal of International Economics* 76: 33-47.

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

Morrow, Peter M., (2008), "East is East and West is West: A Ricardian-Heckscher-Ohlin Model of Comparative Advantage," University of Toronto mimeo.

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," *Quarterly Journal of Economics* 122: 569-600.

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.

Ranjan, Priya, and Jae Young Lee, (2007), "Contract Enforcement and International Trade," *Economics and Politics* 19: 191-218.

Ramondo, Natalia, (2008), "Size, Geography, and Multinational Production," University of Texas at Austin mimeo.

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, (2007), "Capital, technology, and specialization in the neoclassical model," Suffolk University mimeo.

Svaleryd, Helena, and Jonas Vlachos, (2005), "Financial Markets, the Pattern of Industrial Specialization and Comparative Advantage," *European Economic Review* 49: 113-144.

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. Converted from SITC Rev 2 into US 1987 SIC format using a concordance based on 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. This is used to derive concordance weights to map SITC Rev 2 categories into SIC-87 format, following the procedure in Cuñat and Melitz (2007).

Two complications arise. First, classification for the SIC-87 categories is based on observed finished products, but the distinction between SIC industries is often defined according to the production process. To cite an example from 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 FRS 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. I break up the merged trade flows for the affected categories in proportion to the value of US total shipments in 1990 reported in the NBER-CES Manufacturing Industry database (Bartelsman et al. 2000).[33] Then, 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 codes with suffixes 'A' and 'X', for trade flows not observed at a more disaggregate level. I assign the trade in these 'A' and 'X' categories to the truncated (more aggregate) SITC code. In other words, I treat 111A and 111X as coming from the 3-digit SITC category 111, and then use FRS to construct weights to map SITC 111 into SIC categories.

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

Distance: Physical distance is measured by the great circle formula distance between countries' capital cities, taken from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII).[34] A country's log distance to itself is set to zero, so that physical distance does not impose an iceberg cost for internal trade. The following binary variables are also from the CEPII: (i) "Common Language", equal to 1 if at least 9% of each country's population speaks a shared language; and (ii) "Colony", equal to 1 if one of the countries had ever colonized the other. The "Border" dummy (equal to 1 if the countries share a land border) is coded using the CIA World Factbook.[35] The following two measures are based on Rose (2004), augmented by direct reference to the WTO website[36] to cover all country pairs in my sample for the year 1990: (i) "RTA", equal to 1 if the countries are joint signatories in any of the regional trade agreements reported to the WTO; and (ii) "GATT", equal to 1 if both countries are GATT/WTO members. A value of 1 is assigned for all five dummies for a country's distance from itself.

[33] 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.

[34] http://www.cepii.fr/anglaisgraph/bdd/distances.htm

[35] http://www.odci.gov/cia/publications/factbook/index.html

[36] http://www.wto.org/english/thewto_e/gattmem_e.htm

B. INDUSTRY CHARACTERISTICS

Factor intensities: From the NBER-CES database. Variables are calculated for each 2-digit SIC-87 industry. Skill intensity is the log of the ratio of non-production workers to total employment. Physical intensity is the log of the ratio of real capital stock to total employment. Both ratios are averages over the period 1980-89.

The welfare counterfactuals require information on industry factor payment shares. These are obtained from the same NBER-CES database, using averages over 1980-89. The share of payments to skilled labor (s_h) and unskilled labor (s_l) are calculated by the ratios of non-production worker payroll and production worker payroll to total industry value-added respectively. The factor share of physical capital (s_k) is the ratio of residual payments (total value-added minus total payroll) to total value-added.

External capital dependence ($CAPDEP$): Constructed following the methodology in Rajan and Zingales (1998). Data from Compustat is used, which covers all publicly-traded firms in North America. A given firm's dependence on external capital is the fraction of total capital expenditures over the period 1980-89 not financed by internal cash flow. The median value across firms in each SIC-87 2-digit category is used as the industry measure of $CAPDEP$. (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 (2007). Equal to the Herfindahl index of intermediate input use, based on the 1987 US Input-Output (IO) Use Table. The IO-87 6-digit level categories map cleanly into the SIC-87 4-digit categories based on the correspondence table provided by the Bureau of Economic Analysis (BEA).[37] 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 total shipments reported in 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). RS is 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. Moving from (i) to (iii), one has 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 with the procedure described for the HI variable. The measure is aggregated 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. (The measure in Nunn (2007) is constructed for IO-87 industries instead.)

Job complexity ($COMPL$): Based on Costinot (2007). The 1985 and 1993 instalments of the US Panel Survey of Income Dynamics (PSID) contain 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 (2007) calculates the average response for SIC-1972

[37]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.

3-digit industries, normalized to a maximum value of 1. I assign these values to the corresponding 4-digit sub-categories. 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). These are then transformed from SIC-1972 to SIC-1987 categories using the weights in the correspondence table developed by Bartelsman, Becker and Gray.[38] The value of $COMPL$ for each SIC-1987 2-digit industry is then taken as the median over all its 4-digit sub-categories. There are two industry groups for which this imputation procedure may seem too liberal, namely SIC 21 and 29, for which direct information on complexity is not available in the PSID for any of the 3-digit sub-categories. The OLS results are similar if I omit these two industry groups.

Sales Volatility ($SVOL$): From Cuñat and Melitz (2007). Equal to the employment-weighted standard deviation of sales growth for firms in the 1980-2004 Compustat sample. Only firms with at least 5 years of data are used. Observations where the absolute sales growth rate exceeds 300% are omitted as outliers.

C. COUNTRY CHARACTERISTICS

Factor endowments: Physical capital per worker ($\log(K/L)_i$) and human capital per worker ($\log(H/L)_i$) are from Hall and Jones (1999), for the year 1988.

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

Legal System ($LEGAL$): From Gwartney and Lawson (2004). Index measure of "Legal System and Property Rights" for 1985, rescaled between 0 and 1, which is 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 sub-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 an international survey of business executives.

Employment Flexibility ($FLEX$): From the World Bank's *Doing Business* database. Index of "Rigidity of Employment", averaged over 2003-06, rescaled to be increasing in labor market flexibility and to lie between 0 and 1. Calculated as the 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 on the methodology in Botero et al. (2004).

GDP: Both GDP and GDP per capita are taken from the World Development Indicators (WDI), in current US dollars.

Population: From the WDI.

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

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 (83)

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); Venezuela (VEN); South Africa (ZAF); Zaire (ZAR); Zambia (ZMB); Zimbabwe (ZWE)

Table 2
Baseline OLS Regression Model of Bilateral Industry Trade Flows
(Gravity equation estimation, with fixed effects)

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

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

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

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

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

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

Notes: ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively. For the OLS and probit regressions, exporter and importer-industry fixed effects are included, with robust standard errors clustered by exporter-importer pair. Column (1a) reports standardized beta coefficients from the Column (1) specification, while Column (1b) reports the factor increase in trade in the 75th compared to the 25th percentile exporter and industry. Column (2) performs a probit regression on the probability of observing positive trade, with Column (2a) standardizing these to report the probability change from a one standard deviation increase in the right-hand side variable. Column (3) presents the SMM coefficients, where the exporter and industry fixed effects have been grouped (as discussed in the text); these group fixed effects are not reported.

Table 4
Counterfactuals I: Reducing distance barriers

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

Notes: The mean percentage welfare change across countries is reported in the "Mean" column. The decomposition breaks this down into the contributions from changes in country GDP, changes in the differentiated goods price index ($k \geq 1$), and changes in the price of domestic non-tradables ($k = 0$). These are also reported for the countries at the 5th, 25th, 50th, 75th, and 95th percentiles of GDP per capita (in US dollars) among the 83 countries. All percentages are calculated as $100 \ln(x'/x)$, where x' and x are the final and initial values respectively.

Table 5
Counterfactuals II: Sources of Comparative Advantage

Comparative advantage force(s) switched off	% Welfare Change Min.	Max.	Std. Dev.	Mean	Decomposition Due to change in: Country GDP	Prices ($k \geq 1$)	Prices ($k = 0$)	Correlation with cty. char.
Ricardian forces:								
Stochastic component	−55.7	19.1	9.3	**−5.8**	−44.8	−1.0	40.1	
Systematic component	−145.0	0.9	40.5	**−36.6**	−321.9	−3.9	289.2	
Both stochastic and systematic	−145.0	0.9	40.5	**−36.6**	−321.9	−3.9	289.2	
Heckscher-Ohlin forces:								
$\log(H/L)^k \times \log(H/L)_i$	−69.5	−0.1	19.5	**−20.8**	−169.1	−1.8	150.1	−0.46***
$\log(K/L)^k \times \log(K/L)_i$	−117.4	−0.1	30.9	**−28.7**	−229.6	−2.4	203.2	−0.63***
All Heckscher-Ohlin forces	−135.9	−0.2	37.0	**−34.5**	−288.6	−3.3	257.4	
Institutional determinants:								
$CAPDEP \times FINDEV$	−75.8	−0.1	19.6	**−19.2**	−152.3	−1.6	134.6	−0.58***
$HI \times LEGAL$	−140.3	−0.1	37.4	**−33.5**	−269.9	−3.0	239.4	−0.87***
$RS \times LEGAL$	−134.6	−0.1	35.4	**−31.8**	−253.5	−2.8	224.4	−0.87***
$COMPL \times LEGAL$	−67.2	−0.0	17.5	**−15.7**	−120.7	−1.0	105.9	−0.66***
$COMPL \times \log(H/L)_i$	−19.0	−0.0	4.6	**−3.9**	−29.4	−0.2	25.8	−0.84***
$SVOL \times FLEX$	−50.0	−0.1	10.7	**−9.4**	−73.0	−0.7	64.3	−0.40***
All institutional determinants	−145.0	0.9	40.5	**−36.6**	−321.4	−3.9	288.6	
Distance: (for comparison)								
Physical distance only	−126.1	−1.2	37.4	**−37.1**	−301.4	−6.0	270.3	
All distance barriers	−132.4	−1.9	38.0	**−37.8**	−305.5	−6.5	274.2	

Notes: For each row, the comparative advantage force is neutralized country-by-country. The mean percentage welfare change for the country for which the comparative advantage force is shut down is reported in the "Mean" column. The decomposition breaks this down into the contributions from changes in country GDP, changes in the differentiated goods price index ($k \geq 1$), and changes in the price of domestic non-tradables ($k = 0$). The final column reports the cross-country Pearson correlation between the percent welfare change and the initial level of the corresponding country characteristic; *** denotes significance at the 1% level respectively. All percentages are calculated as $100 \ln(x'/x)$, where x' and x are the final and initial values respectively.

Table 6
Counterfactuals IIIA: Country Policy Experiments for IDN

	IDN rank (out of 83)	% Welfare Change: IDN Total	Due to change in: Country GDP	Prices ($k \geq 1$)	Prices ($k = 0$)	ROW Mean
Raising:						
$\log(H/L)^k \times \max(\log(H/L)_i)$	30	9.8	−46.3	−0.8	56.9	0.002
$\log(K/L)^k \times \max(\log(K/L)_i)$	33	18.5	69.1	1.1	−51.7	−0.005
$CAPDEP \times \max(FINDEV)$	23	1.0	11.1	0.0	−10.2	0.001
$HI \times \max(LEGAL)$	36	36.8	291.7	3.8	−258.7	−0.011
$RS \times \max(LEGAL)$	36	26.6	216.8	3.0	−193.2	−0.011
$COMPL \times \max(LEGAL)$	36	4.8	47.3	0.6	−43.1	−0.001
$COMPL \times \max(\log(H/L)_i)$	30	1.8	13.9	0.2	0.7	−0.001
$SVOL \times \max(FLEX)$	34	4.1	40.2	0.6	−36.7	−0.003
Joint Effects:						
$\max(\log(H/L)_i)$	30	10.2	−39.5	−0.7	50.4	0.003
$\max(LEGAL)$	36	106.8	795.0	8.6	−696.8	−0.023

Notes: The decomposition breaks down the percentage 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 ($k = 0$). The final column reports the mean welfare change in the 82 other countries in the sample (Rest of the World: ROW). All percentages are calculated as $100 \ln(x'/x)$, where x' and x are the final and initial values respectively.

Table 7
Counterfactuals IIIB: Impact on IDN's industry structure

SIC	Industry description	Raising $\log(H/L)_i$	Raising $\log(K/L)_i$	Raising $FINDEV$	Raising $LEGAL$	Raising $FLEX$
		% change in IDN industry share				
20	Food products	12.7	−18.3	−14.3	−45.6	−14.0
21	Tobacco products	20.0	−20.8	−48.8	−59.4	−16.4
22	Textile mills products	−19.3	−19.8	−6.0	−47.4	−11.1
23	Apparel	−10.2	−30.2	−23.3	−47.0	−16.2
24	Wood products	−2.6	−24.2	−6.8	−58.3	−13.8
25	Furniture	−19.4	−15.5	−9.3	−23.3	−6.3
26	Paper products	−7.9	−3.1	−12.4	−23.4	−7.0
27	Printing	26.7	3.5	−25.9	16.8	0.8
28	Chemical products	21.8	26.5	31.1	32.4	6.4
29	Petroleum refining	16.1	−2.3	−12.2	−41.4	−7.1
30	Rubber and misc plastics	−15.2	−6.5	−9.0	−17.8	−5.3
31	Leather products	−3.6	−29.5	−41.1	−50.7	−18.6
32	Stone, clay, glass, concrete	−9.6	−0.4	−5.4	−10.9	−7.5
33	Primary metal industries	−18.5	4.6	−11.5	−9.0	3.0
34	Fabricated metal products	−8.2	−5.9	−16.8	−14.8	−0.6
35	Machinery and computers	7.1	20.7	29.1	43.2	18.1
36	Electronic products	−13.1	13.5	30.4	41.1	19.5
37	Transportation equipment	3.4	9.8	−10.7	17.8	1.4
38	Instruments	20.4	30.1	73.8	113.7	27.1
39	Misc manufacturing	−10.6	4.1	6.1	13.6	18.5
	Correlation with:	$\log(H/L)^k$	$\log(K/L)^k$	$CAPDEP$	HI	$SVOL$
		0.69***	0.34	0.93***	0.74***	0.43*
		$COMPL$			RS	
		0.46**			0.47***	
					$COMPL$	
					0.88***	

Notes: Policy experiments considered involve raising each of IDN's country characteristics to the maximum level in the sample. The $\log(H/L)_i$ column raises the human capital endowment for both the interactions involving $\log(H/L)^k$ and $COMPL$. The $LEGAL$ column raises the legal institutions index for all three of the interactions involving HI, RS and $COMPL$. The percentage change of each industry's output as a share of total IDN production is reported. The bottom of the table reports the Pearson linear correlations between the percentage changes and the corresponding industry characteristic(s); ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. All percentages are calculated as $100\ln(x'/x)$, where x' and x are the final and initial values respectively.

Appendix Table 1A
Summary of Country Characteristics

	Min.	10th	25th	Med.	75th	90th	Max.	Std. Dev.
$\log(H/L)_i$	0.072	0.257	0.392	0.592	0.807	1.039	1.215	0.290
$\log(K/L)_i$	5.763	7.050	8.332	9.723	10.828	11.318	11.589	1.584
Financial Devt. ($FINDEV$)	0.007	0.100	0.157	0.279	0.515	0.790	1.378	0.296
Legal Quality ($LEGAL$)	0.17	0.26	0.35	0.5	0.67	0.79	0.83	0.185
Labor Mkt. Flexibility ($FLEX$)	0.225	0.39	0.49	0.615	0.76	0.87	1	0.184

Appendix Table 1B
Pairwise Correlation of Country Characteristics

	$\log(H/L)_i$	$\log(K/L)_i$	$FINDEV$	$LEGAL$
$\log(K/L)_i$	0.81***			
$FINDEV$	0.58***	0.66***		
$LEGAL$	0.69***	0.63***	0.68***	
$FLEX$	0.34***	0.28**	0.21*	0.23**

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

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

	Min.	10th	25th	Med.	75th	90th	Max.	Std. Dev.
Skill intensity $(\log(H/L)^k)$	−1.971	−1.906	−1.576	−1.395	−1.082	−0.831	−0.759	0.370
Capital intensity $(\log(K/L)^k)$	2.316	2.891	3.499	3.906	4.589	5.071	6.127	0.884
Ext. Capital Dep. $(CAPDEP)$	−1.206	−0.751	−0.148	−0.028	0.165	0.587	0.941	0.498
Input Concentration (HI)	0.724	0.783	0.794	0.834	0.908	0.932	0.943	0.064
Input Relationship-Spec. (RS)	0.594	0.673	0.818	0.946	0.969	0.988	0.991	0.125
Job Complexity $(COMPL)$	0.148	0.153	0.311	0.384	0.615	0.732	1	0.221
Sales Volatility $(SVOL)$	0.124	0.130	0.144	0.152	0.179	0.198	0.219	0.026

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

	$\log(H/L)^k$	$\log(K/L)^k$	$CAPDEP$	HI	RS	$COMPL$
$\log(K/L)^k$	0.39*					
$CAPDEP$	0.52**	0.10				
HI	0.46**	−0.34	0.63***			
RS	0.09	−0.54**	0.13	0.55**		
$COMPL$	0.82***	0.19	0.65***	0.54**	0.21	
$SVOL$	−0.08	0.06	0.38*	0.11	−0.20	0.07

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

Appendix Table 3A
List of Country and Industry Groups for SMM Estimation

Exporter Groups: (Grouped by total export volumes)

Group 1: TCD; MLI; TGO; BDI; MWI; CAF; NIC; UGA; SLE; NER; MDG; BOL; SYR; SLV; HTI; PRY; JOR; PNG; SEN; NGA; HND; GHA; ZWE; CMR; KEN; GTM $(< \text{US}\$1,000,000)$
Group 2: JAM; ZMB; IRN; CRI; ECU; CIV; LKA; EGY; URY; ZAR; PAN; DOM; PER; TUN; COL; PAK; MAR $(> \text{US}\$1,000,000$ and $< \text{US}\$5,000,000)$
Group 3: DZA; HUN; CHL; GRC; VEN; NZL; POL; PHL; ZAF; ARG; TUR; ISR $(> \text{US}\$5,000,000$ and $< \text{US}\$10,000,000)$
Group 4: IND; IDN; PRT; NOR; THA; AUS; IRL; FIN; MYS; BRA; MEX; DNK; SGP; AUT; CHN; ESP; KOR; SWE; CHE $(> \text{US}\$10,000,000$ and $< \text{US}\$90,000,000)$
Group 5: CAN; BEL; NLD; GBR; ITA; FRA; JPN; USA; DEU $(> \text{US}\$90,000,000)$

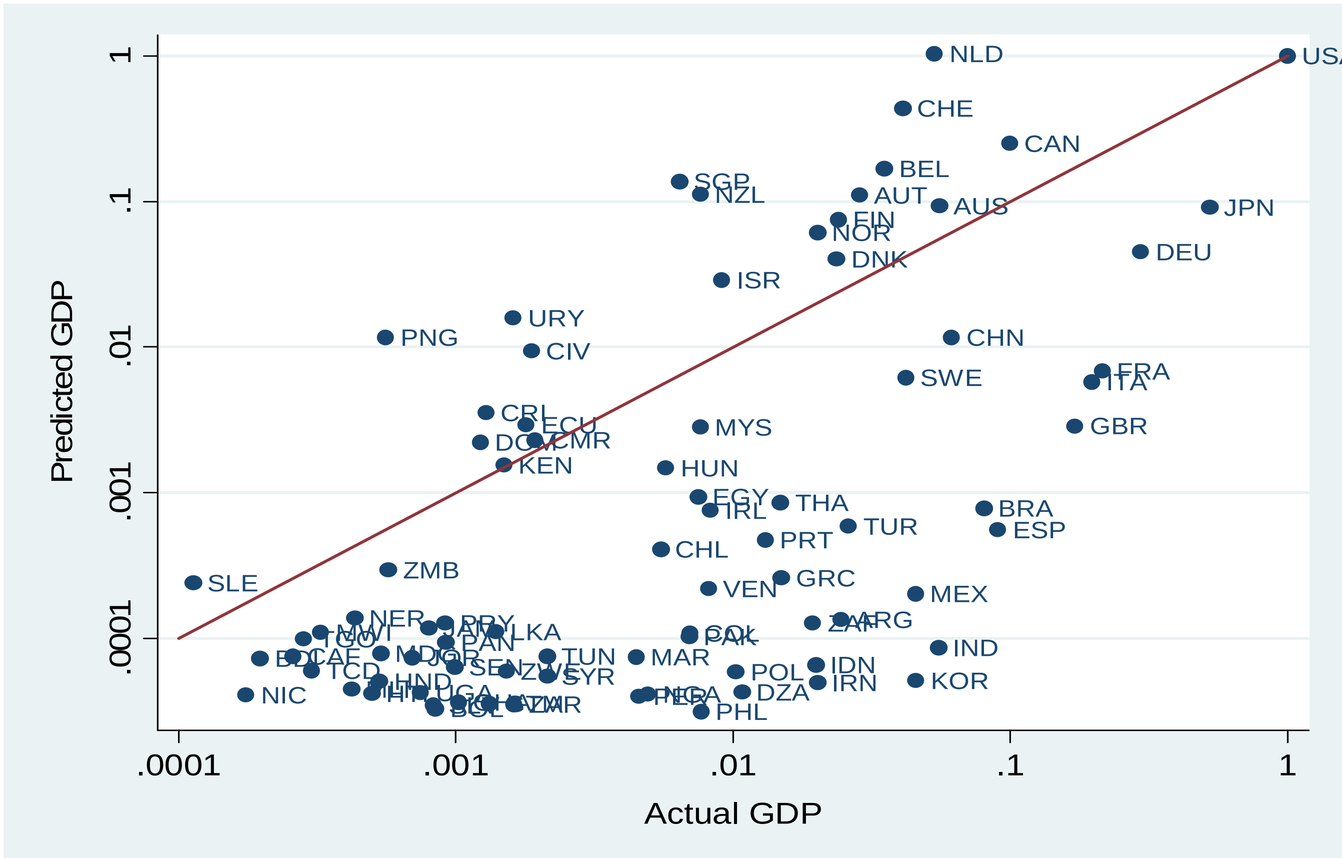
SIC Industry Groups: (Grouped by total trade volumes)

Group 1: 21; 27; 25; 31; 32; 24 $(< \text{US}\$50,000,000)$
Group 2: 30; 29; 22; 39; 26; 23; 34; 38; 33; 20 $(> \text{US}\$50,000,000$ and $< \text{US}\$200,000,000)$
Group 3: 36; 28; 37; 35 $(> \text{US}\$200,000,000)$

Appendix Table 3B
Comparison of Data Moments and Matched Simulated Moments

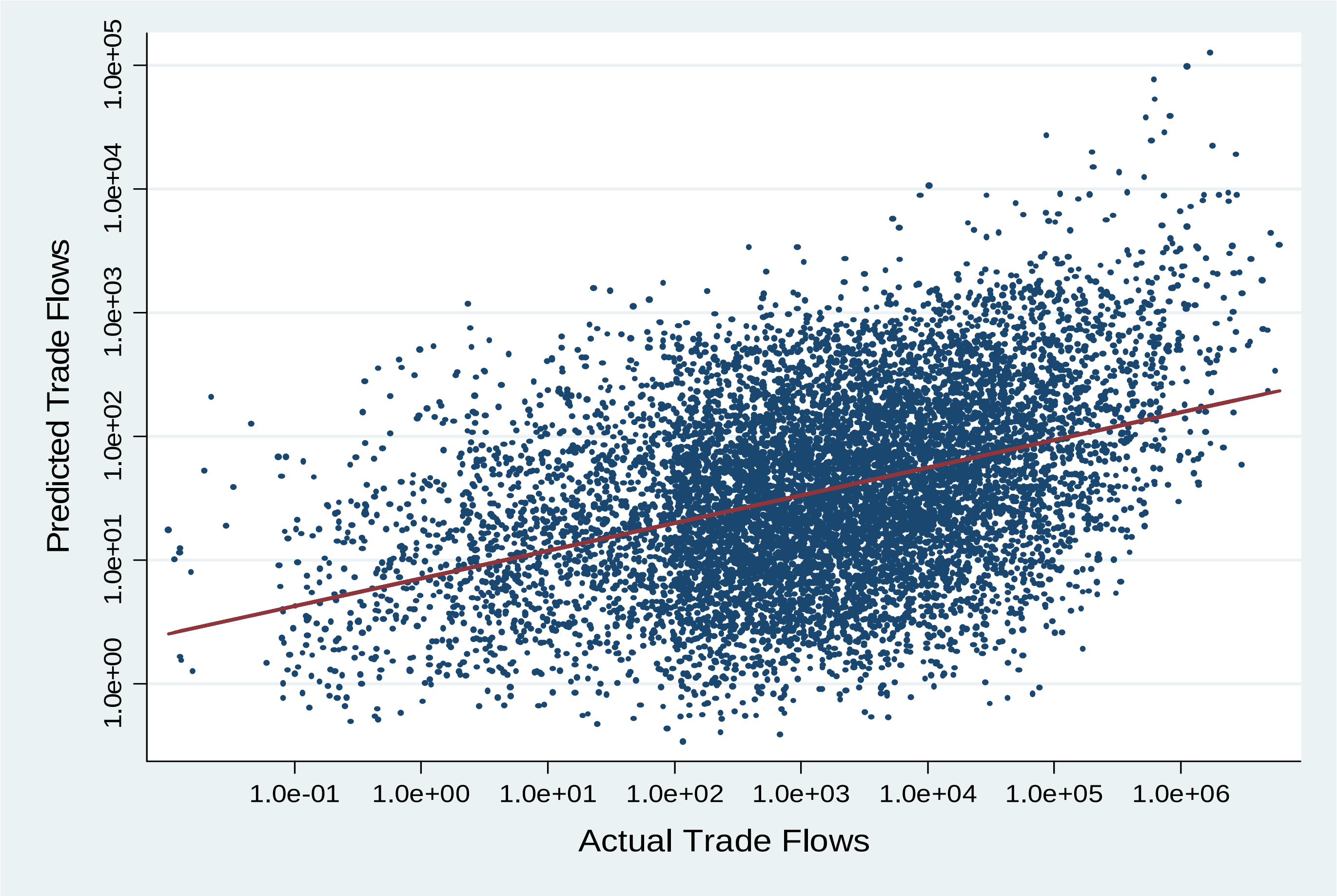
	Data moment	Matched moment (based on $\hat{\Theta}^{SMM}$)
Regression coefficients:		
β_{d1}: Log (Distance)	−1.16085	−1.16066
β_{d2}: Common Language	0.50195	0.50212
β_{d3}: Colony	0.76554	0.76529
β_{d4}: Border	0.18948	0.18979
β_{d5}: RTA	0.29025	0.29092
β_{d6}: GATT	0.21723	0.21572
β_{f1}: $\log(H/L)^k \times \log(H/L)_i$	1.24570	1.24543
β_{f2}: $\log(K/L)^k \times \log(K/L)_i$	0.16413	0.16408
β_{lm1}: $CAPDEP \times FINDEV$	1.27883	1.27978
β_{lm2}: $HI \times LEGAL$	14.30727	14.30744
β_{lm3}: $RS \times LEGAL$	9.63769	9.63810
β_{lm4}: $COMPL \times LEGAL$	2.91853	2.91879
β_{lm5}: $COMPL \times \log(H/L)$	1.46171	1.46231
β_{lm6}: $SVOL \times FLEX$	9.04316	9.04320
Trade shares:		
Exporter Group 2:	0.01398	0.01400
Exporter Group 3	0.03829	0.03833
Exporter Group 4	0.23768	0.23812
Exporter Group 5	0.70615	0.70660
SIC Group 2	0.41362	0.41393
SIC Group 3	0.51905	0.51933

Figure 1
Assessing the Goodness of Fit: Predicted vs Actual Country GDPs
(normalized, US=1)



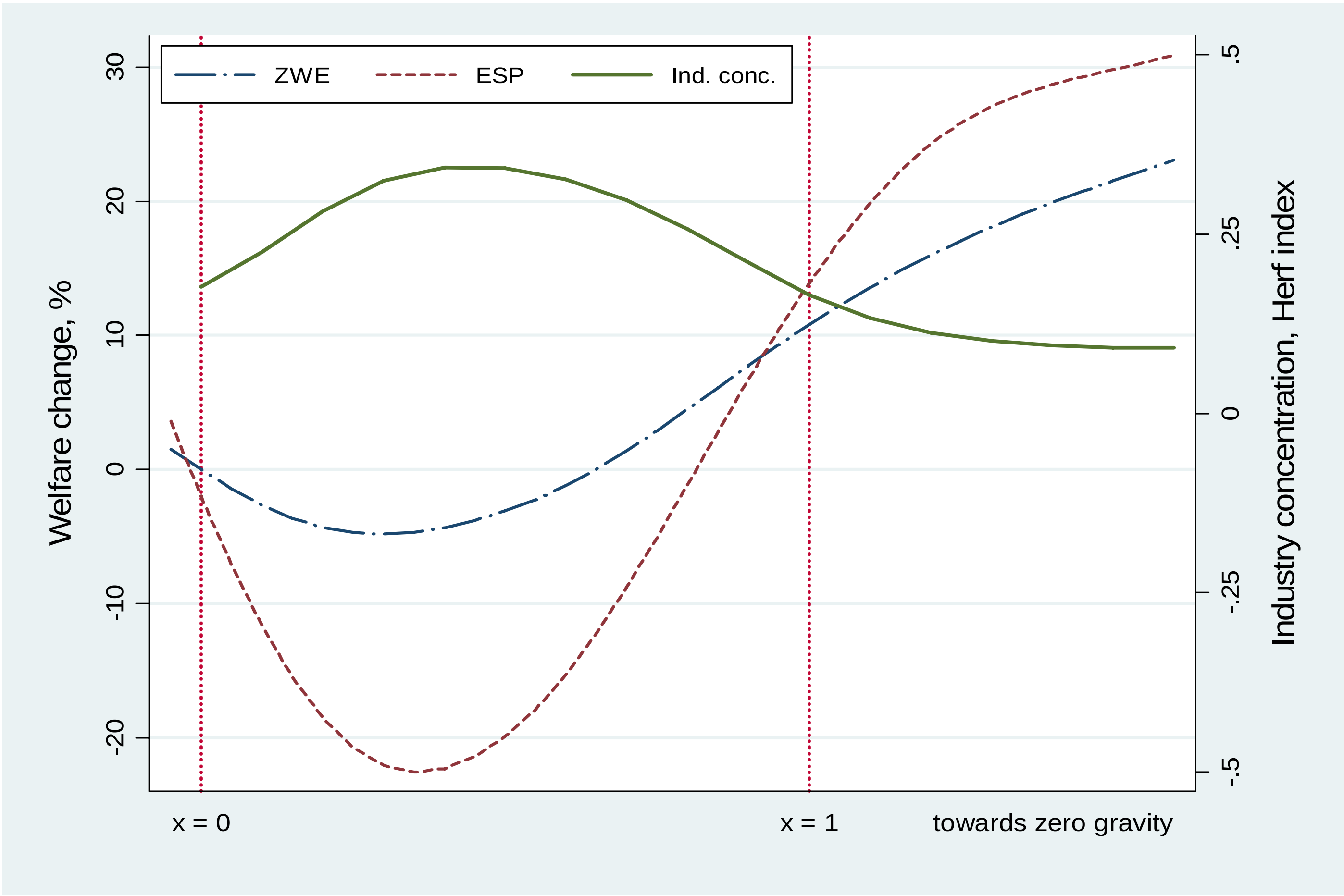
Notes: Actual GDP levels on the horizontal axis are from the World Development Indicators (WDI). The predicted country GDPs on the vertical axis are computed based on the SMM estimates. US GDP is normalized to 1. Both axes are on a log-scale. The 45-degree line is plotted for reference. The Pearson correlation between the two log-income variables is 0.58, while the Spearman rank correlation is 0.54, both significant at the 1% level.

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



Notes: Actual 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. Both axes employ a log-scale; the original units are in thousands of current (1990) US dollars. The log-linear regression line is illustrated (slope = 0.22, significant at the 1% level).

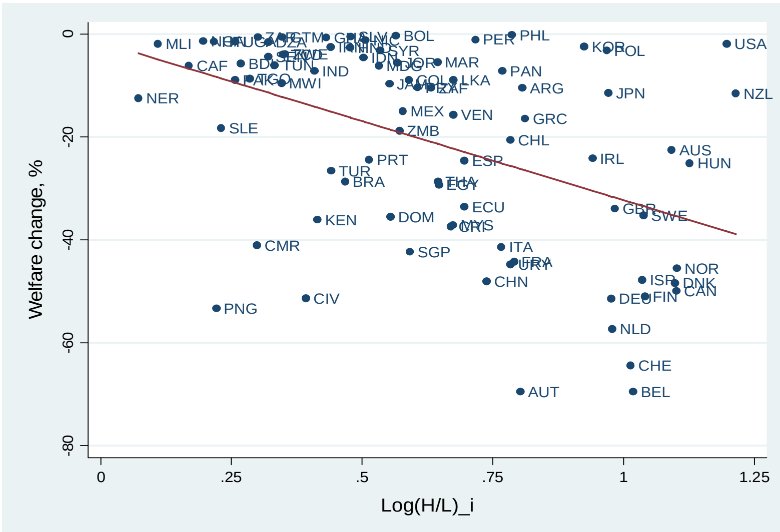
Figure 3
Counterfactual I: The Effects of Reducing Distance Barriers



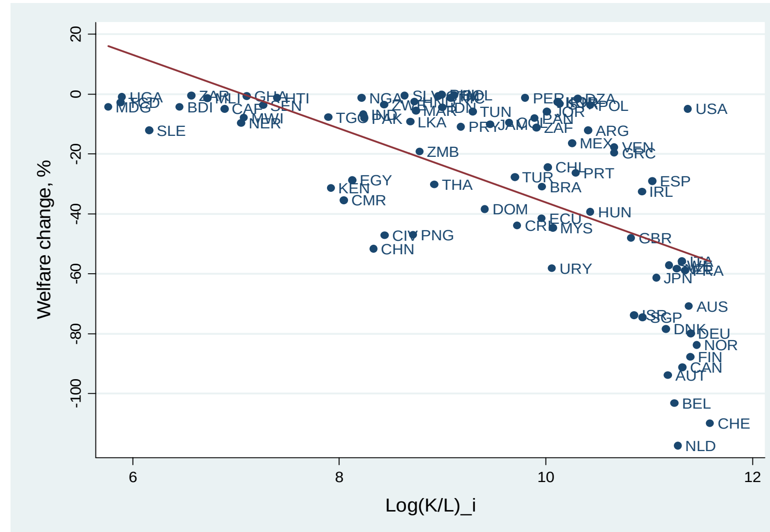
Notes: Welfare effects (measured on the left vertical axis) are illustrated for two countries: Zimbabwe (ZWE), the 25th percentile per capita income country in the sample, and Spain (ESP), the 75th percentile country. The decline in industry concentration (Herfindahl index of producer shares in each industry; mean across industries) is plotted by the solid curve (measured on the right vertical axis). The smooth curves drawn are based on locally weighted regressions.

Figure 4
Counterfactual II: Correlating Welfare Changes against Initial Country Positions

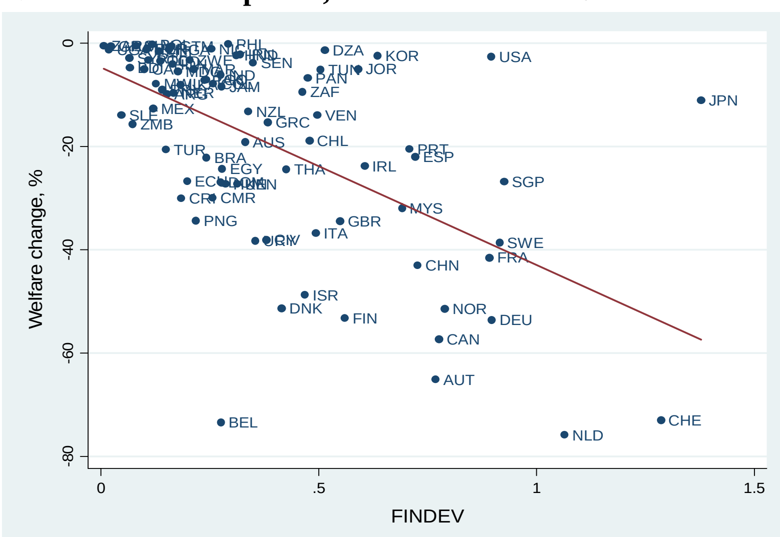
A: Human capital, $log(H/L)^k \times log(H/L)_i$



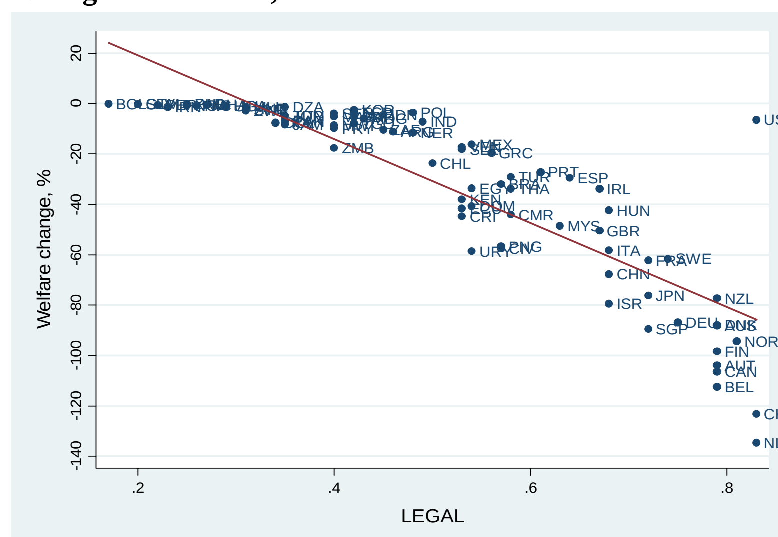
B: Physical capital, $log(K/L)^k \times log(K/L)_i$



C: Financial Development, *CAPDEP* × *FINDEV*



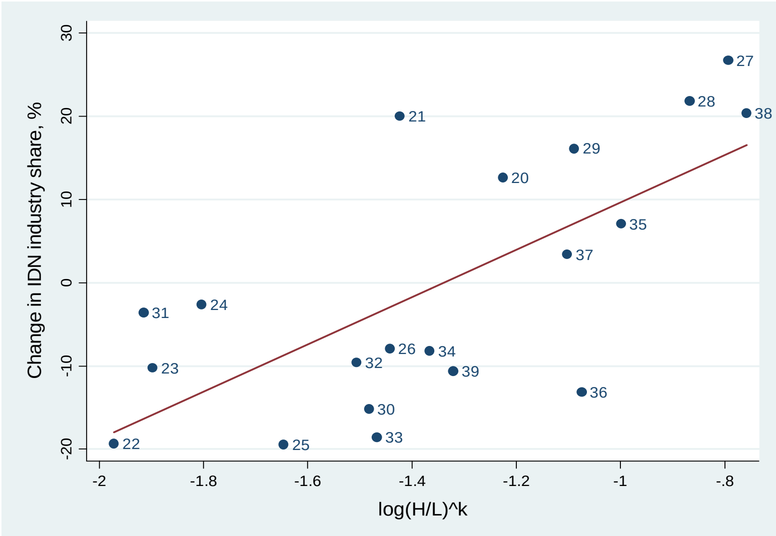
D: Legal Institutions, *RS* × *LEGAL*



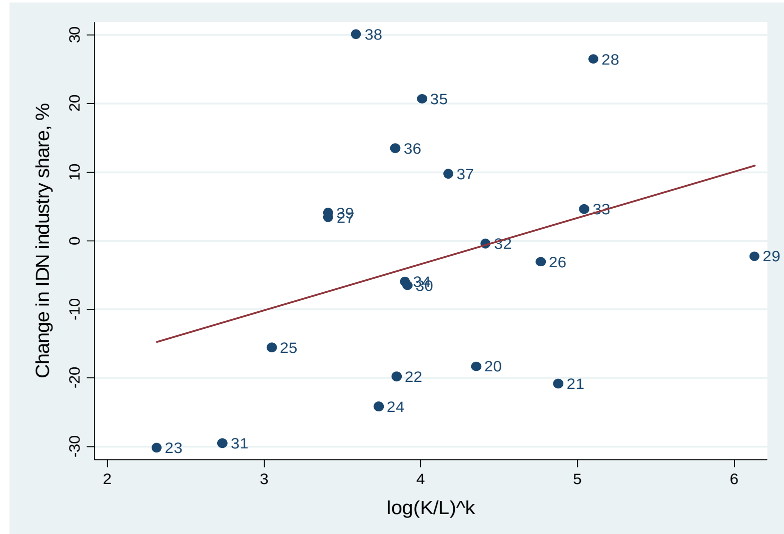
Notes: In each panel, I plot the welfare change from neutralizing the source of comparative advantage in question country-by-country. This is plotted against the initial value of the corresponding country characteristic. A linear regression line is added in each plot (each slope coefficient is negative and significant at the 1% level).

Figure 5
Counterfactual III: Shift in Industry Composition for Indonesia

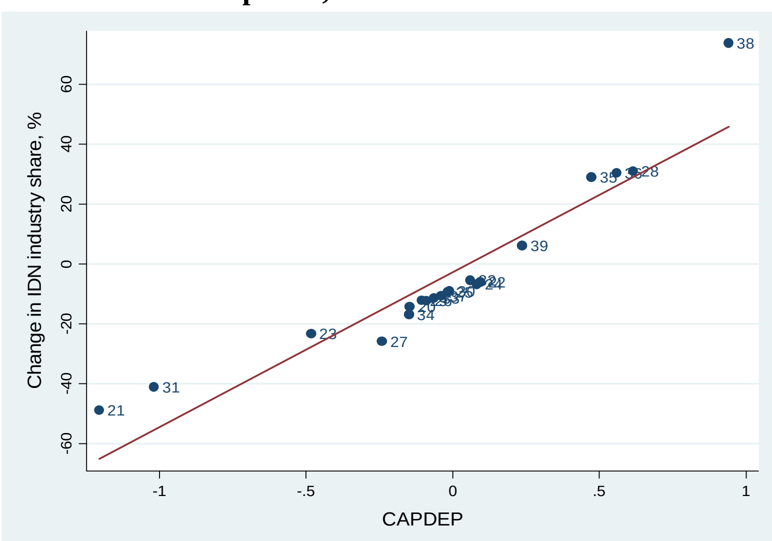
A: Human capital, *log(H/L)ᵢ*



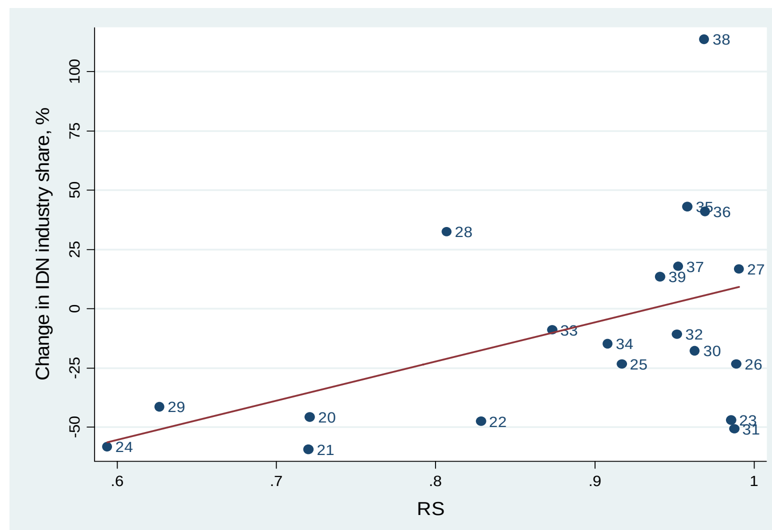
B: Physical capital, *log(K/L)ᵢ*



C: Financial Development, *FINDEV*



D: Legal Institutions, *LEGAL*



Notes: The vertical axis plots the percent change in each industry's share of Indonesia's manufacturing output following each counterfactual exercise of raising a country characteristic for Indonesia to the world frontier (the highest value observed in the sample), as described in Section 5.3. This is plotted against the corresponding industry characteristic. A linear best fit line is illustrated; in each panel, the slope is statistically significant at the 1% level (robust standard errors), except in panel B where it is marginally insignificant at the 10% level.