Beyond Gravity: The Composition of Multilateral Trade Flows

Ahmad Lashkaripour*

January 2016

Abstract

Standard gravity models describe the volume of trade in a multilateral world, but overlook three basic facts about the composition margin: (i) the effect of geography on the price composition of trade, (ii) the effect of income per capita on the price composition of exports, and (iii) the systematically higher trade-to-GDP ratio of rich countries. I develop a novel view of comparative advantage that reconciles the gravity equation with these basic facts. My approach delivers a unified model that fully describes both the volume and the composition of multilateral trade flows. A remarkable feature of the model is its ability to replicate consumption differences across rich and poor countries while adopting standard homothetic preferences. I estimate the unified model using bilateral trade data on 100 countries and compare it to a special case: the standard gravity model without the composition margin. The unified model fits the data significantly better with an $R^2$ that is 43% higher than the standard gravity model. Further, the unified model implies gains from trade that are both substantially larger and systematically biased in favor poor and remote nations.

1 Introduction

The contemporary theories of international trade deliver the gravity equation, which relates the characteristics of a country to the volume of its trade. These theories, however, overlook the composition margin; they do not systematically characterize the mix of goods a country trades. With the dramatic growth of trade between dissimilar nations, and the availability of microlevel data, it has become increasingly evident that dissimilar countries trade different types of goods. Evidence indicates that:

*Many people have provided helpful comments and suggestions that improved this paper. I am especially thankful to Jonathan Eaton, James Tybout, and Stephen Yeaple for their advice and encouragement. I also thank seminar participants at the University of British Columbia, University of California Santa Cruz, Drexel University, Indiana University, Pennsylvania State University, the New Faces in International Economics conference, the Midwest Trade Meetings, the InsTed workshop, and UECE Lisbon Meetings for helpful comments and suggestions. All errors are my own. Correspondence: alashkar@indiana.edu, http://pages.iu.edu/~alashkar/.
i. High-income countries import/export a higher share of their GDP.

ii. High-income countries specialize in high-price tradables.

iii. Countries export higher-price tradables to faraway markets (the “Washington apples” effect).\(^1\)

The first two facts point to a systematic relationship between income per capita and the composition of trade, i.e. rich countries trade goods that exhibit higher prices and higher degrees of tradability. The third fact points to a systematic relationship between geography and the composition of trade. All three facts are beyond the scope of both standard gravity models and Neoclassical trade theories.\(^2\) Three independent bodies of literature have emerged, addressing each fact individually. However, there is no unified theory that accounts for all three facts.

I address this void by developing a novel view of comparative advantage that collectively explains the three facts highlighted above. My view of comparative advantage is integrated with national product differentiation to construct a unified model that describes both the volume and the composition of multilateral trade flows. When I eliminate comparative advantage from the unified model, it reduces to a standard gravity model that overlooks the composition margin. I estimate the unified model and compare it to this special case. The comparison reveals that the composition margin is quantitatively important and has profound effects on the gains from trade. Specifically, accounting for the role of composition more than triples the gains from trade, and shifts them in favor of poor and remote nations.

To explain the effect of geography and income on the composition of trade, I take an alternative view from the literature. Theories that address the composition margin, usually require non-homothetic demand to explain the role of per capita income and rely on non-iceberg trade costs to account for the effect of geography. To reconcile the role of geography with income, I purposefully abstract from both non-homotheticity and non-iceberg trade costs. I instead confront a common assumption of the gravity models that is inconsistent with micro level evidence. Standard gravity models assume that the scope for product differentiation is the same across all goods.\(^3\) I relax this assumption, and allow for two types of goods that offer different scopes for product differentiation: a highly-differentiated (low-\(\sigma\)) type and a less-differentiated (high-\(\sigma\)) type.

In equilibrium, patterns of trade are determined by how countries concentrate their production and consumption across the two types of goods. Countries are characterized by their National

\(^{1}\)The first fact is documented by Limao and Venables (2001), Waugh (2010), and Caron, Fally, and Markusen (2014). The second fact is documented at various levels of aggregation: Schott (2004) shows that rich countries have higher within-category export prices; Hummels and Klenow (2005) show that rich countries have higher aggregate export prices. The third fact is documented by Hummels and Skiba (2004) and Baldwin and Harrigan (2011), among others.

\(^{2}\)The Neoclassical view of comparative advantage can only explain fact (ii) under the extra assumption that high-quality, high-price goods are capital-intensive (Schott (2004)).

\(^{3}\)Broda and Weinstein (2006) estimate that different HS10 product categories exhibit substantially different degrees of product differentiation. At an even more disaggregated level, Berry, Levinsohn, and Pakes (1995) estimate that luxury cars have systematically lower price elasticities of demand than economy cars, which implies that luxury cars are highly-differentiated whereas economy cars are less-differentiated.
Product Quality and Labor endowment. Advanced countries (by definition) are endowed with higher National Product Qualities and produce more appealing varieties of both types. This entails more global demand, which drags up their equilibrium wage. The higher wage paid by advanced countries makes them comparatively disadvantaged in the less-differentiated, price-sensitive type. Therefore (even in the absence of technical comparative advantage) high-wage countries have revealed comparative advantage in the highly-differentiated (low-σ) type; whereas low-wage countries have comparative advantage in the less-differentiated (high-σ) type. Put differently, the autarky relative price index of the highly-differentiated type is lower in high-wage countries. As a result, in equilibrium, rich countries concentrate their production on the highly-differentiated type, whereas poor countries specialize in the less-differentiated type.

Consumption patterns are determined by what I term the home-production effect on local consumption. In the presence of trade costs, consumers spend relatively more on the locally abundant type, which is relatively cheaper. Specifically, due to costly trade, the relative price index of the highly-differentiated type is lower in rich countries than in poor countries. This induces households in rich countries to consume relatively more of the highly-differentiated type. Hence, countries with identical, homothetic preferences concentrate their consumption on different types of goods. This effect is the opposite of the “home-market effect” formalized by Krugman (1980). The “home-market effect” states that local demand determines patterns of local production, whereas here (in face of costly trade) local production determines the structure of local consumption.

In equilibrium, rich countries are both the net exporters and the main consumers of the highly-differentiated type, which has a lower demand elasticity and exhibits two equilibrium properties: (i) it comes with a higher markup, and (ii) it is traded more heavily (as it is subject to a lower trade elasticity). Rich countries have higher trade-to-GDP ratios because they produce and consume relatively more of the highly-differentiated type, which is traded more intensively. Rich countries have higher aggregate export prices due to two reasons. First, rich countries sell both types of goods at a higher price due to their higher National Product Quality. Second, rich countries export relatively more of the highly-differentiated, high-markup type. The second channel is novel and points to a purely compositional effect. The effect of geography on the price composition of trade can be explained along similar lines. In equilibrium, distant countries trade relatively more of the highly differentiated, high-markup type. This is due to remote exporters facing higher trade costs and being price-disadvantaged; inducing them to sell relatively more of the high-markup type that is price-insensitive. This behavior gives rise to the “Washington apples” effect, even in the

\[4\] Unlike Neoclassical models, in the unified model comparative advantage is regulated by demand. In the Heckscher-Ohlin model goods differ in the labor or capital intensity of their production. In the present model, however, there are two types of goods: the low-σ type has a quality-intensive demand and the high-σ type has a quantity-intensive demand. In the trade equilibrium, quality-abundant (labor-abundant) nations specialize in the quality-intensive (quantity-intensive) type.

\[5\] Preferences are nested CES, and expenditure shares across types are determined by the relative price indexes.

\[6\] Atkin (2013) finds strong support for this effect using micro-level consumption data from India. Moreover, this effect would prevail in any multi-sector gravity framework with across-sector substitutability. However, to my knowledge, no previous study has highlighted and quantified the importance of this effect in a multi-country gravity setting.
presence of iceberg trade costs.\footnote{In Lashkaripour (2015), I analyze detailed trade data and find that the theory highlighted above is the main driver of the “Washington apples” effect in US imports.}

In the unified model, \textit{national product differentiation} governs the trade of the same type (of good) between similar countries. I accommodate \textit{national product differentiation} by letting firm-specific varieties from the same country to be closer substitutes (the unified model, therefore, nests the Armington gravity model.\footnote{The Armington model implicitly assumes that varieties from the same country are perfect substitutes (which entails perfect competition). I relax this restriction, and develop a method that tractably combines \textit{national product differentiation} with monopolistic competition.} Altogether, the unified model combines trade across types that is driven by comparative advantage, with within-type trade driven by standard gravity forces.\footnote{Specialization across highly- and less-differentiated types (like specialization across quality types) corresponds to \textit{within-product} specialization. Schott (2004) finds strong support for the kind of specialization in US imports.} This gives rise to two distinct welfare-improving effects of foreign trade. First, like gravity models, trade increases the number of differentiated varieties. Second, like Neoclassical models, trade allows for production specialization, which lowers the relative price index of the comparatively disadvantaged type in each country. More specifically, \textit{quality-abundant} countries shift production towards the highly-differentiated type, whereas \textit{labor-abundant} countries shift production towards the less-differentiated type. Welfare improves universally because the relative price index of the highly-differentiated type falls in poor (labor-abundant) countries, and the relative price index of the less-differentiated type falls in rich (quality-abundant) countries.

I estimate the unified model using bilateral trade data for 100 countries. The sample represents more than 95% of the world trade, and includes countries that are vastly dissimilar in their GDP per capita. To demonstrate the merits of the unified model, I compare it to a special case (with one type of good) that delivers the pure gravity equation. The unified model fits the data significantly better, with an $R^2$ that is 43\% higher than the pure gravity model.\footnote{The pure gravity model, itself, fits the data substantially better than the standard Armington model since it estimates the scope for national product differentiation instead of forcing it to be complete.} The superior fit of the unified model stems from its ability to explain trade between both similar and dissimilar nations. Specifically, the unified model matches two margins of the data that are beyond the scope of the pure gravity model. First, the unified model allows for South-South trade to be systematically smaller than North-North trade. As a result, it correctly predicts the higher trade-to-GDP ratio of rich countries, whereas the gravity model predicts the opposite. Second, the unified model allows for (endogenously determined) country-specific trade elasticities and, thus, accommodates the smaller distance elasticity of export flows from rich and remote nations. The estimation-fit, however, does not fully manifest the merits of the model. The unified model also performs remarkably well in matching (\textit{out-of-sample}) facts regarding the price composition of exports and the evolution of trade patterns over time.

I use the estimated model to demonstrate the importance of the composition margin on the gains from trade. Arkolakis, Costinot, and Rodriguez (2012) show that in gravity models the trade-to-GDP ratio and trade elasticity are sufficient statistics for evaluating the gains from trade relative
to autarky. Gravity models, however, systematically understate the trade-to-GDP ratio of rich countries and counter-factually assume a common trade elasticity for all trading nations. The unified model corrects the gravity model along both dimensions, by accommodating comparative advantage-driven trade and pining down the composition of aggregate trade flows. As a result, the gains from trade are more than four-times larger in the unified model. The gains are systematically larger for rich countries, because the model could match their higher trade-to-GDP ratios. The gains are systematically larger for poor countries because they import relatively more of the highly-differentiated type and have systematically lower trade elasticities. These findings could help close the gap between the relatively small gains implied by structural gravity models and those implied by direct cross-country comparisons in Frankel and Romer (1999).

The composition margin not only influences the overall size of the gains from trade, but also their distribution across nations. Comparative advantage-driven trade (which is absent in the pure gravity model) increases the welfare of the average country by 28%. The gains from comparative advantage are, however, systematically biased in favor of poor and remote nations. Further, in stark contrast to the pure gravity model, the unified model predicts that liberalizing trade could reduce international income inequality by more than 10%. The intuition is that, poor countries are affected more severely by trade impediments as they specialize in less-differentiated goods with low degrees of tradability. Eliminating these impediments would, therefore, make poor countries relatively more competitive, and would improve their income relative to rich countries.

This paper contributes to a contemporary literature that describes the composition of multilateral trade flows. In general, existing studies either focus on the role of geography by assuming non-iceberg trade costs (e.g. Hummels and Skiba (2004); Irarrazabal, Moxnes, and Opronolla (2014)) or explain the effect of per capita income with non-homothetic preferences (e.g. Flam and Helpman (1987); Matsuyama (2000); Hallak (2006); Fajgelbaum, Grossman, and Helpman (2011); Fieler (2011); Caron et al. (2014); Feenstra and Romalis (2014)). I contribute to this literature by developing a unified perspective that integrates geography and income. Apart from generality, the unified model is distinct from existing theories in two aspects: (i) it assumes standard homothetic preferences and iceberg trade costs, which make it amenable to straightforward estimation, and (ii) it is consistent with how the composition of trade has transformed over time. More importantly, the unified model contributes to our understanding of the welfare gains from trade. To my knowledge, this paper is the first to show that embedding systematic specialization into a gravity model (to accommodate the composition margin) tremendously magnifies the gains from trade;

---

11Feenstra and Romalis (2014) address both the effect of geography and income on the price composition of trade, by assuming both non-homothetic preferences and non-iceberg trade costs. Their model, however, does not accommodate the higher trade-to-GDP ratio in rich countries. In general, papers that confront the role of per capita income either explain international differences in trade-to-GDP ratios or export prices. To my knowledge, however, there is no existing framework that accommodates both.

12With all their merits, theories based on non-homothetic preferences do not explain the recent rise of dissimilar-dissimilar (North-South) trade relative to similar-similar trade, which is quite substantial and is highlighted extensively in Krugman (2009) and Hanson (2012), among others (see section 5.2 for a detailed discussion).
especially for poor and remote nations.\footnote{\cite{Costinot2012} argue that accommodating sectoral specialization increases the gains from trade only marginally. They impose, however, that specialization is regulated by sector-specific technologies. Their model, therefore, does not accommodate the three (composition-related) facts highlighted in this paper. \cite{Costinot2013} and \cite{Ossa2012} show that embedding multiple sectors into a gravity model magnifies the gains from trade. Both studies, however, exogenously fix expenditure shares on heavily imported sectors. I show that embedding production specialization into a multi-sector gravity model tremendously magnifies the gains from trade, even when consumption shares are endogenously determined.}

The notion that countries specialize across goods with different degrees of differentiation has been also put forward by \cite{Helpman1985}, \cite{Hanson2004}, and \cite{Faigelbaum2011}.\footnote{Similarly, \cite{Caron2014} and \cite{Fieler2011} postulate that countries specialize across goods with different degrees of technology differentiation. This assertion combined with the assumption that technologically differentiated goods are more income elastic explains the higher trade-to-GDP ratio of rich countries.} In these studies specialization is driven by the home-market effect and patterns of specialization are regulated primarily by market size. These frameworks, though, rely on non-homotheticity to isolate the role of per capita income, and overlook the effect geography on export prices. This paper, by contrast, highlights an alternative channel of comparative advantage that integrates the role of geography, per capita income, and market size. My paper is also related to a vibrant literature that studies within-product specialization in international trade. Most studies have underlined the role of specialization across quality types, which is driven by differences in factor abundance, non-homotheticities and non-iceberg trade costs (\cite{Schott2004}; \cite{Hummels2004}; \cite{Hallak2006}; \cite{Faigelbaum2011}; \cite{Feenstra2014}; \cite{Sutton2014}). I contribute to this literature by highlighting a distinct channel through which countries specialize across low- and high-markup types. Importantly, my view of specialization reconciles several salient facts regarding export unit values, and receives support from micro-level data.\footnote{More generally, this paper is related to a literature that integrates production specialization into gravity models (\cite{Helpman1985}; \cite{Markusen1986}; \cite{Davis1995}; \cite{Bernard2007}; \cite{Romalis2004}). I contribute to this literature by (i) proposing a novel view of production specialization that is consistent with micro-level evidence, (ii) my view of production specialization delivers three modern facts about the composition of aggregate trade flows, whereas existing studies build on the Noeclassical view that does not accommodate these facts, and (iii) Existing theories are generally confined to basic settings with two countries, while this paper develops and estimates a multi-country, general equilibrium model that tractably combines within and across-category trade among many dissimilar countries. This paper has less in common with the multi-sector gravity frameworks in \cite{Caliendo2014}, \cite{Costinot2012}, \cite{Fieler2011}, and \cite{Levchenko2011}. First, unlike these studies, I accommodate monopolistic competition and across-markup specialization. Second, in these frameworks, consumption shares are either fixed across types or regulated by non-homotheticities. Therefore, even when I shut down across-markup specialization, the unified model delivers predictions that are distinct from the above frameworks (see appendix B).}

At a broader level, this paper contributes to a literature that examines structural differences in trading behavior across rich and poor countries. There is a long-standing consensus that (i) rich and poor countries have systematically different consumption structures (\cite{Markusen1986}; \cite{Hunter1988}) and (ii) poor countries face larger export frictions (\cite{Rodrik1998}; \cite{Limao2001}; \cite{Waugh2010}). Less research has been done, however, on the source of these structural differences. This paper points to one potential source: comparative advantage. Specifically, I show that (i) comparative advantage could induce consumption dissimilarity across rich and poor countries, and (ii) poor countries are affected more severely by trade impediments.
as they specialize in price-elastic goods with low degrees of tradability. My view of consumption dissimilarity could explain the recent rise of North-South relative to North-North trade. Moreover, in line with Waugh (2010), my model predicts that eliminating the global trade impediments could reduce international income inequality by more than 10%.

2 Theory

The unified model combines monopolistic competition and national product differentiation in a multi-country general equilibrium framework with two types of goods. There are two driving forces behind trade: production specialization and national product differentiation. Production specialization governs trade across types and is motivated by comparative advantage. National product differentiation is the driver of within-type trade. The two forces, combined, determine the volume and the composition of aggregate trade flows.

2.1 The Environment

There are $N$ countries; $C = \{1, \ldots, N\}$ denotes the set of countries. Population $L_i$, and National Product Quality $\alpha_i$ characterize country $i \in C$. There are two types of goods: $H$ and $L$. Each type comes in a continuum of firm-specific varieties. Firm-specific varieties are (horizontally) differentiated both at the firm level and at the national level.

Demand. The demand structure accommodates two features that are implied by micro-level evidence: (i) the scope for product differentiation varies across different types of goods (Broda and Weinstein (2006)), and (ii) quality and quantity are not isomorphic (Baldwin and Ito (2008); Rodrik (1994); Bils and Klenow (2001)). Both futures could be incorporated with a standard, homothetic nested-CES utility function. Specifically, consumers in country $i$ maximize the following utility

$$U_i = \left[ \sum_{z \in \{H, L\}} (U_i^z)^{\frac{\gamma}{2}} \right]^{\frac{2}{\gamma-1}}$$

16 When quality and quantity are not isomorphic, some goods/industries are more quality-intensive than others. Baldwin and Ito (2008) show that some industries are characterized by quality-competition, whereas others operate on the basis of price competition. Similarly, Rodrik (1994) argues that some goods are more quality-intensive than others, and that this feature is quantitatively important in explaining growth patterns across developing countries. Fan, Li, and Yeaple (2015) show that quality upgrading and tariff reductions have differential effects on revenues across different products. Bils and Klenow (2001) show that the elasticities of demand with respect to product quality and price could diverge depending on income and the product class.
where $U_i^z$ is the sub-utility corresponding to type $z \in \{H, L\}$. $U_i^z$ is a CES aggregator across all national varieties of type $z$

$$U_i^z = \left[ \sum_{j=1}^{N} \alpha_j^{1-p_z} \left( Q_{ji}^z \right)^{p_z} \right]^{\frac{1}{p_z}}$$

(1)

The utility attained from type $z$ varieties produced in country $j$ is a Cobb-Douglas combination of their aggregate quantity, $Q_{ji}^z$, and their quality, $\alpha_j$.\(^\text{17}\) Note that $\alpha_j$ is the National Product Quality of country $j$, and applies to all varieties produced in that country — i.e. countries endowed with a higher $\alpha_j$, produce more appealing varieties of both type $H$ and $L$. Aggregate quantity, $Q_{ji}^z$, is a CES aggregator across quantities purchased from individual firms in country $j$:

$$Q_{ji}^z = \left[ \int_{\omega \in \Omega_{ji}} \left( q_{ji}^z \right)^{\bar{p}_z} d\omega \right]^{\frac{1}{\bar{p}_z}} = M_{ji}^{\frac{1}{\bar{p}_z}} q_{ji}^z,$$

where $q_{ji}^z$ denotes the quantity purchased (of type $z$) from a typical firm $\omega$ located in country $j$ (note that firms are homogeneous). $M_{ji}$ and $\Omega_{ji}$ denote the mass and the set of firms that sell from country $j$ to $i$, respectively. In the above three-tier nested-CES utility, $\varepsilon$ is the elasticity of substitution between types $H$ and $L$; $\sigma_z = 1 / (1 - p_z)$ is the elasticity of substitution between (aggregated) national varieties of type $z$; $\bar{\sigma}_z = 1 / (1 - \bar{p}_z)$ is the intra-national elasticity of substitution between firm-specific varieties of type $z$.

The above demand structure nests both the Krugman and Armington models. When $\bar{\sigma}_z = \sigma_z$ there is no scope for national product differentiation, and the demand structure (for each type) reduces to that of Krugman (1980). If $\bar{\sigma}_z \to \infty$, the scope for national product differentiation is complete, similar to the Armington model. In this paper I adopt a middle ground. Specifically, I allow for some

\(^{17}\)Product Quality, $\alpha_j$, and quantity are assumed to be Cobb-Douglas complements. This specification has been used in earlier gravity models (Anderson (1979) and Deardorff (1998)) and more recently in Eaton, Kortum, and Kramarz (2011). The gravity equation that emerges from this specification is the same as the one implied by the technology-driven model in Eaton and Kortum (2002). The alternative specification adopted by Hallak and Schott (2011) and Kugler and Verhoogen (2012) is the following:

$$U_i^z = \left[ \sum_{j=1}^{N} \left( \alpha_j Q_{ji}^z \right)^{p_z} \right]^{\frac{1}{p_z}}$$

The above specification is generally employed in settings where the scope of product differentiation is the same across all goods. When applied to a multi-type environment, it implies that the importance of product quality (relative to price) is the same for all types or sectors, which is inconsistent with the findings of Baldwin and Ito (2008), Rodrik (1994), and Fan et al. (2015). Moreover, the above specification would imply that high-quality producers (like low-cost producers) sell relatively more of the high-price-elasticity types, which contradicts the findings of Hausman, Leonard, and Zona (1994), Berry et al. (1995), and Goldberg (1995) — These studies find that within a narrowly defined markets, high-quality producers face a lower price elasticity. Additionally, from a theoretical perspective, Sutton (2007) and Hallak and Sivadasan (2013) show that the isomorphism between quality and productivity breaks down under realistic assumptions. Hallak and Sivadasan (2013) argue that models with quality-productivity isomorphism explain the exporter premia, but fail to account for the conditional exporter premia. Roberts, Xu, Fan, and Zhang (2012) too highlight the distinction between cost-shifters and demand-shifters using firm-level evidence.
degree of national product differentiation, which is the same for both types:

$$\frac{\bar{\sigma}_{H} - 1}{\sigma_{H} - 1} = \frac{\bar{\sigma}_{L} - 1}{\sigma_{L} - 1} \equiv \eta > 1$$

The above specification indicates that firm-specific varieties produced in the same country are closer substitutes, and $\eta$ denotes the degree of national product differentiation ($\eta \to \infty$ corresponds to the standard Armington case where varieties from the same country are perfect substitutes).\(^{18}\)

Importantly, type $H$ and type $L$ goods are systematically different. Type $H$ offers a greater scope for product differentiation than type $L$:

$$\sigma_{H} < \sigma_{L} \iff \rho_{H} < \rho_{L}$$

Hence, by definition, preferences for type $L$ are quantity-intensive, whereas preferences for type $H$ are quality-intensive (this is implied by the fact that $\rho_{H} < \rho_{L}$ in equation 1). From now on, I will refer to type $H$ as the highly differentiated (low-$\sigma$) type, and to type $L$ as the less-differentiated (high-$\sigma$) type. In summary, on the demand side, $\sigma_z$ regulates the scope for product differentiation in type (or sector) $z$, while $\eta$ regulates the degree of national product differentiation in the economy.

**Supply.** As in Krugman (1980), firms are monopolistically competitive and homogeneous.\(^{19}\) Unlike the Krugman model, entry cost is paid locally (and separately) for each market, so there are no global economies of scale.\(^{20}\) Labor is the only factor of production. One unit of labor is required to produce one unit of each type $z \in \{H, L\}$. The unit labor cost is the same for both types and for all countries. Exports from country $j$ to $i$ are subject to an iceberg trade cost, $\tau_{ji}$. Altogether, the marginal cost of producing type $z$ in country $j$, and selling it in country $i$ is

$$mc_{ji}^z = \tau_{ji}w_j$$

where $w_j$ is wage in country $j$. Let $p_{ji}^z$ denote the price of type $z$, produced in country $j$, and sold in country $i$. A typical firm exporting $q_{ji}^z$ units of type $z$ from country $j$ to $i$ collects a variable profit equal to

$$\pi_{ji}^z = \left( p_{ji}^z - \tau_{ji}w_j \right) q_{ji}^z$$

\(^{18}\)In section 3, I estimate $\eta$ without imposing any restrictions, which delivers $\eta = 3.16$. When I impose the Armington assumption that $\eta \to \infty$, the fit of the model decreases by 20%.

\(^{19}\)Appendix A demonstrates that the predictions of the model are robust to adding firm-level heterogeneity.

\(^{20}\)The results presented in this paper are not sensitive to adding global economies of scale. The only reason I abstract from economies of scale is (i) to isolate the role of per capita income from population, and (ii) to avoid the counter-factual prediction that domestic trade shares and relative income levels increase too steeply with country size (see Ramondo, Rodríguez-Clare, and Saborío-Rodríguez (2012)).
The firms use the combined variable profits from selling both types to pay the local entry cost, which is $f^e$ units of home labor. The free entry condition (for market $i$) is, therefore, given by

$$\sum_{z \in \{H, L\}} \pi_{ji}^z = w_j f^e$$

The free entry condition determines $M_{ji}$: the mass of firms that export from country $j$ to $i$.$^{21}$

**Equilibrium.** Let $X_{ji}^z = M_{ji} p_{ji}^z q_{ji}^z$ denote the amount spent by country $i$ on type $z$ goods from country $j$. Utility maximization implies that:

$$X_{ji}^z = \alpha_j \left( \frac{p_{ji}^z}{p_i^z} \right)^{1-\sigma_z} \left( \frac{p_i^z}{P_i} \right)^{1-\epsilon} w_j L_i \tag{2}$$

where $p_{ji}^z$ denotes the price index of exports from country $j$ to $i$ of type $z \in \{H, L\}$:

$$p_{ji}^z \equiv \left[ \int_{0 \in \Omega_j} \left( p_{ji}^z \right)^{1-\alpha_z} d\omega \right]^{1-\sigma_z} = M_{ji}^{1-\sigma_z} p_{ji}^z,$$

$p_i^z$ denotes the price index of type $z$ in country $i$:

$$p_i^z \equiv \left[ \sum_{k \in C} \alpha_k \left( p_{ki}^z \right)^{1-\sigma_z} \right]^{1-\sigma_z}.$$

$P_i$ is the aggregate price index in country $i$ (aggregated across both types):

$$P_i \equiv \left[ \sum_{z \in \{H, L\}} \left( p_i^z \right)^{1-\epsilon} \right]^{1-\epsilon}.$$

In equation 2, $\left( \frac{p_i^z}{P_i} \right)^{1-\epsilon}$ is the share of spending on type $z$; $\alpha_j \left( \frac{p_{ji}^z}{p_i^z} \right)^{1-\sigma_z}$ is the share of spending on country $j$ varieties of type $z$. A typical firm from country $j$, therefore, sells $x_{ji}^z \equiv p_{ji}^z q_{ji}^z = \frac{X_{ji}^z}{M_{ji}}$ dollars of types $z$.

---

$^{21}$The assumption that firms are multi-product implies that $M_{ji}^H = M_{ji}^L = M_{ji}$. This assumption is adopted only for expositional purposes, and is conservative in terms of explaining the patterns I highlighted earlier. Section 2.2 will demonstrate that the superior explanatory power of the unified model comes from embedding comparative advantage into a gravity model. Assuming $M_{ji}^H = M_{ji}^L = M_{ji}$, implies that comparative advantage does not regulate the relative number of exported varieties ($\frac{M_{ji}^H}{M_{ji}}$). Therefore, if I relax this assumption, the effect of comparative advantage would be magnified, and so would the explanatory power of the unified model with respect to the composition margin.
Firms are Monopolistically competitive, and charge a type-specific markup over marginal cost

\[ p_{ji} = \frac{\bar{\sigma}_z}{\bar{\sigma}_z - 1} \tau_{ji}w_j = \left[ 1 + \frac{1}{\eta (\sigma_z - 1)} \right] \tau_{ji}w_j \]

Note that firms charge a higher markup on the highly differentiated (low-\(\sigma\)) type: \(\frac{\bar{\sigma}_{ji}}{\bar{\sigma}_{il}} > \frac{\bar{\sigma}_{ji}}{\bar{\sigma}_{il}}\).

Plugging the equilibrium price into the demand function (equation 2) delivers a type-specific gravity equation:

\[ X_{zi}^e = \frac{\alpha_j M_{ji}^{1/\sigma_z} (\tau_{ji}w_j)^{1-\sigma_z} X_i^e}{\sum_{k \in C} \alpha_k M_{ki}^{1/\sigma_z} (\tau_{ki}w_k)^{1-\sigma_z}} \]

where \(X_i^e \equiv \left( \frac{p_i}{\bar{p}_i} \right)^{1-\epsilon} w_iL_i\), denotes total spending on type \(z\) in country \(i\). The above gravity formulation indicates that in the less-differentiated (high-\(\sigma\)) sector, \(L\), trade shares are relatively more sensitive to price, whereas in the highly-differentiated (low-\(\sigma\)) sector, \(H\), trade shares are relatively more sensitive to National Product Quality.

The number of firms entering market \(i\) from country \(j\), \(M_{ji}\), is pinned down by the free entry condition:

\[ M_{ji} = \frac{1}{w_jf_{e}} \left[ \frac{X_{ji}^H}{\bar{\sigma}_H} + \frac{X_{ji}^L}{\bar{\sigma}_L} \right] \]

where, by assumption, \(\bar{\sigma}_z - 1 = \eta (\sigma_z - 1)\). Note that upon entry, firms collect higher profits from selling the highly-differentiated, high-markup type, \(H\). Therefore, countries that export relatively more of type \(H\) are represented by more firms in the global markets. Finally, balanced trade implies that

\[ w_jL_j = \sum_{i \in C} X_{ji}^H + X_{ji}^L \]

The above equation insures that for any country \(j\), total spending equals totals sales.

### 2.2 Four Equilibrium Outcomes that Describe the Global Economy

This section describes four basic outcomes that arise in the TRADE EQUILIBRIUM. These outcomes describe the structure of production and consumption across dissimilar countries.

**Outcome 1.** In the trade equilibrium, all else equal, countries with higher National Product Quality pay higher wages. Basically, what separates high-income countries from low-income ones is their National Product Quality. To demonstrate this, consider two countries, \(N\) (North) and \(S\) (South) that face the same set of trade costs. The North is endowed with a higher National Product Quality, which (all else equal) implies more demand for Northern varieties. Balanced trade (equation 5), therefore, entails that the North pays higher equilibrium wages than the South:

\[ \alpha_N > \alpha_S \implies w_N > w_S \]
Outcome 2. High-wage countries have revealed comparative advantage in type $H$. To see this, note that the gravity equation (3) implies that

$$\frac{X^H_{ji}}{X^L_{ji}} = \left( \frac{\tau_{ji}w_j}{\tau_{ki}w_k} \right)^{\sigma_L - \sigma_H}.$$

Equation 6 implies that the North exports relatively more of type $H$ to $i$, whereas South exports relatively more of type $L$.

$$\frac{X^H_{Ni}}{X^L_{Ni}} = \left( \frac{w_N}{w_S} \right)^{\sigma_L - \sigma_H} > 1$$

Intuitively, North has absolute quality-advantage in both types: $\alpha_N > \alpha_S$. This results in more demand for Northern varieties, which drags up its equilibrium wage: $w_N > w_S$. As a result, the North is comparatively disadvantaged in type $L$ (which is price-sensitive), and has revealed comparative advantage in type $H$. Comparative advantage in the unified model is similar to the Heckscher-Ohlin model in that it is endogenously determined, but different in that it is regulated by demand. Specifically, quality-abundant North has comparative advantage in type $H$ for which demand is quality-intensive. Labor-abundant South, meanwhile, has comparative advantage in type $L$, which has a quantity-intensive demand.\footnote{Note that while high-income countries are net exporters of the highly-differentiated type ($H$) and net importers of the less-differentiated type ($L$), they still do export their comparative disadvantage type, $L$. This is driven by national product differentiation, and is consistent with micro-level evidence (Schott (2004)).}

These patterns of comparative advantage are robust to adding firm level heterogeneity (see appendix A) and are supported by micro-level evidence (see section 5.1).

The above view of comparative advantage is more general than the classical view, as incorporates the role production cost, product quality and product variety. In the classical view, a country has comparative advantage in a good for which it has a lower autarky relative price (Deardorff (1980)). Here, comparative advantage is determined by the relative price index, which is nominal price adjusted by quality and variety. To see this, note that the autarky relative price index of type $H$ in country $i$ is given by

$$\left( \frac{p^H_i}{p^L_i} \right)^{\text{Autarky}} = \phi \left[ \alpha_i \left( \frac{L_i}{\bar{e}_i} \right)^{1/\bar{\sigma}_i} \right]^{\frac{1}{\bar{\sigma}_L - \bar{\sigma}_H}}$$

where $\frac{1}{\bar{\sigma}_i} \equiv \frac{1}{\sigma_i} s_i + \frac{1}{\sigma_L} (1 - s_i)$ (where $s_i$ is the autarky expenditure share on type-$H$ in country $i$), and $\phi \equiv \frac{\beta_{ij}(\bar{\sigma}_i - 1)}{\sigma_L(\bar{\sigma}_H - 1)}$. Given that $\alpha_N > \alpha_S$, equation 7 entails that the autarky relative price index of
type \( H \) is lower in the North.\(^{23}\)

\[
\left( \frac{P_H}{P_L} \right)_{\text{Autarky}}^{X^H_i \gg X^L_i} < \left( \frac{P_H}{P_L} \right)_{\text{Autarky}}^{X^H_i \gg X^L_i}
\]

Therefore, North has comparative advantage in the good for which it has a lower autarky relative price index.\(^{24}\) Furthermore, the comparative advantage of the North in type \( H \) implies that (i) the North is a net exporter of type \( H \) to the South, and (ii) the North sells relatively more of type \( H \) while the South sells relatively more of type \( L \) to any country \( i \) that is located at equal distance from both:

\[
\frac{X^H_{Ni}}{X^H_{Si}} > \frac{X^L_{Ni}}{X^L_{Si}}
\]

**Outcome 3.** In the trade equilibrium, the price index of type \( H \) relative to type \( L \) is lower in high-income countries: \( \frac{\partial}{\partial x_i} \frac{p_H^{ij}}{p_L^{ij}} < 0 \). As before, I will demonstrate this with the North-South example. Similar to a (two-country) Neoclassical trade model, if North and South (countries \( N \) and \( S \)) engage in trade, the relative price index of type \( H \) rises in the North and falls in the South. However, in the presence of trade costs, prices are not equalized across countries. Specifically, the relative price index of type \( H \) remains lower in North even after trade:

\[
\left( \frac{P_H}{P_L} \right)_{\text{Free Trade}} < \left( \frac{P_H}{P_L} \right)_{\text{Free Trade}}
\]

The above result follows from the fact that the price index under trade is a weighted CES average across all international prices, with more weight assigned to the domestic price.\(^{25}\)

**Outcome 4.** All else equal, high-income countries consume relatively more of the highly differentiated, high markup type. That is, the North (\( N \)) spends relatively more on type \( H \), than the

\(^{23}\)Note that \( \frac{\partial q_i}{\partial x_i} > 0 \), which implies that \( \frac{\partial q_i}{\partial x_i} > 0 \). More specifically, an increase in National Product Quality induces consumption reallocation from type \( L \) to \( H \), which creates a greater scope for firm entry. Therefore, inequality 8 follows from the fact that \( \frac{1}{\alpha_N} < \frac{1}{\alpha_S} \) and \( \alpha_N > \alpha_S \):

\[
\alpha_N > \alpha_S \implies \phi \left[ \alpha_N \left( \frac{L_N}{\sigma_N} \right)^{\frac{1}{\eta_N}} \right] < \phi \left[ \alpha_S \left( \frac{L_S}{\sigma_S} \right)^{\frac{1}{\eta_S}} \right]
\]

\(^{24}\)Equation 8 also implies that larger economies have comparative advantage in the highly-differentiated type, a pattern highlighted in Helpman and Krugman (1985). In fact, if I shut down differences in product quality, eliminate national product differentiation (i.e. \( \eta = 1 \)), and allow for global economies of scale (\( f^* \) is paid once and for all market), the unified model would become a generalized case of Helpman and Krugman (1985).

\(^{25}\)Under free trade (\( \tau^{ij} = 1, \forall i, j \)), prices equalize in North and South:

\[
\left( \frac{P_H}{P_L} \right)_{\text{Free Trade}} = \left( \frac{P_H}{P_L} \right)_{\text{Free Trade}}
\]
South (S):
\[
\frac{X_N^H}{X_N^L} \gg \frac{X_S^L}{X_S^S}
\]

The above result is a direct consequence of outcome 3. The relative price index of type \( H \) is lower in the North. Hence, the relative consumption of type \( H \) is higher. This result emerges despite the fact that preferences are homothetic and uniform across countries. This outcome could be further generalized as follows: if type \( H \) and \( L \) exhibit some degree of substitutability and trade is costly, countries consume relatively more of the locally abundant type. I refer to this as the home-production effect on local consumption — this effect is the opposite of the “home-market effect” highlighted by Krugman (1980).\(^{26}\) While the home-production effect on consumption is backed by micro-level evidence (Atkin (2013)), no previous study (to my knowledge) has highlighted or quantified the importance of this effect for aggregate, multilateral trade flows.\(^{27}\)

2.3 The Three Modern Facts about the Composition of Aggregate Trade Flows

The four equilibrium outcomes, highlighted in section 2.2, determine the composition of aggregate trade flows. The predicted composition collectively explains three modern facts that are beyond the scope of pure gravity models. While extended gravity models have been developed to confront these facts individually, the present framework is the first to collectively explain all three facts.

**Income per capita \( \times \) trade-to-GDP ratios.** Rich countries have systematically higher trade-to-GDP ratios because they produce and consume relatively more of type \( H \), which is highly-differentiated and subject to lower effective trade costs: \( \tau_{ji}^{mH} \ll \tau_{ji}^{mL} \). Let \( \lambda^z_{ii} \equiv \frac{X_i^z}{X_i} \), denote share of domestic expenditure on type \( z \). The trade-to-GDP ratio of country \( i \) can be written as

\[
\left( \frac{\text{Trade}}{\text{GDP}} \right)_i = \left( 1 - \lambda^L_{ii} \right) \frac{X_i^L}{X_i} + \left( 1 - \lambda^H_{ii} \right) \frac{X_i^H}{X_i}
\]

In the South, consumption is dominated by type \( L \): \( \frac{X_S^L}{X_S} \approx 1 \).\(^{28}\) The effective trade costs are, however, sizable for type \( L \) entailing negligible import flows:

\[
\left( \frac{\text{Trade}}{\text{GDP}} \right)_S \approx \left( 1 - \lambda^L_{SS} \right) \approx 0
\]

\(^{26}\)The “home-market effect” states that local demand determines patterns of local production, whereas here (in face of costly trade) local production determines the structure of local consumption.

\(^{27}\)While the home-production effect on consumption is implicit in the non-homothetic gravity frameworks of Fieler (2011) and Caron et al. (2014), it cannot be separately identified from non-homotheticity in cross-sectional data. The aforementioned studies, therefore, normalize the home-production effect to focus on non-homotheticity (see appendix B). The home-production effect, however, delivers distinct predictions regrading the evolution of trade composition over time (see section 5.2).

\(^{28}\)To demonstrate how the unified model replicates differences in trade-to-GDP ratios across countries, I adopt the approximation method used in Fieler (2011).
In the North, consumption is dominated by type $H$: $\frac{X^H}{X^N} \approx 1$. The effective trade costs, however, are relatively small for type $H$, resulting in a relatively large trade-to-GDP:

$$\left(\frac{\text{Trade}}{\text{GDP}}\right)_N \approx (1 - \lambda^H_{NN}) \approx 1 - \left(\frac{\alpha^H_N}{w_N}\right)^\eta \sum_k \left(\frac{\alpha^k}{w_k}\right)^\eta$$

The above result follows from the fact that $\frac{X^H}{X^N} \approx \frac{\alpha^H}{\alpha^N}$ and $\frac{M^H}{M^N} \approx \frac{X^H_{NN}}{X^N_{NN}/w_N}$, when $\sigma_H$ approaches unity. Overall, these results indicate the trade-to-GDP ratio is systematically lower in the (low-wage) South relative to the (high-wage) North.

The above result contributes to the seminal works of Waugh (2010) and Fieler (2011), by endogenizing the systematic differences between rich and poor countries. Specifically, to explain across country differences in trade-to-GDP ratios, Waugh (2010) assumes that rich countries face exogenously lower export costs. Similarly, Fieler (2011) and Caron et al. (2014) assume that rich countries consume more of the highly-tradable types with non-homothetic preferences. In the present model, both outcomes emerge endogenously in equilibrium. Specifically, comparative advantage induces rich countries to both specialize in the production of the highly-tradable type and to consume relatively more of the highly-tradable type.

**Income per capita $\times$ the price composition of exports.** In the unified model, two factors contribute to the higher f.o.b. price of exports from high-income countries.\(^{29}\) First, high-income countries sell both types of goods ($H$ and $L$) at a higher equilibrium price, due to their higher National Product Quality:

$$\alpha_N > \alpha_S; \quad \tau_N = \tau_S \implies \frac{p^H_{Ni}}{p^H_{Si}} = \frac{p^L_{Ni}}{p^L_{Si}} = \frac{w_N}{w_S} > 1$$

This channel corresponds to quality-differentiation, and presents a standard explanation that is well-established in the literature (Schott (2004); Hallak and Schott (2011)).

The second channel, however, is novel and corresponds to across-markup specialization. That is, Northern countries specialize in the high-markup type $H$, and have a more differentiated and markup-intensive export-mix than the South. More specifically, the average price of exports from country $j$ to $i$ is

$$\bar{p}_{ji} = \left(\frac{X^H_{ji}}{X^L_{ji}}\right) p^H_{ji} + \left(\frac{X^L_{ji}}{X^H_{ji}}\right) p^L_{ji}; \quad j = N, S$$

Type $H$ exhibits a higher equilibrium price ($p^H_{ji} > p^L_{ji}$), and North ($N$) sells relatively more of type $H$ ($\frac{X^H_{Ni}}{X^N_{Ni}} > \frac{X^H_{Si}}{X^S_{Si}}$), which implies: $\bar{p}_{Ni} > \bar{p}_{Si}$. Hence, all else the equal, the share of the highly-

\(^{29}\)As noted by Schott (2004) gravity models (e.g. Eaton and Kortum (2002) and Krugman (1980)) do not accommodate within-product specialization across low and high-price varieties, which is implied by micro-level evidence. However, in a standard gravity framework, one could attribute the competitiveness of rich countries to their higher-quality rather than lower-cost. In that case, even though gravity model neglect systematic within within-product specialization, they could accommodate the higher aggregate export unit values from rich countries.
differentiated, high-price type (H) increases in a nation’s export-mix, the higher its National Product Quality. This contributes to the higher export prices of rich countries:\(^{30}\)

\[
\frac{\partial}{\partial \tau_{ji}} X_{ji}^H > 0 \implies \frac{\partial \bar{p}_{ji}}{\partial \tau_{ji}} > 0
\]

**Distance \times the price composition of exports**  According to the gravity equation trade volumes decrease with bilateral distance. However, when one decomposes volume into quantity and price, export-quantity decreases with distance whereas export-price increases (Bernard, Jensen, Redding, and Schott (2007)). The observation that countries export higher-price tradables to faraway locations is referred to as the “Washington apples” effect. Surprisingly, and despite being one of most-documented and robust patterns in trade, the “Washington apples” effect is inconsistent with mainstream gravity models (see Baldwin and Harrigan (2011)). The unified model, however, accommodates the “Washington apples” effect. In fact, the model offers a novel explanation for the effect that (unlike the standard explanation) is robust to how one formulates the trade costs.\(^{31}\)

Equation 6 implies that countries export relatively more of the highly-differentiated, high-markup-type (H) when facing larger iceberg trade costs. More specifically, all else the same, higher trade costs increase the demand for type H relative to type L exports:

\[
\frac{\partial}{\partial \tau_{ji}} X_{ji}^H > 0 \implies \frac{\partial \bar{p}_{ji}}{\partial \tau_{ji}} > 0
\]

The above effect is driven by the higher price elasticity of type L. Demand for type L is price-sensitive, which puts the high-cost exporters from distant sources at a disadvantage. As a result, geographically distant trading partners exchange relatively less of type L, and relatively more of the high-price type, H.\(^{32,33}\)

\(^{30}\)For example, consider auto exports from Korea and Germany. Suppose there are two types of cars: Luxury (high-markup) cars, and Economy (low-markup) cars. The present model predicts that Germany sells both types of cars at a higher price point due to its higher National Product Quality. Additionally, Germany sells relatively more luxury cars and Korea sells relatively more Economy cars. This additional composition effect contributes to Germany’s higher export prices relative to Korea.

\(^{31}\)The standard explanation is founded on additive (non-iceberg) trade costs, and is due to Alchian and Allen (1983).

\(^{32}\)Firms also collect higher profits from exporting type H (since type H offers a high markup). Introducing fixed exporting costs would, therefore, introduce an additional channel that contributes to the “Washington Apples” effect. That is, in the presence of fixed costs, disadvantaged exporters would generate profits only from selling type the high-profit type, H. Therefore, firms export only type H to faraway locations.

\(^{33}\)A real world example that corresponds to this effect, is auto exports from Europe. Europe exports the luxury, high-markup brands (e.g. Audi, BMW, Volvo) to the US, whereas the economy, low-markup brands (Opel, Renault, Peugeot) are not exported to the US market, but sold mostly in the local European market. Further, in Lashkaripour (2015) I provide strong evidence that the theory highlighted in this paper is a major driver of the “Washington apples” effect in US imports.
2.4 A Special Case: The Pure Gravity Model

If both types of goods are identical ($H = L = \sigma$), the present model reduces to a (one-sector) Pure Gravity Model. The gravity equation then becomes

$$X_{ji} = \frac{\alpha_j M_j^{1-\eta} (\tau_j w_j)^{1-\sigma}}{\sum_{k \in C} \alpha_k M_k^{1-\eta} (\tau_k w_k)^{1-\sigma}} X_i$$  \hspace{1cm} (9)

The pure gravity model, characterized by equation 9, nests the ARMINGTON MODEL. Specifically, if firms are perfectly competitive ($\eta \to \infty$, and $f^e = 0$), equation 9 reduces to the standard Armington gravity equation:

$$X_{ji} = \frac{\alpha_j (\tau_j w_j)^{1-\sigma}}{\sum_{k \in C} \alpha_k (\tau_k w_k)^{1-\sigma}} X_i$$

Contrary to the unified model, the pure gravity model does not systematically characterize the composition of aggregate trade flows. As a result, the pure gravity model generates three counterfactual patterns: First, in the pure gravity model, high-income countries have lower trade-to-GDP ratios (high-income countries have a lower equilibrium effective wage, $w_i^{e-1}/\alpha_i$, which makes them globally more competitive, but also less likely to import). Second, in the gravity model, aggregate export flows from all countries are counterfactually subject to the same trade elasticity, $\sigma - 1$. Third, bilateral distance has no systematic effect on the price-mix of exports — the pure gravity model predicts that countries sell the same type with the same f.o.b. price across all locations. Additionally, the gravity model provides no scope for systematic (within-product) specialization across low and high-price types, which is inconsistent with micro-level evidence.

3 Mapping the Model to Data

This section estimates the structural parameters of the model by fitting it to data on bilateral trade volumes and per capita income across 100 countries. To demonstrate the merits of the unified model, I compare it to a pure gravity model that is fitted to the same data. The unified model matches the trade volumes remarkably better than the pure gravity model. Furthermore, the unified model (in contrast to the pure gravity model) correctly predicts the price of composition

---

$^{34}$Another special case arises when $\eta = 1$. In this case, National Product Differentiation is eliminated, and the only remaining force behind trade is comparative advantage across types. Suppose there are a continuum of types, $z \in [0, 1]$, and two countries: North (N) and (South). In that case, the model becomes analogue to Dornbusch, Fischer, and Samuelson (1977). Specifically, let North have a higher national product quality: $\mu_N > \mu_S$. There exist two cut-offs, $\overline{\sigma}$ and $\overline{\sigma}$, such that North is the sole producer (and exporter) of type $z$ if $\sigma < \overline{\sigma}$; South is the sole producer (and exporter) if $\sigma > \overline{\sigma}$; and type $z$ is not traded if $\overline{\sigma} < \sigma < \overline{\sigma}$.

$^{35}$Note that most multi-sector or multi-type gravity frameworks do not predict a clear pattern of specialization across low and high-price types. In Fieler (2011), for example, while rich countries export more technologically differentiated types, there is no direct mapping between the level of technology differentiation and price. Studies that do make systematic predictions regarding the price composition of exports, generally highlight the role of specialization across quality types (e.g. Fajgelbaum et al. (2011))
of exports across exporters and over space. I then perform a counter-factual analysis to estimate the gains from trade in both models. The estimated gains from trade are substantially larger in the unified model, and more unequally distributed across nations. Overall, these results highlight the quantitative significance of the composition margin.

**Data.** I use data on bilateral merchandise trade flows in 2000 from the U.N. Comtrade database (Comtrade (2010)). The data on population and GDP, and the price of tradables are from the World Bank database (World-Bank (2012)). The sample consists of the 100 largest economies (in terms of real GDP), which account for more than 95% of the world trade in 2000. Data corresponding to country pairs (distance, common official language, and borders) are compiled by Mayer and Zignago (2011).

### 3.1 Estimation Strategy

This section describes the estimation procedure and addresses identification issues. Equation 3 characterizes the bilateral trade flows for each type of good. We can, therefore, calculate total flows from country \( j \) to \( i \) as

\[
X_{ji} = X^H_{ji} + X^L_{ji}
\]

where \( X^H_{ji} \) and \( X^L_{ji} \) are given by equation 3. Aggregate trade flows depend on the number of exporting firms \( \{M_{ij}\}_{i,j \in C} \), population \( \{L_i\}_{i \in C} \), wage \( \{w_i\}_{i \in C} \), National Product Quality \( \alpha \equiv \{\alpha_i\}_{i \in C} \), iceberg trade costs \( \tau \equiv \{\tau_{ji}\}_{j,i \in C} \), and demand parameters \( \sigma_L, \sigma_H, \epsilon, \) and \( \eta \).

I take populations \( L_i \) and wages \( w_i \) from the data; solve for a vector of National Product Qualities, \( \alpha \), that imply wages consistent with the data; solve for the equilibrium number of firms \( M_{ji} \) that satisfy the free entry condition; and estimate \( \tau, \sigma_L, \sigma_H, \epsilon, \) and \( \eta \). The procedure is the following:

1. I parametrize the iceberg trade costs as follows:

\[
\tau_{ji} = 1 + \left[ \kappa_{const} + \kappa_{dist} \operatorname{dist}_{ji} \right] \kappa_{border} \kappa_{lang} \kappa_{agreement}
\]

where \( \operatorname{dist}_{ji} \) denotes the distance (in thousands of kilometers) between countries \( j \) and \( i \). \( \kappa_{border} \) is one if countries do not share a border, and an estimated parameter otherwise. \(^{37}\) Similarly, \( \kappa_{agreement} \) and \( \kappa_{lang} \) are one if a country pair do not have a trade agreement or a common-language, and estimated otherwise. Altogether, \( \kappa \equiv \{\kappa_{border}, \kappa_{lang}, \kappa_{agreement}, \kappa_{const}, \kappa_{dist}\} \) denotes the vector of parameters describing the iceberg trade costs. For a given \( \kappa \), and data on distance, trade agreements, common-language and borders, I can construct a matrix of iceberg trade costs.

\(^{36}\)The entry cost parameter, \( f^e \), governs the scale of entry and is normalized to one. The normalization does not affect trade flows, but normalizes the mass of firms in each markets. Putting it differently, \( f^e \) cannot be identified from trade flows; it can be identified with data on the number of firms.

\(^{37}\)For example, if \( \kappa_{border} \) is, say, 0.9, sharing a border reduces \( \tau_{ji} = 1 \) by 10%.
ii. Given parameters \{κ, α, σ_L, σ_H, ε, η\} and data (D) on the wages, population, distance, trade agreements, common languages and borders I can solve for the mass of firms using the \textit{free entry condition} (equation 4):

\[ M_{ji} = M_{ji}(D; κ, α, σ_L, σ_H, ε, η), \quad i, j ∈ C \]

iii. Given \(M \equiv \{M_{ji}\}_{i,j∈C}\) from the previous step, parameters \{κ, σ_L, σ_H, ε, η\}, and data (D), I solve for a vector of \textit{National Product Qualities}, \(α\), that satisfy the balanced trade condition (equation 5):

\[ α_j = α_j(D; M, κ, σ_L, σ_H, ε, η), \quad j ∈ C \]

That is, \(α_j\) is chosen so that the market clearing wage equals data on GDP per capita.

iv. For any set of parameters, \{κ, σ_L, σ_H, ε, η\}, and data, D, I iterate over steps 2 and 3 to find an implicit solution for \(α_j\) and \(M_{ji}\). Using the implicit solution, I calculate bilateral trade flows \(X_{ji}\) from equation 10, and the matrix of trade shares: \(λ_{ji} = X_{ji} / X_j\). The gravity equation in stochastic form becomes

\[ \ln λ_{ji} = g(D; κ, σ_L, σ_H, ε, η) + ε_{ji} \quad (11) \]

The above equation indicates that trade shares (\(λ_{ji}\)) are a function of data, D, the estimated parameters, \{\(κ_{\text{border}}, κ_{\text{lang}}, κ_{\text{agreement}}, κ_{\text{const}}, κ_{\text{dist}}, σ_L, σ_H, ε, η\}\}, and the error term \(ε_{ji}\). I estimate equation 11 by minimizing the residual sum of squares (Non-linear Least Squares (NLLS)). Anderson and Van Wincoop (2003) argue that the NLLS estimator is unbiased if \(ε\) is uncorrelated with the derivative of \(g(\cdot)\) with respect to D. This would be the case if \(ε\) represents measurement errors.

**Identification of parameters.** The vector \(κ = \{κ_{\text{border}}, κ_{\text{lang}}, κ_{\text{agreement}}, κ_{\text{const}}, κ_{\text{dist}}\}\) determines the size of the iceberg trade costs. More specifically, \(κ_{\text{const}}\) governs the overall level of home bias, while \(κ_{\text{dist}}\) governs the effect of bilateral distance on bilateral trade flows. \(η\) governs the scope for \textit{national product differentiation}, and could be identified from iceberg trade costs given our proxy for national wage.\(^{38}\) More specifically, A larger \(η\) induces countries to diversify their imports, whereas a lower \(η\) induces countries to concentrate their imports on more competitive partners with lower quality-adjusted wages.\(^{39}\) The type-specific trade elasticities (\(σ_H\) and \(σ_L\)) are not jointly identified from \(η\). However, if we set \(σ_L = 6\), we can separately identify \(σ_H\). The relative magnitude of \(σ_L\) to \(σ_H\) governs the differential effects of distance on export flows from rich and poor countries. Parameter \(ε\) (the elasticity of substitution across types) governs the size of the \textit{home-production effect}

\(^{38}\)\(η\) is analogous to the trade elasticity in a standard gravity model. As noted in Head and Mayer (2014) (chapter 4.2), we could separately identify the trade elasticity by using data on wage or productivity. Alternatively, one could normalize the trade elasticity, and infer the importer/exporter fixed effects (i.e. price indexes) from the structural model — this approach is adopted by Anderson and Van Wincoop (2003).

\(^{39}\)When \(η = 1\), countries import each type from the most competitive supplier.
on local consumption, \( \epsilon \), therefore, regulates the systematic differences in consumption structure and trade-to-GDP ratios between rich and poor countries.

### 3.2 Estimation Results

The estimation results are presented in table 1. The first column reports the estimation results for the **unified model**. Column two reports estimation results corresponding to the **pure gravity model**, a special case of the unified model. Column three reports estimation results for the **Armington model**, a special case of the pure gravity model. When estimating the unified model I normalize \( \sigma_L \) (the elasticity of substitution for the type \( L \)) to 6. When estimating the pure and restricted gravity models, I normalize the trade elasticity to 4.97: the trade-weighted average of \( \sigma_H \) and \( \sigma_L \) from the unified model.

In both the unified and the pure gravity models, countries diversify their imports due to *national product differentiation*. However, in the unified model comparative advantage induces countries to concentrate their imports relatively more (and on dissimilar partners). Therefore, to match the observed levels of import diversification, the unified model estimates a greater scope for *national product differentiation*. In the Armington (restricted gravity) model, the scope for *national product differentiation* is assumed to be complete (i.e. \( \eta \to \infty \)). Thus, bilateral trade costs are the only force preventing countries from fully diversifying their imports. As a result, the Armington model over-estimates the distance elasticity to match the factual levels of import diversification.

With only two additional degrees of freedom, the unified model has an \( R^2 \) that is 43 percent higher than the pure gravity model (table 1). The improved fit of the unified model comes from matching two quantitatively important margins: (i) the higher trade-to-GDP ratios of rich countries and (ii) the higher distance elasticity of exports from poor nations. Note that these two margins could be fixed with alternative models. The unified model is, nevertheless, distinct in several aspects. First, it improves the fit of the gravity model with fewer degrees of freedom.\(^{40}\) Second, unlike alternatives, the unified model accommodates the factual rise of North-South trade over time (see section 5.2). Finally, the superior fit of the unified model is not confined to *trade volumes*, and extends to salient out-of-sample patterns regarding export *unit values*.

**Income per capita \( \times \) trade-to-GDP ratio:** Figure 2 displays the relationship between trade-to-GDP and per capita income in the data, and compares it to the ones implied by the unified model and the pure gravity model. The gravity model counterfactually predicts that trade-to-GDP drops with income per capita.\(^{41}\) This feature makes the gravity model inapt for analyzing trade between

\(^{40}\)For example, non-homothetic gravity models (Fieler (2011) and Caron et al. (2014)) could explain the higher trade-to-GDP of rich countries. While these models require more degrees of freedom, they do not nest the model estimated in this paper (see appendix B).

\(^{41}\)The trade-to-GDP ratios are also on average lower in the estimated pure gravity model. This is due to the pure gravity model’s inability to accommodate the small volume of South-South trade. Therefore, with many poor countries in the sample, it compensate for this shortcoming by estimating a large degree of home bias for all countries.
rich and poor countries — especially that the gains from trade are determined primarily by trade-to-GDP ratios. The unified model, though, correctly predicts the positive relationship between trade-to-GDP and per capita income. This prediction is driven by South-South trade being systematically smaller than North-North trade. A non-homothetic gravity model would also deliver this prediction. However, non-homotheticity would over-state the overall volume of dissimilar-dissimilar trade (North-North + South-South) compared to the unified model (figure 3).

**Income per capita × trade elasticity:** Figure 4 plots the normalized export flows $\frac{X_{ji}}{X_i X_j}$ against bilateral distance $\text{dist}_{ji}$. Exporters are divided in to two groups: North (the richest 50 countries) and South (the poorest 50 countries). In the data, export flows from the North are significantly less sensitive to distance compared to export flows from the South.\(^{42}\) The unified model correctly captures this regularity (In the unified model trade elasticities are country-specific and *endogenously determined* by across-type specialization). The pure gravity model, by contrast, counter-factually predicts a similar elasticity for North and South. Given that trade elasticities govern the size of the welfare gains from trade, this is an undesirable feature stemming from the fact that the pure gravity model overlooks the composition margin.

**Out-of-sample performance:** When estimating the unified model, I exclusively targeted trade volumes. Therefore, the estimation-fit does not fully manifest the merits of the unified model. Another advantage of the unified model is that it pins down a systematic relationship between geography, income and the price composition of trade, which I discuss below:

i. **Income per capita and the price composition of exports:** high-income countries export higher price goods within narrowly defined categories, and have higher aggregate export-prices. This pattern is mostly driven by within-product specialization across low- and high-price varieties (Schott (2004)). Standard gravity models do not accommodate this pattern, as they offer no scope for (systematic) specialization.\(^{43}\) The unified model, however, allows for systematic specialization and predicts that rich countries export relatively more of the high-price (high-markup) type. Figure 5 illustrates the (stand-alone) effect of international specialization on aggregate export prices, in the estimated model.

\(^{42}\)To formally illustrate the (mediation) effect of income per worker on trade elasticities, I can run a conventional gravity regression on my sample of 100 countries. Specifically, I allow for interaction between bilateral distance and the exporter’s income per worker:

$$\ln X_{ji} = \left( \frac{-3.26 \pm 0.20 \ln w_j}{0.15 \pm 0.02} \right) \ln \text{DIST}_{ji} + S_j + M_i + \epsilon_{ji}$$

where $S_j$ and $M_i$ are exporter and importer fixed effects. The robust standard errors are reported in the parenthesis, and the $R^2$ is 0.47. The results confirm that export flows from rich countries are significantly less sensitive to distance.

\(^{43}\)Standard Gravity models, predict that high-income countries have lower export prices unless one attributes competitiveness to product quality (rather than productivity). In comparison, the unified model predicts that high-income countries have higher export prices due to: (1) their higher *National Product Quality*, which is reflected in their higher wage, and (2) their markup-intensive exports.
ii. **The “Washington apples” effect**: Countries tend to export higher-priced tradables to far-away markets. This pattern is inconsistent with standard gravity models (see Baldwin and Harrigan (2011)).\(^{44}\) In the unified model, on the other hand, countries export the highly-differentiated, high-markup type \(H\) relatively more to far-away locations. This gives rise to the “Washington apples” effect. Figure 6 illustrates this result for US and German exports: the estimated model predicts that the (aggregate) f.o.b. price of exports increase significantly with distance to a given market.

## 4 The Gains from Trade

This section compares the gains from trade implied by the unified model to those implied by the pure gravity model. While the pure gravity model accommodates only the standard gains from product variety, the unified model integrates the gains from product variety with the gains from comparative advantage. My counterfactual analysis reveals that: (i) the gains from trade are more than four-times larger in the unified model, (ii) the gains from comparative advantage are systematically biased in favor of disadvantaged (i.e. poor and remote) nations, and (iii) trade liberalization could decrease international income inequality by around 10%.

### 4.1 The realized gains from trade

In the pure gravity model, when \(\eta\) is sufficiently large, the gains from trade relative to autarky are determined solely by the intensity of trade. Specifically, let \(\lambda_{ii}\) denote the share of domestic consumption in country \(i\)'s GDP \((1 - \frac{\text{trade}}{\text{GDP}})\). The gains from trade relative to autarky in the gravity model are given by

\[
\text{Gains}_i = \lambda_{ii}^{-\frac{1}{2}}
\]

where \(e = \sigma - 1\) is the trade elasticity, and common across countries. Arkolakis et al. (2012) show that the above equation characterizes the gains from trade across a wide range of workhorse gravity models. By contrast, in the unified model the gains from trade not only depend on the intensity of trade, but also the composition of aggregate trade flows. In fact, trade elasticities are country-specific and endogenously determined by the composition of a nation’s imports. When \(\eta\) is sufficiently large, the gains from trade in the unified model are given by

\[
\text{Gains}_i = \left[ s_i \left( \frac{H}{\lambda_{ii}} \right)^{-\frac{e-1}{s-1}} + (1 - s_i) \left( \frac{L}{\lambda_{ii}} \right)^{-\frac{e-1}{s-1}} \right]^{\frac{1}{e}}
\]

\(^{44}\)As Baldwin and Harrigan (2011) argue, in heterogeneous gravity models (e.g. Chaney (2008); Eaton and Kortum (2002)) bilateral distance lowers the f.o.b. price of exports. In homogeneous gravity models (e.g. the Armington model), f.o.b. export prices are the same across all export destinations.
where \( s_i \equiv \frac{\alpha_i^{H-1}}{\left(\alpha_i^{H-1} + \alpha_i^{\ell-1}\right)} \) denotes the autarky expenditure share on type \( H \) in country \( i \). Equation 12 indicates that the gains from trade in the unified model are a weighted average of the gains across the two types, with the gains being considerably larger for the highly-differentiated type, \( H \). The overall gains from trade are, therefore, determined primarily by the term \( s_i \left( \frac{H}{H_{ii}} \right)^{\frac{1}{s_{ii}-1}} \). These overall gains are sizable due to the inclusion comparative advantage-driven trade. The gains are large for rich countries as they consume relatively more of type \( H \), which corresponds to a high \( s_i \). For poor countries, the gains are large as these countries are net importers of type \( H \), which corresponds to a high \( \lambda_{ii}^{H} \).

The pure gravity model, in comparison, systematically understates the gains from trade for all countries. For rich countries the gravity model systematically under-states the gains, since it counter-factually asserts that rich countries have lower trade-to-GDP ratios. The gravity model systematically understates the gains for poor countries as it forces their imports to have the same elasticity as rich countries. From an alternative perspective, gravity models understate the gains from trade by overlooking the gains from production specialization or comparative advantage.\(^{45}\)

To quantify the importance of the composition margin, I estimate the gains from trade in the unified model and compare them with the gains implied by the gravity model. To this end, I compare the factual real wage with the counter-factual autarky real wage in both models. I solve for the counter-factual real wages by resolving the general equilibrium with no trade. The estimated gains are compared in figure 7. A summary of the estimated gains is reported in table 2. The gains from trade are more than four-times larger in the unified model and more unequally distributed across countries. These results suggest that the composition margin is central to our understanding of the gains from trade. This is compelling given that accommodating margins such as firm heterogeneity do not magnify the gains from trade to comparable degrees (Arkolakis et al. (2012)).

To highlight the economic significance of the above results, I would like to emphasize two points. First, previous studies have pointed out that allowing for multiple sectors magnifies the gains from trade (Costinot and Rodríguez-Clare (2013) and Ossa (2012)). These studies, however, do not account for systematic specialization across sectors and endogenously fix the consumption share on the heavily imported sectors.\(^{46}\) In this paper, I demonstrate that the gains from trade are magnified, even when production specialization and consumption shares (across sectors) are endogenously determined. Second, the results above do not suffer from the critique of Costinot and Rodríguez-Clare (2013). Specifically, Costinot and Rodríguez-Clare (2013) argue that to quantify the contribution of an added margin to the gains from trade, one should compare the full model to a benchmark, which is independently estimated. Here, I adopt this exact approach to quantify

\(^{45}\)Eaton and Kortum (2002) derive a gravity equation built on Ricardian comparative advantage. Their framework, however, does not deliver systematic patterns of production specialization. That is, their theory does not pin down a relationship between the characteristics of a country and the type of goods they have comparative advantage in.

\(^{46}\)These issues have raised skepticism regarding the robustness of these results. Costinot and Rodríguez-Clare (2013) demonstrate that replacing the across-sector Cobb-Douglas utility aggregator with a CES aggregator tremendously reduces the role of multiple sectors. Here, I argue that allowing for systematic production specialization (across-sectors) tremendously magnifies the gains from trade even in the presence of a CES utility aggregator.
the welfare contribution of the composition margin.

4.2 The gains from comparative advantage

The previous section demonstrated that the gains from trade are four-times larger in a model that accommodates comparative advantage. The analysis, however, relied on the gains quantified using two independently estimated models. These estimated gains are, therefore, contaminated with estimation issues. To quantify the pure benefits of comparative advantage, I estimate the losses (or gains) from eliminating comparative advantage in the unified model. The results indicate that the average country’s welfare improves by 28% when one accommodates comparative advantage-driven trade. More importantly, the gains from comparative advantage are systematically biased in favor of poor and remote nations (figures 8 and 9). Overall, these findings suggest that incorporating comparative advantage into a pure gravity model not only magnifies the gains from trade; but shifts them in favor of disadvantaged nations.

4.3 Trade Liberalization and International Income Inequality

A key implication of the unified model is that trade liberalization could reduce international income inequality. The inequality-reducing effects of trade are driven by comparative advantage, and are distinct from previously emphasized channels. Specifically, existing studies (most notably Waugh (2010)) have argued that poor countries face systematically larger export costs, and that eliminating these asymmetries could reduce international income inequality. The unified model predicts that trade liberalization could reduce international income inequality, even when trade costs are symmetric across rich and poor countries. Specifically, the unified model asserts that poor countries are affected more severely by (symmetric) export costs as they specialize in price-elastic goods with low degrees of tradability. Eliminating international trade costs will, therefore, make poor countries relatively more competitive and will drag up their wage relative to rich countries. Figure 10 illustrates this effect. The pure gravity model, in comparison, predicts that trade liberalization has only a weak effect on international income inequality.

5 Micro-level Evidence and Alternative Theories

5.1 Evidence from product-level data

So far I have evaluated the unified model with data on aggregate trade flows. This section contrasts the prediction of the model directly with micro-level data. The unified model predicts that,

\(^{47}\)I eliminate comparative advantage by counter-factually forcing both types to provide the same scope for product differentiation, which is equal to the trade-weighted average of \(\sigma_L\) and \(\sigma_H\).
all else equal, high-income countries export relatively more of the high-markup, low-\(\sigma\) goods to a given market. That is, the export-mix from high-income countries should have a higher markup content than that of the low-income countries. I verify this prediction using product-level US import data.\(^{48}\) For US imports, Broda and Weinstein (2006) have estimated the scope for product differentiation (\(\sigma_z\)) for various 10-digit product categories.\(^{49}\) Using their estimates and product-level import data, I can calculate the average markup (\(\frac{\sigma}{\sigma + 1}\)) embedded in the exports of a country to the US. Figure 11 plots the markup content of exports against the (average) income per capita of an exporter during the period of 1989 to 2011. The graph clearly supports the prediction of the unified model that high-income countries export relatively more high-markup goods. The second test I perform is similar to the one conducted in Hanson and Xiang (2004). Specifically, I look at the relative export share of Northern countries to Southern countries across various product categories.\(^{50}\) Consistent with my theory, I find that the export share of Northern countries is systematically higher in more differentiated, low-\(\sigma\) product categories (table 3).

### 5.2 The transformation of trade: the home-production effect vs. non-homotheticity

Many studies (e.g. Flam and Helpman (1987), Stokey (1991), Matsuyama (2000)) have underlined the role of non-homothetic preferences in explaining trade among rich and poor countries. Markusen (1986) demonstrated that non-homotheticity could explain why, in the 1980s, North-North trade dominated North-South trade. Recently, Fieler (2011) and Caron et al. (2014) have extended this idea to explain the higher trade-to-GDP ratio of rich countries. The underlying thesis is that rich countries, by assumption, demand more of the goods for which other rich countries have comparative advantage in. The unified model, meanwhile, highlights an alternative explanation for these differences: the home-production effect on local consumption. While the home-production effect and non-homotheticity cannot be separately identified in cross-sectional data (see appendix B), they offer distinct predictions about the evolution of trade (in response to lower trade costs) over time. Non-homotheticity gives rise to a Linder effect that would persist even under free trade — that is, even with costless trade, rich countries concentrate their imports on goods that are exported by other rich countries. The unified model, by contrast, predicts that countries spend relatively more on their comparative advantage type, only if trade costs are sufficiently large. As trade is liberalized, countries diversify their consumption and import relatively more from dissimilar partners. Hence, the Linder effect becomes significantly weaker with lower trade costs.\(^{51}\)

---

\(^{48}\) The product-level US import data is compiled by Schott (2008), and publicly available.

\(^{49}\) See Soderbery (2015) for updated estimates.

\(^{50}\) I use the North-South categorization employed in Romalis (2004).

\(^{51}\) The unified model delivers the Linder effect when trade costs are sufficiently large. Intuitively, rich countries have incentive to import type \(L\) from poor countries due to comparative advantage, but are discouraged by high trade costs given that type \(L\) is price-sensitive. Hence, when trade costs are sizable, rich countries spend mostly on type \(H\), and their imports are mostly from other rich countries and driven by national product differentiation (figure 12). Figure 13 displays a crude illustration of the Linder effect in the estimated model (the illustration is crude as it does not control for geography).
The transformation of *trade composition* over time closely resembles the predictions of the unified model. Hanson (2012) shows that in the past decade, trade liberalization has brought about hyper-specialization with North-South trade growing dramatically relative to North-North trade. Specifically, by 2008 (in stark contrast to 1998) developed countries where sourcing most of their imports from developing economies.\(^{52}\) Krugman (2009), too, points to a similar pattern stating that “in 2006, for the first time, the United States did more trade in manufactured goods with developing countries than with other advanced nations.” Figure 12 illustrates this transformation from 1986 to 2006.\(^{53}\) The rise of North-South trade is more than a mere scale effect. The 2013 WTO world trade report (WTO, 2013) indicates that from 1980 to 2011 the share of developing countries in global trade grew by 45%, while their GDP-share grew by only 27%. At the same time, the average income gap between rich countries and their direct trading partners has increased significantly, suggesting that the rise of North-South trade is not solely driven by poor countries becoming richer (Subramanian and Kessler (2013)).\(^{54}\)

As illustrated in Figure 12, the unified model could explain this transformation. Specifically, The unified model predicts that when trade costs are sufficiently large, the 20-richest countries source the majority of their imports from other rich countries. As one counter-factually lowers the trade costs, the *home-production effect* on consumption becomes weaker, and countries diversify their imports. Consequently, North-South trade (driven by comparative advantage) dominates North-North trade. Purely non-homothetic models, by contrast, do not allow for the same level of diversification. That is, even with free trade, rich countries would consume more of the income-elastic good which is exported by other rich countries.\(^{55}\) Similarly, multi-sector gravity models that exogenously fix sectoral expenditure shares do not accomoated this transformation.\(^{56}\)

---

\(^{52}\) Hallak (2010) shows that even in 1995 the Linder effect did not hold for *aggregate* trade flows, but held only when formulated as a sector-level prediction.

\(^{53}\) Taking all directions of trade into account, the trend has been a rise of similar-similar (N-S) trade relative to dissimilar-dissimilar (N-N + S-S) trade. Specifically, the share of dissimilar-dissimilar trade in world trade has grown by more than 15% since 1980 (the 2013 WTO world trade report (WTO, 2013)).

\(^{54}\) The hypothesis that North-South trade has risen due to poor economies becoming relatively richer has two further caveats: First, international income inequality has increased between 1980 and 2010 — mostly due to declining incomes in Latin America during the ‘lost decade’ of the 1980s and the prolonged economic implosion of countries in sub-Saharan Africa (Schneider, Buehn, and Montenegro (2013)). Second, if dissimilarities between North and South were diminishing, it would have two opposing effects: (i) Southern countries would demand more skill-intensive goods, which encourages North-South trade and (ii) The production structure in the South would become more similar to that of the North, reducing the scope for comparative advantage-driven North-South trade.

\(^{55}\) The non-homothetic model estimated in Fieler (2011) implies that the richest countries spend close to a 100% of their income on the good for which other rich countries have comparative advantage in (see Figure 6 in Fieler (2011)). Since these consumption shares are regulated primarily by non-homotheticity, they would not transform substantially in response to lower trade costs. Therefore, the richest countries would source their imports mostly from other rich countries even when trade cost are relatively small.

\(^{56}\) Caliendo and Parro (2014) and Costinot et al. (2012), for example, assume exogenously fixed expenditure shares across sectors.
6 Conclusion

I argue that both the volume and the composition of aggregate trade flows should be taken into account when quantifying the welfare gains from trade. To this end, I propose a simple model that reconciles the gravity equation with three basic facts about the composition margin. The composition margin is regulated by a novel view of comparative advantage that is backed by micro-level evidence. I estimate the unified model and compare it to a restricted case of the model that delivers the standard gravity equation without the composition margin. The comparison implies that accommodating the composition margin and comparative advantage-driven trade magnifies the (average) gains from trade by a factor of 4. Moreover, the gains from comparative advantage are systematically pro poor and remote countries. These findings have direct implications for a vibrant literature that studies multilateral trade agreements.

Two aspects of the unified model merit further research. In the unified model National Product Quality determines the comparative advantage of a nation and the composition of its exports. The model, however, takes no stance on the origins of National Product Quality. One could extend the unified model so that National Product Quality is endogenously determined by the capital- or skill-abundance of a country. Second, the unified model can be extended to allow for intermediate trade and multi-national production. Such extensions could provide a tractable framework for studying those phenomena across dissimilar countries. Existing models of intermediate trade and multi-national production are generally implemented within standard gravity frameworks and, therefore, miss out on principal moments when applied to trade between rich and poor countries.

References


Appendix

A Firm-level heterogeneity

This section demonstrates that the main predictions of the unified model are robust to adding firm-level heterogeneity. The analysis closely follows Chaney (2008). Specifically, suppose that the mass of firms in country $i$ is fixed to $M_i$. Firms are heterogeneous in quality, $\varphi$, which is drawn (independently) from a country-specific Fréchet distribution:

$$G_i(\varphi) = 1 - \alpha_i \varphi^{-\gamma}$$

Here, $\alpha_i$ denotes the overall quality of firms/varieties from country $i$. Moreover, firms should pay a local entry cost $f^e$ for each type they want to sell in a given market.\(^{57}\) Type-specific bilateral

\(^{57}\)The entry cost could also be paid for both types and the main outcomes would still be the same — see footnote 58.
trade flows are, therefore, given by

$$X_{ji}^z = M_j^{\frac{1}{\gamma}} \left( \frac{\bar{\sigma}_z}{\bar{\sigma}_z - 1} \frac{\tau_{ji}w_j}{p_i^z} \right)^{1-\sigma_z} \left( \int_{\varphi_{ji}^{z \ast}}^{\infty} \varphi dG_j(\varphi) \right) X_i^z \tag{13}$$

where $\varphi_{ji}^{z \ast}$ denotes the lowest quality of type $z$ that would be exported from country $j$ to $i$. $\varphi_{ji}^{z \ast}$ is pinned down by the zero cut-off profit condition:

$$\varphi_{ji}^{z \ast} \left( \frac{\bar{\sigma}_z}{\bar{\sigma}_z - 1} \frac{\tau_{ji}w_j}{p_i^z} \right)^{1-\sigma_z} X_i^z = w_i f^e \implies \varphi_{ji}^{z \ast} = \left( \frac{\bar{\sigma}_z}{\bar{\sigma}_z - 1} \frac{\tau_{ji}w_j}{p_i^z} \right)^{\sigma_z - 1} \frac{\bar{\sigma}_z f^e}{\lambda_i^2 L_i}$$

The average quality of type $z$ exports from country $j$ to $i$ would, therefore, be:

$$\left( \int_{\varphi_{ji}^{z \ast}}^{\infty} \varphi dG_j(\varphi) \right) = \frac{\gamma}{\gamma - 1} \left( \varphi_{ji}^{z \ast} \right)^{1-\gamma} = \frac{\gamma}{\gamma - 1} \alpha_j \left( \frac{\bar{\sigma}_z}{\bar{\sigma}_z - 1} \frac{\tau_{ji}w_j}{p_i^z} \right)^{(\sigma_z - 1)(1-\gamma)} \left( \frac{\bar{\sigma}_z f^e}{\lambda_i^2 L_i} \right)^{1-\gamma}$$

Plugging the above expression into the trade flow equation (13) would give us:

$$X_{ji}^z = \alpha_i M_j^{\frac{1}{\gamma}} \frac{\gamma}{\gamma - 1} \left( \frac{\bar{\sigma}_z}{\bar{\sigma}_z - 1} \frac{\tau_{ji}w_j}{p_i^z} \right)^{\gamma(1-\sigma_z)} \left( \frac{\bar{\sigma}_z f^e}{\lambda_i^2 L_i} \right)^{1-\gamma} X_i^z$$

The above equation combined with the market clearing condition implies that

$$X_{ji}^z = \frac{\alpha_i M_j^{\frac{1}{\gamma}} \left( \tau_{ji}w_j \right)^{\gamma(1-\sigma_z)}}{\sum_{k=1}^{N} \alpha_k M_k^{\frac{1}{\gamma}} \left( \tau_{ki}w_k \right)^{\gamma(1-\sigma_z)}} X_i^z \tag{14}$$

The above gravity equation, like the benchmark case, implies that trade flows associated with the highly-differentiated type are less sensitive to trade costs. Further, equation 14 implies that rich countries have comparative advantage in the highly-differentiated type, $H$. To see this, note the relative exports of type $H$ is given by: \(^{58}\)

$$\frac{X_{ni}^H}{X_{ni}^L} / \frac{X_{si}^H}{X_{si}^L} = \left( \frac{\tau_{ni}w_n}{\tau_{si}w_s} \right)^{\gamma(\sigma_l - \sigma_H)}$$

\(^{58}\text{If entry costs were paid per type, then we would have a common quality cut-off, } \varphi_{ji}^\ast, \text{ and a common average quality, } \int_{\varphi_{ji}^\ast}^{\infty} \varphi dG_j(\varphi), \text{ for both types. This would imply that for any country } j \text{ exporting to } i:\n
$$\frac{X_{ji}^H}{X_{ji}^L} = \left( \frac{p_{ji}^H}{\bar{p}_i} \right)^{\sigma_H - 1} \left( \tau_{ji}w_j \right)^{\sigma_l - \sigma_H} X_{ji}^H / X_{ji}^L \implies \frac{X_{ni}^H}{X_{ni}^L} / \frac{X_{si}^H}{X_{si}^L} = \left( \frac{\tau_{ni}w_n}{\tau_{si}w_s} \right)^{\sigma_l - \sigma_H}$$

The above equation would give rise to the same revealed pattern of comparative advantage. That is, high-wage countries would export relatively more of type $H$. 

32
Hence, all else equal, the North will export relatively more of type $H$:

\[
\alpha_n > \alpha_s, \quad \tau_{n\bar{i}} = \tau_{s\bar{i}} \forall \bar{i} \implies \begin{cases} w_n > w_s \\ \frac{X_{n\bar{i}}}{X_{s\bar{i}}} > \frac{X_{n\bar{i}}^{H}}{X_{s\bar{i}}^{H}} \forall \bar{i} \end{cases}
\]

The other claims of the paper will follow once the above pattern of comparative advantage is established.

**B The estimated model versus non-homothetic alternatives**

The unified model corrects the standard gravity model along two distinct dimensions: (i) it accommodates the systematic differences in trade volumes across rich and poor countries, and (ii) it accommodates the systematic variations in export unit values across exporters and over space. If I shut down the second dimension, the unified model will reduce to a restricted case with perfect competition, which is also a special case of the non-homothetic models in Fieler (2011) and Caron et al. (2014).\(^{59}\) My estimation, though, has no overlap with the aforementioned models. More specifically, the estimation in Fieler (2011) and Caron et al. (2014) normalizes the home-production effect on consumption and focuses instead on non-homotheticity. My estimation, by contrast, focuses exclusively on the home-production effect, and demonstrates that this channel performs equally well in matching cross-sectional data.\(^{60}\) The following, lays out this argument in detail.

Consider Fieler (2011) with two types of goods: type 1 and 2. The relative expenditure on type 1 in country $i$ is given by

\[
\frac{X^1_i}{X^2_i} = \lambda^{\sigma_2 - \sigma_1} \left( \frac{\alpha_1 (p^1_i)^{1 - \sigma_1}}{\alpha_2 (p^2_i)^{1 - \sigma_2}} \right) \tag{15}
\]

Moreover, for each type $z = \{1, 2\}$ the trade shares are given by

\[
\frac{X^z_{ji}}{X^z_{ii}} = \frac{T_j (\tau_{ji} w_j)^{-\theta_z}}{T_i (\omega_i)^{-\theta_z}} \tag{16}
\]

The perfectly competitive case of the unified model (with no across-markup specialization; $\eta = 1$ and $f = 0$) is also a special case of the above model. Specifically, if $\alpha_1 = \alpha_2 = 1$, $\sigma_1 = \sigma_2 = \epsilon$ and $\theta_1 = \sigma_H - 1$ and $\theta_1 = \sigma_L - 1$, the model in Fieler (2011) could generate the same trade flows as the restricted (perfectly competitive) case of unified model. While the two models overlap in theory, in estimation they identify distinct channels and, thus, do not overlap. Specifically, in the demand equation (15), the magnitude of $\sigma_2$ regulates the force of the home-production effect and the

\(^{59}\)In terms of micro-level predictions, the unified model is always distinct from non-homothetic models, even in the perfectly competitive case. Therefore, the micro-level evidence presented in section 5.1 do not apply to non-homothetic models.

\(^{60}\)The non-homotehtic model in Fieler (2011) improves the fit ($R^2$) of the standard gravity model by 25% with 4 extra parameters. The unified model improves the fit by 43% with only 2 extra parameters.
wedge, $\frac{\sigma_1}{\sigma_2}$, governs the degree of non-homotheticity. One cannot separately identify both. Fieler (2011) sets $\sigma_2 = 5$ and estimates $\frac{\sigma_1}{\sigma_2}$.\footnote{Similarly, Caron et al. (2014) normalize $\sigma_{\text{trade}} = 1$, and control for the normalized home-production effect by constructing price indexes from the first-stage estimation.} My estimation strategy, by contrast, is analogous to setting $\frac{\sigma_1}{\sigma_2} = 1$ (shutting down non-homotheticity) and estimating $\sigma_2 \sim \varepsilon$, which delivers $\varepsilon = 2.78$.\footnote{This does not mean that home production effect implicit in the Fieler model is stronger than in the unified model. The force of the home production effect is determined collectively by $\varepsilon \sim \sigma_2$ and the price indexes which depend on (i) the trade costs, (ii) $\sigma_H = 1 \sim \theta_1$ and (iii) $\sigma_L = 1 \sim \theta_2$.} Put differently, my estimation underlines the role of the home-production effect, whereas the other approach underlines the role of non-homotheticity.

These arguments are related to a more general identification issue faced by the gravity literature. In cross-sectional data, taste cannot be separately identified from trade costs (note that non-homotheticity is driven by taste, whereas the home-production effect is driven by trade costs). Recently, several studies have employed richer data to disentangle these two forces (e.g. Cosar, Grieco, Li, and Tintelnot (2015); Head and Mayer (2015)). The analysis I perform in section 5.2, while raw, has a similar flavor. The idea is that if trade shares are regulated by taste (non-homotheticity) they should not undergo major transformation. Whereas, if they are driven by trade costs, they are subject to transformation given the tremendous decline of trade impediments.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Unified model</th>
<th>Pure gravity</th>
<th>Restricted gravity (Armington)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_L$ (Normalized)</td>
<td>6</td>
<td>4.97</td>
<td>4.97</td>
</tr>
<tr>
<td>$\sigma_H$</td>
<td>3.27</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>2.78</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$</td>
<td>3.16</td>
<td>2.43</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>$\kappa_{\text{const}}$</td>
<td>2.16</td>
<td>1.71</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\kappa_{\text{dist}}$</td>
<td>0.11</td>
<td>0.13</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\kappa_{\text{border}}$</td>
<td>0.57</td>
<td>0.79</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.013)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>$\kappa_{\text{lang}}$</td>
<td>0.87</td>
<td>0.81</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\kappa_{\text{agreement}}$</td>
<td>0.71</td>
<td>0.84</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td>0.43</td>
<td>0.30</td>
<td>0.21</td>
</tr>
<tr>
<td>(R-squared)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Estimation Results – Standard error are reported in parenthesis.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>The average gains from trade relative to autarky</th>
<th>The coefficient of variation of the gains across countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE UNIFIED MODEL</td>
<td>4.27 %</td>
<td>95.8</td>
</tr>
<tr>
<td>THE PURE GRAVITY MODEL</td>
<td>1.05 %</td>
<td>69.6</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics of the estimated gains from trade relative to autarky. The gains from trade correspond to percentage changes in real wage when moving from the counter-factual autarky equilibrium to the factual trade equilibrium.
Figure 1: The four equilibrium outcomes that characterize the global economy: (i) countries with high national product quality, pay higher wages; (ii) high-wage countries have revealed comparative advantage in type H, and specialize in the production of type H; (iii) due to trade costs, the relative price index of type H is lower in high-income countries; and (vi) high-income countries spend relatively more on type H.
Figure 2: Trade-to-GDP ratio increases with GDP per capita in the data. The unified model captures this pattern, whereas the gravity model does not.
**Figure 3:** The composition of foreign trade as predicted by various models versus the data from 2000. The non-homothetic model’s predictions are calculated using the estimated parameters in Fieler (2011). North corresponds to the 20 richest countries in the sample.

---

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Export share of the North ( \ln X_{N,US} )</th>
<th>Export share of the South ( \ln X_{S,US} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of Product Differentiation (logs)</td>
<td>0.10*** (0.003)</td>
<td>0.03*** (0.003)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.75*** (0.005)</td>
<td>0.74*** (0.004)</td>
</tr>
<tr>
<td>Observations (product×years)</td>
<td>261,021</td>
<td>252,856</td>
</tr>
<tr>
<td>Industry fixed-effect</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.004</td>
<td>0.004</td>
</tr>
</tbody>
</table>

**Table 3:** The export share of high-income countries to the US is significantly higher in highly-differentiated product categories. The high versus low-income categorization is taken from Romalis (2004). The export shares are constructed using 10-digit product-level US import data from 1989 to 2011. The standard errors are reported in parenthesis.
Figure 4: Normalized export flows \( \left( \frac{X_{ij}}{N_i N_j} \right) \) from the richest 50 countries (North) are less sensitive to distance than export flows of the poorest 50 countries (South). The unified model captures this pattern, whereas the gravity model does not.
Figure 5: The unified model predicts that export prices (net of trade costs and wage) strongly increase with GDP per capita. The above figure displays the compositional effect of GDP per capita on export prices in the estimated model.

Figure 6: The unified model delivers the “Washington Apples” effect: countries export higher-priced tradables to geographically distant markets.
Figure 7: The estimated gains from trade relative to autarky. The gains from trade are both systematically larger and more dispersed in the unified model relative to the pure gravity model, which does not accommodate comparative advantage and production specialization.
Figure 8: The gains from comparative advantage-driven trade $\times$ GDP per capita

Figure 9: The gains from comparative advantage-driven trade $\times$ remoteness. Remoteness is calculated as the trade-weighted distance (in thousands of kilometers) to all trading partners.
Figure 10: The effect of trade liberalization on international income inequality. The above graph is generated by lowering the estimated trade costs and calculating international income inequality in the counterfactual models. International income inequality is calculated as the variance of $\ln w_i$ across countries.

Figure 11: Micro-level evidence support the patterns of comparative advantage predicted by the unified model. The graph is constructed using product-level US import data from 1989 to 2011. The markup content is calculated using the estimates in Broda and Weinstein (2006) (see Soderbery (2015) for updated estimates).
Figure 12: The rise of North-South trade relative to North-North trade. The top figure illustrates this trend in the data (the data is from Head, Mayer, and Ries (2010)). The figure decomposes the overall trade of the 21 richest countries over time into: (i) trade with other rich countries (North-North trade) and (ii) trade with middle-income and poor countries (North-South trade). A similar transformation is documented in the 2013 WTO world trade report (WTO, 2013, p. 65, figure B.8). The bottom figure illustrates the unified model’s ability to replicate this pattern. That is, when I counter-factually lower international trade costs North-South trade emerges as the main component of a rich countries’ trade.
Figure 13: The Linder effect prevails in the unified model, which is estimated using data from 2000. The share of imports from rich countries corresponds to imports from the richest 21-richest countries, which account for roughly 70% of the world GDP in 2000.