Quantifying a vertical differentiation trade model*

Monastyrenko E.§ and Picard P.M.‡

July 17, 2025

Abstract

We build a trade model that simultaneously embeds vertical product differentiation, within-country heterogeneous income, heterogeneous goods, and many countries. Under some specifications of costs and preferences, we can establish the existence of the general equilibrium and obtain a very tractable quantification model. We estimate all of the model parameters by applying the model properties on OECD countries. We finally quantify the effect of trade costs and economic shocks – like Brexit – on each country's share of high-quality goods.

JEL Codes: F12, F16, L11, L15

Keywords: vertical differentiation, general equilibrium, international trade

*We thank Esteban Jaimovich and the participants in seminars and workshops at Department of Economics and Management (Luxembourg), and Workshop on the Impact of Global Trade Shocks (Luxembourg) for their comments and suggestions. This research has been supported by the grants TSPQ C20/SC/14755507 and Inter/ Mobility/2021/LE/16527808 of the Fonds National de la Recherche of Luxembourg.

§STATEC, Luxembourg. E-mail: evgenii.monastyrenko@gmail.com

‡Department of Economics and Management, University of Luxembourg. Corresponding author. E-mail: pierre.picard@uni.lu, Tel.: +352 46 66 44 6289, 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg.

1 Introduction

In the past decade, researchers have highlighted important patterns in the quality of traded goods. Yet, the literature has focused on horizontal differentiation models that, to discuss product quality, have been augmented with idiosyncratic demand shifters. In this approach, each good remains horizontally differentiated so that researchers are unable to explain how consumers switch between low and high-quality versions of the same good as it is widely studied in the vertical differentiation literature in Industrial Organization (see the seminal paper by Mussa and Rosen (1978)). To fill this gap, Picard and Tampieri (2021a) and Picard and Tampieri (2021b) present general equilibrium models with vertical differentiation where goods are declined in several quality versions. They show that the theoretical equilibrium properties of such models match the existing empirical regularities.

In this paper, we study the quantitative properties of trade equilibrium models with vertical differentiation and heterogeneous goods. We propose a model of vertical differentiation with many heterogeneous goods and many countries. Following Armington (1969), each country produces a continuous set of differentiated goods that are consumed everywhere. However, here, each good is vertically differentiated with a high and a low-quality version while households purchase a single version of every good. While higher-quality goods give higher utility, they are more costly to produce. On the demand side, workers have heterogeneous skills, which is reflected by heterogeneous endowments of "labor productivity units". As a consequence, they obtain heterogeneous income from their work and richer ones can consume a larger range of high-quality goods. We focus on a class of costs and preferences for goods with two quality versions that make expenditures linear in terms of the inverse marginal utility of income and therefore render the general equilibrium model highly tractable. The expenditures of heterogeneous consumers indeed linearly aggregate within and between countries. So, the equilibrium can be represented by a system of linear equations, which accepts a (nondegenerate) solution and strongly facilitates quantitative exercises.

To focus on the effect of product quality, we sterilize the effect of extensive margins. We can then express the average quality of imports – in quantity and values – as function of a key statistics that relate the consumer's inverse marginal utility of income to the trade cost and producers' labor price. The latter statistics can also be expressed as a function of local income and remoteness indices. This allows us to set up a gravity equation that relates exports to trade costs, wages and populations, and remoteness indices.

The conventional approach is to model horizontal quality through a demand shifter (Flam and Helpman, 1987; Verhoogen, 2008; Baldwin and Harrigan, 2011; Fajgelbaum et al., 2011, 2015; Dingel, 2017). Khandelwal (2010) was the first to propose to measure the quality of traded goods as the residual of the demand. This approach has been further developed in the studies of Hallak and Schott (2011) and Khandelwal et al. (2013). Gervais (2015) investigates the role of product quality for exporting and levels of foreign sales of U.S. manufacturing plants. The quality is proxied by estimated idiosyncratic demand from price and quantity information. Piveteau and Smagghue (2019) developed an instrumental variable approach to estimate the horizontal quality that is free of productivity variation. According to their methodology, the import-weighted real exchange rates are used as an instrument for export prices. Then, the quality equals the residual export variations in a regression after controlling for prices.

When households purchase a single vertically differentiated product the quality and price of consumption goods rises in the income level, creating the a positive relationship between prices and per capita income observed in the data (Hallak, 2006; Choi et al., 2009; Fajgelbaum et al., 2011).

We contribute to the literature that implements non-homothetic preferences to the trade models. Fieler (2011) is one of the pioneering applications of nonhomotheticity across traded goods. In this extension to a Ricardian model, the consumption of goods is higher in rich countries with more diversified technologies.

Bernasconi and Wuergler (2012) extend the model of Krugman (1980) with non-homothetic preferences. The aim at interpreting the fact that richer countries import more along the extensive and quality margin. Feenstra and Romalis (2014) build the monopolistic competition model where firms simultaneously decide on product quality in the context of non-homothetic demand. Their estimations suggest a positive correlation with a magnitude of 14% between countries' incomes and export quality. Jaimovich and Merella (2012) assume nonhomotheticity within goods, with richer consumers preferring higher qualities of each good. At the same time, the nonhomotheticity across goods means that richer consumers are allocating larger expenditure shares towards goods with higher scope for quality upgrading. Jaimovich and Merella (2015) build a Ricardian model with nonhomotheticities featuring horizontal and vertical differentiation, where willingness to pay for quality increases with income. Eaton and Fieler (2019) construct two-tier CES preferences nesting horizontal and vertical dimensions of goods.

A main contribution lies in the application of the above theoretical model on OECD countries using only three datasets: the Trade Unit Value Database, BACI database, and Historical Bilateral Trade and Gravity Data set (Berthou and Emlinger, 2011; Gaulier and Zignago, 2010; Fouquin et al., 2016). Following the structure of the above theoretical model, we can estimate the average quality upgrade cost, the bilateral trade costs between country pairs, and the endowments and prices of labor productivity units in each country.

In the spirit of Fajgelbaum and Khandelwal (2016), we recover a key parameter of the general equilibrium from the estimation of a gravity equation that is consistent with the theoretical model. We then calibrate the model using the estimated coefficients and validate it with simulated and measured values of countries' shares of high-quality goods. Finally, the quantitative model allows us to assess the effect of trade shocks and reforms. For the sake of conciseness, we focus on the impact of trade costs and the effect of the Brexit on the quality of goods imported in and exported from OECD countries.

In our opinion, the paper presents three methodological contributions. It first offers an approach to quantify a general equilibrium model embedding the the framework of vertical product differentiation framework, which is subject to strong interest in Industrial Organization literature. Second, it presents the quantification of a model with non-homothetic preferences, which contrasts with much of the quantification literature that is based on homothetic preferences. Third, the estimation of the model parameters is consistently made using the same data set, without importing parameters from other data or literature insets.

The remainder of this paper is organized as follows. Section 2 introduces the model. Section 3 describes the data sources and provides summary statistics. In Section 4, we detail our estimated empirical specifications. We report estimation results in Section 5. In Section 6, we run a series of counterfactual exercises. Finally, Section 7 concludes.

2 Model

We consider an economy with N trading countries $i \in \{1, ..., N\}$ populated by a mass of households m_i, with $\sum_i m_i = 1$, where in this paper the summation over i applies over the N countries. Each country i hosts a continuum of households h who are each endowed with labor productivity units s_{ih}, which reflects local productive skills and education and is distributed with c.d.f. F_i. Denoting the wage per productivity unit by w_i, a household h in country i earns an income $w_i s_{ih}$. The country's average labor productivity is denoted by s_i (i.e. $s_i = \int s_{ih} \mathrm{d}F_i(s_{ih})$). The country's average income is given by $w_i s_i$.

Each household in country i consumes a set of differentiated 'goods' produced in country $j \in \{1, ..., N\}$. We model product differentiation in three dimensions: horizontal differentiation between goods and then both horizontal and vertical differentiation within each good. At a high level, each good is horizontally differentiated according to the parameter $z \in [0, 1]$ (e.g., HS4, motor car HS-8703). At a lower level, each good includes a set of horizontally differentiated 'varieties' $\nu \in [0, n_{ij}]$ (e.g., combustion car or electric car) and are offered at high and low-quality levels $k \in \{H, L\}$ (e.g. combustion car with low or high cylinder capacity). For simplicity, the horizontal differentiation of the variety is assumed to be symmetric and implies the same preference and cost structure.

By contrast, vertical differentiation implies heterogeneous preference and cost structures towards the quality levels. In the spirit of Armington (1969), we assume that each variety and quality of a good is produced only in one country and is consumed by all consumers in every country while the number of consumed varieties n_{ij} is exogenous. In line with the vertical differentiation literature, we assume that each household consumes a single unit of each variety but chooses its quality version. Hence, the mass of goods offered in the country i is equal to $N * 1$ and the mass of varieties (ν, k) offered there is given by $2\sum_j n_{ij}$ whereas the mass of varieties consumed by a household is given by $\sum_j n_{ij}$. Finally, for the sake of simplicity and realism, we assume that the consumer buys a basket mixing high and low qualities. In the industrial organization literature on vertical differentiation, the latter assumption corresponds to the "full market coverage" condition.

A consumer gets $b_H(z) > 0$ utility units for the high-quality version of a variety of good z and $b_L(z) > 0$ for its low-quality version. Since each household consumes a unit of every variety z produced in every country j, a household in country i maximizes the utility

$$U_i = \sum_j n_{ij} \int_0^1 \left(\sum_{k=H,L} b_k(z)\, x_{ijk}(z) \right) \mathrm{d}z, \tag{1}$$

subject to the budget constraint

$$\sum_j n_{ij} \int_0^1 \left(\sum_{k=H,L} p_{ijk}(z)\, x_{ijk}(z) \right) \mathrm{d}z = w_i s_{ih}, \tag{2}$$

where $p_{ijk}(z) > 0$ is the (destination) consumer prices, $x_{ijk}(z) \in \{0, 1\}$ the unitary consumption decision of variety z ($x_{ijH} + x_{ijL} = 1$) and $w_i s_{ih}$ the household's income.

Production technologies depend on each good z and the quality k of its varieties.[1] In particular, the production of each variety z requires $a_H(z)$ and $a_L(z)$ labor productivity units for the high and low-quality versions. We assume proportionate cost and utility upgrades, i.e., $a_H(z)/a_L(z) = \alpha/(\alpha - 1) > 1$ and $b_H(z)/b_L(z) = \beta/(\beta - 1) > 1$ where $\alpha > 1$ and $\beta > 1$.

When it is exported, the variety incurs an iceberg trade cost $\tau_{ij} \geq 1$ in terms of labor input, where a share $1/\tau_{ij}$ of the variety arrives at the destination after shipment from country j. Trade costs are symmetric across countries and nil within countries: $\tau_{ji} = \tau_{ij} > 1$ and $\tau_{ii} = 1$. Under perfect competition, the price of a variety z sold in country i and

[1] As said above, horizontal differentiated varieties ν are symmetric within each good.

produced in country j is equal to its unit cost:

$$p_{ijk}(z) = \tau_{ij} w_j a_k(z), \quad k = H, L, \tag{3}$$

where w_j is the wage per labor productivity unit in the production country j.

2.1 Demands

The above consumption choice problem yields household demands of a variety ν of a good z. In particular, household i buys the high-quality version H of such a variety if

$$b_H(z) - \frac{1}{\mu_{ih}} p_{ijH}(z) \geq b_L(z) - \frac{1}{\mu_{ih}} p_{ijL}(z), \tag{4}$$

and the low-quality L otherwise. In this expression, the scalar μ_{ih} measures the inverse of the marginal utility of income (it is equal to the inverse value of the Lagrange multiplier of the budget constraint). Because higher income $w_i s_{ih}$ relaxes the budget constraint, it also raises μ_{ih}. So, μ_{ih} is a proxy for income.

Using prices eq. (3), the set of high-quality varieties consumed by household h in country i and produced in country j is given by $[0, n_{ij}] \times \mathcal{H}(\mu_{ih}/(\tau_{ij} w_j))$ where

$$\mathcal{H}\left(\frac{\mu_{ih}}{\tau_{ij} w_j}\right) \equiv \left\{z : \frac{\mu_{ih}}{\tau_{ij} w_j} \geq \ell(z)\right\}, \tag{5}$$

where

$$\ell(z) \equiv \frac{a_H(z) - a_L(z)}{b_H(z) - b_L(z)},$$

denotes the per-quality-unit labor input needed to upgrade variety z. The low-quality varieties belong to the complement of this set. W.l.o.g., we rank the varieties z such that $\ell'(z) > 0$. Two restrictions must be satisfied. First, because all consumers buy a mix of high and low qualities, we impose $\mu_{ih}/(\tau_{ij} w_j) \in [\ell(0), \ell(1)]$, $\forall i, j$, which guarantees that the identity $\mu_i/(\tau_{ij} w_j) = \ell(z)$ has a unique interior solution. Second, since all consumers purchase all varieties, it must be that the input per quality schedule ℓ lies above the schedules a_L/b_L and a_H/b_H, which is fulfilled if

$$\frac{\mu_{ih}}{\tau_{ij} w_j} \geq \max\left\{\frac{a_L(1)}{b_L(1)}, \frac{a_H(1)}{b_H(1)}\right\}, \forall i, j. \tag{6}$$

As μ_{ih} is positively related to income, this condition expresses that consumers have a high enough income to purchase all low-quality goods. Note that, since consumers buy all varieties, changes in income do not change extensive margins. Shutting down the extensive margin allows us to highlight the role of quality margin in a general equilibrium setting.

2.2 Expenditure

The expenditure of household h in country i on goods produced in j is given by

$$E_{ijh} = n_{ij}\left[\int_{\mathcal{H}\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)} \tau_{ij}w_j a_H(z)\mathrm{d}z + \int_{\mathcal{L}\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)} \tau_{ij}w_j a_L(z)\mathrm{d}z\right] \tag{7}$$

This can be expressed as $E_{ijh} = n_{ij}\tau_{ij}w_j E\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)$ where $E(\cdot)$ is called the *real expenditure function of a variety* that we define as

$$E\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right) \equiv \int_{\mathcal{H}\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)} a_H(z)\mathrm{d}z + \int_{\mathcal{L}\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)} a_L(z)\mathrm{d}z. \tag{8}$$

This measures the labor content (in terms of productivity units) of the variety consumed by the household.

To ease analytical tractability, we focus on a set of costs and preferences that facilitate aggregation (similarly to Pollak (1969) for divisible goods). In particular, we focus on the cost and utility profiles such that household expenditures E_{ijh} are linear in the household's income $w_i s_{hi}$. To achieve this, we first assume that the real expenditure function is the linear function $E(y) = y - r$, which can be shown to be equivalent to the following assumption on the per-quality input

$$\ell(z) = \frac{a\left(0\right)}{b(0)} + \int_0^z a(\zeta)\mathrm{d}\zeta \tag{9}$$

where $a(z) = a_H(z) - a_L(z)$ and $b(z) = b_H(z) - b_L(z)$ are the profiles of the costs and benefits of quality upgrades. This condition is equivalent to the following assumption on the profile of utility upgrades:

$$b(z) = \frac{a(z)}{\frac{a(0)}{b(0)} + \int_0^z a\left(\zeta\right)\mathrm{d}z}. \tag{10}$$

Under this assumption, the intercept of the real expenditure is equal to $r = \alpha\ell(0) - (\alpha - 1)\ell(1)$. One can check that $r < \ell(0) < \ell(1)$. The value of r is subject to our empirical analysis below. We assume it to be positive for simplicity.

We can then characterize the expenditure of a household h in country i for goods produced in j as

$$\begin{aligned}
E_{ijh} &\equiv n_{ij}\left[\int_{\mathcal{H}\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)} \tau_{ij}w_j a_H(z)dz + \int_{\mathcal{L}\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)} \tau_{ij}w_j a_L(z)dz\right] \qquad (11)\\
&= n_{ij}\tau_{ij}w_j E\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)\\
&= n_{ij}\tau_{ij}w_j\left(\frac{\mu_{ih}}{\tau_{ij}w_j} - r\right)\\
&= n_{ij}\left(\mu_{ih} - r\tau_{ij}w_j\right).
\end{aligned}$$

Finally, we plug this into the budget constraint so that

$$w_i s_{ih} = \sum_{j=1}^{N} E_{ijh} = \mu_{ih}\sum_{j=1}^{N} n_{ij} - r\sum_{j=1}^{N}\tau_{ij}w_j n_{ij}. \qquad (12)$$

and rearrange as

$$\frac{\mu_{ih}}{\tau_{ij}w_j} = \frac{w_i s_{ih}}{\tau_{ij}w_j n_i} + \frac{r}{\tau_{ij}w_j n_i}\sum_{l=1}^{N}\tau_{il}w_l n_{il} \qquad (13)$$

where $n_i = \sum_l n_{il}$ is the number of consumed varieties within each good category. This expression is linear in household income. As $E(\cdot)$ is also linear, the expenditure $E_{ijh} = \tau_{ij}w_j E\left(\mu_{ih}/\tau_{ij}w_j\right)n_{ij}$ is also linear in income. More precisely, the expenditure writes as

$$E_{ijh} = \frac{n_{ij}}{n_i}\left[w_i s_{ih} + r\left(\sum_l \tau_{il}w_l n_{il}\right) - r\tau_{ij}w_j n_i\right], \qquad (14)$$

which is linear in household income as proposed above.

Finally, the above linearity allows us to average expenditures across households of the same country. The average *expenditure* on goods exported from country j to country i is then given by $E_{ij} \equiv \int E_{ijh}\mathrm{d}F_i(s_{ih}) = \int \tau_{ij}w_j E\left(\mu_{ih}/\tau_{ij}w_j\right)n_{ij}\mathrm{d}F_i(s_{ih})$, or equivalently,

$$E_{ij} = n_{ij}\tau_{ij}w_j E\left(\frac{\mu_i}{\tau_{ij}w_j}\right) \qquad (15)$$

where $\mu_i \equiv \int \mu_{ih}\mathrm{d}F_i(s_{ih})$ is the country average of the multiplier μ_{ih}. The latter is simply obtained from expression eq. (13) where we set the average income $w_i s_i$ in the first term on its RHS.

2.3 Key statistics

The statistics $\mu_{ih}/\left(\tau_{ij}w_i\right)$ is sufficient to express the set of high-quality varieties $\mathcal{H}\left(\mu_i/\tau_{ij}w_i\right)$ and can be considered as a proxy for the *quality margin,* which adapts to economic shocks. This statistics also determines the share of high-quality goods from a country j, defined by the quantity of high quality varieties $n_{ij}\int_{\mathcal{H}(\mu_{ih}/\tau_{ij}w_j)}\mathrm{d}z$ divided by the mass of high and low quality varieties n_{ij}. This gives

$$SHQ_{ijh}=\int_{\mathcal{H}(\mu_{ih}/\tau_{ij}w_j)}1*\mathrm{d}z=\ell^{-1}\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right), \tag{16}$$

where ℓ^{-1} is the inverse function of the per-quality-unit labor schedule ℓ. This share increases with statistics $\mu_{ih}/\left(\tau_{ij}w_j\right)$; that is, with higher household's income $w_i s_{ih}$ proxied by μ_{ih}, lower production cost w_j and lower trade costs τ_{ij}. Since it varies with household income, this distribution of this share is given by the income distribution, after transformation by $\ell^{-1}\left(\cdot\right)$.

We can aggregate this to get the share of high-quality goods in cif value from country j to i as

$$SHV_{ij}=\frac{1}{E_{ij}}\int n_{ij}\left(\int_0^{\widehat{z}_{ijh}}\tau_{ij}w_j a_H(z)\mathrm{d}z\right)\mathrm{d}F_i(s_{ih}), \tag{17}$$

where $\widehat{z}_{ijh}$ solves $\ell(\widehat{z}_{ijh})=\mu_{ih}/\left(\tau_{ij}w_j\right)$. This sums up every household's expenditure on each high quality good and divides the results by the country's total expenditure. Given the assumption that households consume a mix of high- and low-quality varieties (i.e. $\ell(0)<\mu_{ih}/\left(\tau_{ij}w_j\right)<\ell(1)$), this expression simplifies to

$$SHV_{ij}=\alpha\frac{\frac{\mu_i}{\tau_{ij}w_j}-\ell(0)}{E\left(\frac{\mu_i}{\tau_{ij}w_j}\right)}=\alpha\frac{\frac{\mu_i}{\tau_{ij}w_j}-\ell(0)}{\frac{\mu_i}{\tau_{ij}w_j}-r} \tag{18}$$

The expression is an increasing function of $\mu_i/\left(\tau_{ij}w_j\right)$ ranging in the interval $[0,\alpha]$ (because $r<\ell(0)<\mu_{ih}/\left(\tau_{ij}w_j\right)<\ell(1)$). It requires only the country statistics on $\mu_i/\left(\tau_{ij}w_j\right)$, but not the household statistics $\mu_{ih}/\left(\tau_{ij}w_j\right)$.

Under linear real expenditures, the indirect utility of household h in country i simplifies to

$$V_{ih}=\sum_j n_{ij}\left[\ln\left(\frac{\mu_{ih}}{\tau_{ij}w_j}\right)+V_0\right] \tag{19}$$

where $V_0=-\beta\ln\ell(0)+(\beta-1)\ln\ell(1)$ is a constant. The indirect utility increases with the statistics $\mu_{ih}/\left(\tau_{ij}w_j\right)$; that is, with higher consumer income, lower production cost and lower trade costs.

The share of high-quality goods and the utility function are not linear functions of $\mu_{ih}/(\tau_{ij}w_j)$ and therefore income $w_i s_{hi}$. So, their averages do not correspond to the choice of a potentially representative consumer with the average income $w_i s_i$ and incentive to purchase $\mu_{ih}/(\tau_{ij}w_j)$. In other words, the model does not admit a representative consumer (Gorman, 1959).

However, the average incentive $\mu_i/(\tau_{ij}w_j)$ is the key statistic that describes the aggregate trade flows. In particular, the fob import value per capita can be written $E_{ij}^{\text{fob}} = w_j E\left(\mu_i/(\tau_{ij}w_j)\right) n_{ij}$. The cif import value per capita in country i is given by $E_{ij}^{\text{cif}} = \tau_{ij} w_j E\left(\mu_i/(\tau_{ij}w_j)\right) n_{ij}$.[2] Given that consumers in country i purchase a mass n_{ij} of varieties produced in country j, average import prices are given by $\overline{p}_{ij}^{\text{fob}} = w_j E\left(\mu_i/(\tau_{ij}w_j)\right)$ and $\overline{p}_{ij}^{\text{cif}} = \tau_{ij}\overline{p}_{ij}^{\text{fob}} = \tau_{ij} w_j E\left(\mu_i/(\tau_{ij}w_j)\right)$.

Finally, the mass of all quality varieties of a good z consumed in country i and produced in country j is equal to the total number of varieties $1 \times n_{ij}$ times the population mass m_i. By contrast, the consumption mass of the high-quality varieties of a specific good z is determined by the mass of households having a high enough income to be willing to consume them, i.e. such that $\mu_{ih}/(\tau_{ij}w_j) \geq \ell(z)$. By eq. (13), those households have a productivity s_{ih} larger than the threshold $\overline{s}_{ij}(z) = \left[\tau_{ij} n_i w_j \ell(z) - r\sum_l \tau_{il} w_l n_{il}\right]/w_i$. Hence, country i consumes an amount of $n_{ij} m_i \left[1 - F_i(\overline{s}_{ij}(z)\right]$ of high quality varieties.

2.4 General equilibrium

In a general equilibrium, the import value of each country i equates to the values of its exports:

$$\sum_{l\neq i} m_i \tau_{il} w_l E\left(\frac{\mu_i}{\tau_{il} w_l}\right) n_{il} = \sum_{l\neq i} m_l \tau_{li} w_i E\left(\frac{\mu_l}{\tau_{li} w_i}\right) n_{li},\ \forall i \tag{20}$$

Given the linear real expenditure, each balanced trade condition simplifies to

$$\sum_j m_i \left(\mu_i - r\tau_{ij} w_j\right) n_{ij} = \sum_j m_j \left(\mu_j - r\tau_{ji} w_i\right) n_{ji}. \tag{21}$$

A trade equilibrium is defined by the vector of inverse marginal utility of representative worker's income $\mu = (\mu_1, ..., \mu_N)$ that matches households' optimal consumption choices eq. (13), the vector of unit wages $w = (w_1, ..., w_N)$ that balances trade conditions eq. (21) and such that consumers buy all varieties and a mix of qualities.

[2]Because of trade costs, the average import prices must be distinguished by whether they are evaluated at the origin or destination. Following international trade terminology, freight on board (fob) prices do not include trade costs while cost, insurance & freight (cif) prices include them. Exports are most generally reported in fob values at the borders of exporting countries and imports are denominated in cif prices at the gates of importing countries.

Expressions eq. (13) and eq. (21) simplify to

$$w_i \left(m_i s_i + r \sum_j m_j n_{ji} \tau_{ji} \right) = \sum_j \frac{m_j n_{ji}}{n_j} \left(w_j s_j + r \sum_{l=1}^{N} \tau_{jl} w_l n_{jl} \right). \tag{22}$$

By the Walras law, an equation is redundant while a variable can be normalized to one. So, normalizing $w_N = 1$, this represents a system of $N-1$ linear equations with $N-1$ variables $\{w_i\}_{i \neq 1}$. It yields a unique solution if the system has a rank $N-1$. Although it has no easy solution, this linear system can readily be solved by algebraic methods.

We end up with a discussion of the gravity equation that expresses trade values as functions of local incomes and distances. Country j's export to country i is captured by the expenditure and number of high-quality varieties, which increases with the statistics $\mu_i / (\tau_{ji} w_j)$. The (nominal per-capita) expenditure on import from j to i (at cif prices) is given by

$$E_{ij}^{\text{cif}} = \tau_{ij} w_j E \left(\frac{\mu_i}{\tau_{ij} w_j} \right) n_{ij}, \tag{23}$$

which gives

$$E_{ij}^{\text{cif}} = \left(w_i s_i \frac{n_{ij}}{n_i} \right) - r \left(\tau_{ij} n_{ij} w_j \right) + r \frac{n_{ij}}{n_i} \left(\sum_{l=1}^{N} \tau_{il} n_{il} w_l \right) \tag{24}$$

From this expression, it comes that trade expenditure rises with higher values of importer's income per capita $s_i w_i$, higher values of exporter's price of labor unit w_j, lower bilateral trade cost τ_{ij} and higher remoteness, here measured by $\sum_{l=1}^{N} \tau_{il} w_l n_{il}$. This will be confirmed with the data in the next sections.

3 Data

This paper includes a number of empirical tests employing different datasets. This section provides details on the datasets for each empirical exercise.

For country-level gravity estimations, we extract the data on trade flows between 2000 and 2014 from the BACI database (Gaulier and Zignago, 2010). In order to focus on similar countries, we select the countries belonging to OECD for which BACI reports a trade flow every year. This narrows our sample to 30 countries and 15 years.[3] This gives $13,050$ observations of cif trade values between pairs of countries. We use the information on the per-capita GDP, the share of the tertiary sector, the border, cultural and colonial links, and

[3] Those countries are Australia, Austria, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Mexico, New Zealand, Norway, Poland, Portugal, Slovenia, South Korea, Spain, Sweden, Switzerland, Turkey, the United Kingdom, United States.

the (spherical) distances between capital cities from the ”Gravity” database (Conte et al., 2022).[4] Table 1 displays a summary of the main variables of interest.

The estimations of the trade costs are based on the disaggregated dataset at the level of HS6 varieties from BACI using cif prices. The latter are computed by dividing total values by total quantity. Using BACI, we count the number of varieties traded within each country pair as the number of HS6 categories with observed trade flows. Because the number of locally produced and consumed varieties are not reported, we assume that this number is equal to the total number of varieties exported by the country.

To estimate quality, we proxy the price of each variety HS6 by its fob unit value. The fob prices are directly observed by exporters at the country border and exclude nontradable components such as transportation costs, tariffs and distribution costs at the destination. We source unit values from Trade Unit Value Database (Berthou and Emlinger, 2011).[5] Feenstra and Romalis (2014) show that the variation in trade unit values at the level of 4-digit in SITC classification could be principally attributed to differences in product quality. Therefore, we construct two quality levels in each HS4 category produced in every country using the Trade Unit Value Database. Towards this aim, we fix each HS4 category produced at the destination country and rank the HS6 varieties according to their unit value at the mill (fob). We discard the top 10% and bottom 10% on the suspicion that those items may not be regular products (e.g. spare parts). We denote the HS6 category as high quality if it has a unit value higher than the median unit value. In the other case, it is denoted as low quality. This procedure allows us to distinguish between basic and advanced items. This allows us to construct the sets of high- and low-quality goods for each HS4 goods and in every trade direction.

[4] The data is maintained and updated on yearly basis by CEPII. The most complete technical description is available here: http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=8.

[5] As an alternative, the unit values might be sourced from BACI dataset. Note that in BACI data imports are reported for C.I.F terms, while exports for F.O.B. To avoid this ambiguity, we use the Trade Unit Value Database where the trade flow is a unique flow between partners.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	N
GDP per capita (country)	31,869.129	18,884.924	13,050
Import value (pair)	231.174	585.368	13,050
Number of varieties (pair)	2.407	0.639	13,050
Common language (pair)	0.074	0.261	13,050
Common border (pair)	0.06	0.237	13,050
Colonial links (pair)	0.028	0.164	13,050

Notes: This table reports summary statistics for the gravity estimates.

We then construct proxies N_{ij} for the mass of horizontally differentiated product n_{ij} as it follows. Using the data for each high- and low-quality good, we first build the consumer's annual unitary weight of a HS4 good, $Q(z)$, as the world weight (kg) of all the HS6 items in this HS4 good divided by the total population considered in the database. We then build a proxy $N_{ij}(z)$ for the mass of horizontally differentiated product for each HS4 good z traded between two countries as the annual weight of all HS6 items traded between these countries divided by $Q(z)$. We finally set N_{ij} as the average of $N_{ij}(z)$ across HS4 goods.

4 Empirical specifications

Our quality model depends on several parameters and variables. In this section, we explain our strategy to recover those elements. We consider observations of varieties ω in different time periods t, with information on destinations i and origins j, horizontal and vertical differentiation indices ν, z, and k. In this section, upper case letters respectively refer to observed values.

Trade cost The first set of parameters we estimate is the trade costs. Trade costs include transport costs and other costs associated with discrepancies in language and cultural relationships. Trade costs can be recovered from the price relationship eq. (3), $p_{ijk}^{\text{cif}}(z) = \tau_{ij} w_j a_k(z)$, $k = H, L$. This gives the the econometric model

$$\log P_{\omega,t}^{\text{cif}} = \theta_0 * dist_{ij} + X_{ij,t} + FE_k \times FE_z + FE_j + FE_t + \varepsilon_{\omega,t} \tag{25}$$

where $\log P_{\omega,t}^{\text{cif}}$ is the measured cif price, $dist_{ij}$ the distance between countries, $X_{ij,t}$ language, cultural and border controls, FE fixed effects on the goods and their quality, on origin countries and time, and $\varepsilon_{\omega,t}$ is the error term. Therefore, $\widehat{T}_{ij} = \exp(\widehat{\theta}_0 * dist_{ij} + X_{ijt})$

expresses the trade cost associated with distance, language, culture, and border. We assume $\widehat{T}_{ii} = 1$. Since we measure the fob expenditure $E^{\text{fob}}_{ij,t}$, we use the estimates of $\widehat{T}_{ij}$ to construct the c.i.f expenditures as $E^{\text{cif}}_{ij,t} = \widehat{T}_{ij} E^{\text{fob}}_{ij,t}$, which will be used below.

Quality cost ratio We recover the relative cost difference between high and low quality, $\gamma \equiv \alpha/(\alpha - 1)$, by regressing the average fob prices of goods in each country pair with respect to the quality dummy of the goods. We add controls for goods and origin countries to ensure the comparability between goods and origins as assumed by Armington (1969). We add controls for trade costs to absorb possible pass-through effects that are not included in the theoretical model.

Prices and endowments of labor productivity units The second set of variables is the system of prices of labor productivity units w_j. The labor productivity units and their prices are however not observed (they are not equal to the hourly productivity measures reported in usual statistics tables). We therefore estimate them using eq. (2), which imposes

$$\frac{p^{\text{fob}}_{ijk}(z)}{p^{\text{fob}}_{ilk}(z)} = \frac{w_j}{w_l} \tag{26}$$

for the same destination i and two different origins j and l, and the same goods z of the same quality k. For each time period, we therefore take the varieties ω and ω' with different origins j and l but the same destination i, goods z and quality k and estimate the values $EW_{n,t}$ such that

$$\log P^{\text{fob}}_{\omega,t} - \log P^{\text{fob}}_{\omega',t} = \sum_{n=1}^{N} LOGW_{n,t} \left(1_{\{n=j\}} - 1_{\{n=l\}}\right) + \varepsilon'_{\omega\omega',t} \tag{27}$$

where $P^{\text{fob}}_{\omega,t}$ and $P^{\text{fob}}_{\omega',t}$ are the measured fob prices of varieties ω and ω' with characteristics (i, j, k, z) and (i, l, k, z), dummy value $1_{\{\text{predicat}\}}$ is equal to 1 if $predicat$ is true and 0 otherwise and finally $\varepsilon'_{\omega\omega',t}$ is the error term. We recover the prices of labor productivity units as $\widehat{W}_{n,t} = \exp\left(\widehat{LOGW}_{n,t}\right)$.

Gravity We estimate the model coefficient r at the macroeconomic level using the gravity eq. (24). We proxy the number of varieties n_{ij} in the country pair (i, j) by the number of HS6 categories divided by the number of HS4 goods in the pair at time t, $N_{ij,t}$. The total number n_i of imported varieties in country i is proxied by $N_{i,t} = \sum_j N_{ij,t}$. The household income $w_i s_i$ is proxied by the annual expenditure on non-service goods $INC_{i,t}$ (i.e. GDP per capita * (one minus the share of service sector)).

We build the terms $\tau_{ij} n_{ij} w_j$ as $R_{ij,t} = \widehat{T}_{ij} N_{ij,t} \widehat{W}_{j,t}$ for each time period and the remoteness factors $\sum_{l=1}^{N} \tau_{il} n_{il} w_l$ as $R_{i,t} = \sum_{l=1}^{N} \widehat{T}_{il} N_{il,t} \widehat{W}_{l,t}$. We therefore can run the OLS regression

$$E_{ij,t}^{\text{cif}} = c_0 \left(INC_{i,t} \frac{N_{ij,t}}{N_{i,t}} \right) + c_1 R_{ij,t} + c_2 \frac{N_{ij,t}}{N_{i,t}} R_{i,t} + \delta_{ij,t} \tag{28}$$

where $\delta_{ij,t}$ is the error term. The coefficient c_0 is expected to be equal to one while c_1 and c_2 are expected to be respectively equal to $-r$ and $+r$. An estimated value of c_0 close to one and opposite estimated values for c_1 and c_2 suggest evidence of a good model specification. The value of r can further be obtained by constraining this model to $c_0 = 1$ and $c_1 = -c_2$.[6]

Quality upgrade cost profile We finally estimate the input profiles of $a_H(z)$ and $a_L(z)$. We use the identity eq. (3) to recover the input $a_k(z)$, $k = H, L$, from the fob price of $p_{ijk}^{\text{fob}}(z)$ in each country pair (i, j) and the estimated values of w_j and τ_{ij}. The important point is to express the input $a_k(z)$ in terms of an annual unit consumption of each good z. In the model, a consumer purchases a unit consumption of each good, which corresponds to a specific weight (kg) reported in the dataset.

We therefore construct two proxies $A_H(z)$ and $A_L(z)$ for the input profiles $a_H(z)$ and $a_L(z)$ of each good z as it follows. We compute the fob average unit values $UV_{ij,k}^{\text{fob}}(z)$, $k = H, L$, of a kg of the high- and low-quality versions of the HS4 good z. The theoretical price $p_{ijk}^{\text{fob}}(z)$ is then proxied by this unit value times the weight $Q(z)$ (kg) of the corresponding HS4 good; that is, $UV_{ij,k}^{\text{fob}}(z)Q(z)$. The proxied input is then given by $A_k(z) = UV_{ij,k}^{\text{fob}}(z)Q(z)/\left(\widehat{T}_{ij} N_{ij} \widehat{W}_j\right)$, $k = H, L$, where $\widehat{T}_{ij}$ and $\widehat{W}_j$ are the projections of the trade cost and price of labor productivity unit discussed above. We rank the goods in increasing order of the average input $A_M(z) \equiv \left[A_H(z) + A_L(z)\right]/2$. We finally smoothen the profiles of $A_L(z)$ and $A_H(z)$ to proxy the input profiles of $a_H(z)$ and $a_L(z)$

5 Estimation results

In this section, we report our estimation results.

Trade cost Table 2 presents the estimates of bilateral trade costs with respect to distance according to the regression model eq. (25). Observations are BACI cif prices for country pairs, years, and HS6 categories. The first column includes no controls and fixed effects

[6] In this case we estimate β in the OLS regression

$$\left(E_{ij,t}^{\text{cif}} - INC_{i,t} \frac{N_{ij,t}}{N_{i,t}} \right) = c_1 \left(R_{ij,t} - \frac{N_{ij,t}}{N_{i,t}} R_{i,t} \right) + \delta'_{ij,t}$$

where $\delta'_{ij,t}$ is the error term. Here, c_1 is expected to be equal to $-r$.

Table 2: Estimations of trade costs τ

	(1)	(2)	(3)	(4)	(5)	(6)
Distance in km	0.0000637***	0.0000555***	0.0000322***	0.0000316***	0.0000308***	0.0000450***
	(9.23e-08)	(0.000000101)	(5.73e-08)	(6.38e-08)	(6.20e-08)	(9.03e-08)
Common language		-0.0851***	-0.0260***	-0.138***	-0.135***	-0.0769***
		(0.00140)	(0.000792)	(0.000852)	(0.000828)	(0.000944)
Colonial links		-0.109***	-0.0508***	-0.0278***	-0.0219***	-0.0121***
		(0.00173)	(0.000977)	(0.000977)	(0.000950)	(0.000966)
Common border		-0.380***	-0.212***	-0.214***	-0.208***	-0.220***
		(0.00140)	(0.000793)	(0.000806)	(0.000783)	(0.000811)
Product k FE			Yes	Yes	Yes	Yes
Destination j FE				Yes	Yes	Yes
Year t FE					Yes	Yes
Origin i FE						Yes
Observations	17,921,006	17,921,006	17,921,006	17,921,006	17,921,006	17,921,006
Adjusted R^2	0.026	0.032	0.692	0.705	0.721	0.729

Notes: This table reports estimations of trade costs. The dependent variable is the log of CIF price sourced from BACI. Robust standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.

and reports a positive effect of distance on trade cost but a low overall fit. The second column introduces proxies for cultural proximity like contiguity, language, and colonial ties, which have coefficients with the expected negative signs, only slightly improving the fit and reducing the coefficient on trade distance. The third column includes additional fixed effects for each HS4 good, which strongly increases the explanatory power and reflects the strong heterogeneity of goods and their shipping cost structures. The fourth column adds destination-fixed effects in order to control for potential market segmentation and pass-through effects that are not included in our model. The fifth column includes year-fixed effects and does not change the coefficient on distance. Finally, in the sixth column, we control for origin countries and get results consistent with the previous columns. Since this column matches the most our regression model eq. (25), we choose it as the preferred specification. The coefficient $\theta_0 = 0.000045$ means that additional 10,000 km of distance increases the log of the CIF price (and therefore trade cost) by 45%.

Quality cost ratio Table 3 presents the estimation of the value of quality ratio $\gamma = \alpha/(\alpha - 1)$. Observations are again taken from BACI for country pairs, years, and HS6 categories. The dependent variable is the log of fob unit value. The first column reports the regression with fixed effects for destination and origin countries and shows a low fit ($R^2 = 4.1\%$). The second column adds a fixed effect for each HS4 good and substantially improves the fit (R^2 rises up to 81.6%) showing that the HS4 product categories explain most of the variation in unit values. The third column captures the effect of quality through a dummy specifying the high (versus low) quality of the HS6 product observation. This coefficient is statistically significant and equals 1.09, which implies a value for γ of

2.97 ($= \exp(1.09)$), which yields $\alpha = 1.507$. The next columns present robustness checks. The fourth column tests the potential impact of product heterogeneity on quality ratio γ. Towards this aim, it includes a dummy for each quality level within each HS4 category. This regression has a similar fit (very similar R^2) so that the potential heterogeneity of γ across HS4 categories does not seem to have an impact on price structures. The fifth and sixth columns check for potential effects of distance on the estimation of the quality ratio γ. They suggest that this ratio remains almost the same whatever the shipment distance, although high-quality goods are more expensive at longer shipping distances, which matches with the Alchian and Allen (1964) effect.

Table 3: Estimation of quality ratio γ

	(1)	(2)	(3)	(4)	(5)	(6)
High Quality			1.090***		1.088***	1.057***
			(0.000)		(0.000)	(0.000)
log Pair distance ij					0.279***	0.170***
					(0.002)	(0.003)
High Quality × *log Pair distance ij*						0.209***
						(0.002)
Fixed effects:						
orig. country	Yes	Yes	Yes	Yes	Yes	Yes
dest. country	Yes	Yes	Yes	Yes	Yes	Yes
years	Yes	Yes	Yes	Yes	Yes	Yes
HS4 product		Yes	Yes		Yes	Yes
High Quality × *HS4 product*				Yes		
Observations	8,694,818	8,694,818	8,694,818	8,694,818	8,694,818	8,694,818
Adjusted R^2	0.041	0.816	0.912	0.923	0.912	0.912

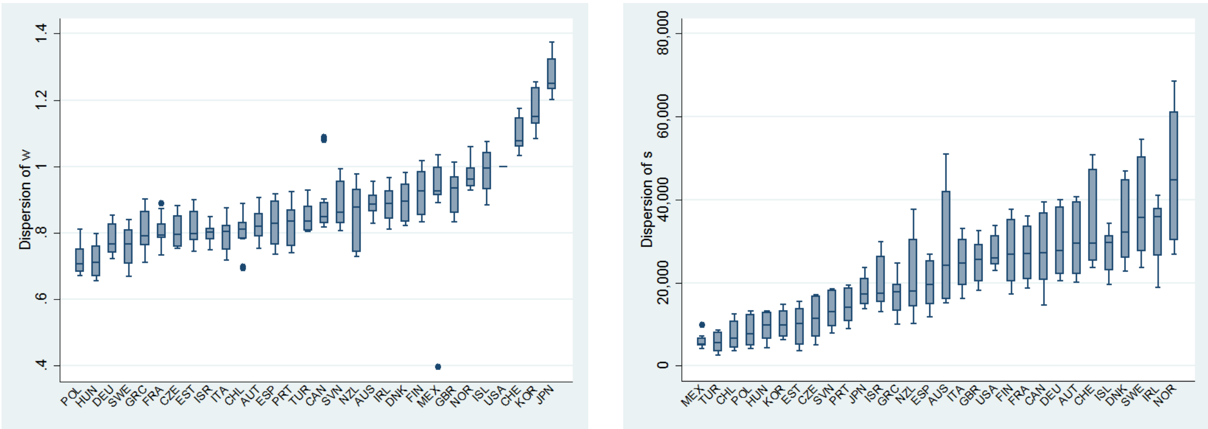
Notes: This table reports estimations of quality ratio $\gamma = \alpha/(\alpha - 1)$. Dependent variable is log of FOB unit values sourced from Trade Unit Values database. Standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.

Labor prices Table 4 reports the estimation of the price of the labor productivity units w in 2014. Observations are taken on each HS6 category for triplets with the same destination i and two alternative origins $j, l \neq i$. The dependent variable is the natural logarithm of relative unit values $\log\left(P^{\text{fob}}_{\omega,t}/P^{\text{fob}}_{\omega',t}\right)$. The independent variables of interest are "exporter dummy differences" whose coefficients estimate the log of w_i/w_{USA}, the labor productivity unit price of the last country, U.S.A., being taken as reference and given a value equal to 0 (i.e. $w_{\text{USA}} = 1$). The first column presents the regression result for those exporter dummy differences only. All estimated coefficients for $\log w_i$ are statistically significant. To control for potential omitted variables, the second column includes fixed effects for HS4 product categories, the third the quality level of each HS6 variety, and the fourth a combination of fixed effects on the number of varieties and bilateral trade costs between destinations and alternative exporters. The country dummies coefficients are robust to those controls whereas the fit substantially improves (R^2 doubles). The fourth specification is our most preferred one. The last two columns include additional fixed effects on the destination country and a combination of product and quality. Estimated coefficients are stable to those controls.

Table 4: Estimation of labor prices w (year 2014)

	(1)	(2)	(3)	(4)	(5)	(6)
D_FRA	-0.139***	-0.138***	-0.138***	-0.123***	-0.118***	-0.118***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_DEU	-0.165***	-0.176***	-0.176***	-0.159***	-0.150***	-0.150***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_IRL	-0.149***	-0.100***	-0.100***	-0.088***	-0.093***	-0.093***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_ITA	-0.168***	-0.165***	-0.164***	-0.150***	-0.145***	-0.145***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_JPN	0.078***	0.105***	0.105***	0.099***	0.086***	0.086***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_GBR	-0.097***	-0.097***	-0.097***	-0.083***	-0.076***	-0.076***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_SWE	-0.210***	-0.191***	-0.191***	-0.177***	-0.176***	-0.176***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_CHE	0.017***	0.030***	0.030***	0.045***	0.048***	0.048***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_ISR	-0.161***	-0.116***	-0.116***	-0.108***	-0.115***	-0.115***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_CAN	-0.100***	-0.078***	-0.078***	-0.077***	-0.082***	-0.082***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_FIN	-0.127***	-0.097***	-0.097***	-0.083***	-0.085***	-0.085***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_NOR	-0.052***	-0.028***	-0.028***	-0.014***	-0.013***	-0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_DNK	-0.134***	-0.111***	-0.111***	-0.096***	-0.096***	-0.096***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_AUT	-0.127***	-0.108***	-0.108***	-0.093***	-0.092***	-0.092***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_ISL	-0.126***	-0.062***	-0.062***	-0.053***	-0.061***	-0.061***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_AUS	-0.066***	-0.041***	-0.041***	-0.066***	-0.089***	-0.089***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_EST	-0.166***	-0.123***	-0.123***	-0.111***	-0.115***	-0.115***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_TUR	-0.143***	-0.116***	-0.116***	-0.105***	-0.107***	-0.107***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_POL	-0.230***	-0.218***	-0.218***	-0.202***	-0.200***	-0.200***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_HUN	-0.251***	-0.216***	-0.216***	-0.201***	-0.204***	-0.204***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_CHL	-0.133***	-0.088***	-0.088***	-0.098***	-0.118***	-0.118***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_CZE	-0.172***	-0.152***	-0.152***	-0.136***	-0.136***	-0.136***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_MEX	-0.073***	-0.041***	-0.041***	-0.046***	-0.058***	-0.058***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_SVN	-0.156***	-0.112***	-0.112***	-0.098***	-0.102***	-0.102***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_PRT	-0.197***	-0.156***	-0.156***	-0.144***	-0.149***	-0.149***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_KOR	0.034***	0.062***	0.062***	0.059***	0.047***	0.047***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_GRC	-0.224***	-0.180***	-0.180***	-0.168***	-0.173***	-0.173***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_NZL	-0.169***	-0.127***	-0.127***	-0.150***	-0.177***	-0.177***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_ESP	-0.176***	-0.165***	-0.165***	-0.152***	-0.151***	-0.151***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Controls:						
HS4 product FE		Yes	Yes	Yes	Yes	
High quality			Yes	Yes	Yes	
Bilateral trade costs and nb. of varieties				Yes	Yes	Yes
$High\ Quality \times HS4\ product$ FE						Yes
Destination country FE					Yes	Yes
Observations	14,808,280	14,808,280	14,808,280	14,808,280	14,808,280	14,808,274
Adjusted R^2	0.022	0.044	0.045	0.049	0.050	0.053

Notes: This table reports estimations of wage per unit of productivity w. Dependent variable is log of FOB unit values sourced from Trade Unit Values. The columns (3-5) include controls for High Quality dummy. The columns (4-6) include the following controls: (a) logarithm of bilateral trade costs between i and j, (b) logarithm of bilateral trade costs between i and l, (c) logarithm of number of varieties. Robust standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.



(a) Price w_i

(b) Endowment s_i

Figure 1: Dispersion of w and s by countries and years.

We re-iterate the above estimation and obtain predicted values of w_i for each year between 2000 and 2014. The estimated prices of labor productivity units w are displayed on Fig. 1(a) (corresponding values for the year 2014 are reported in Table 4). As mentioned above, the price of labor productivity units is normalized to one for the US economy. The countries are ranked by the average values of w_i across years. Those prices vary in a narrow range of $[0.4, 1.37]$. The height of each bar shows the presence of a dispersion of w_i across years. The lowest prices are for Poland and the highest ones for Switzerland. The price of the same labor productivity unit is about 3.4 times higher in the latter, which reflects a much tighter labor market there.

We recover the household endowments of labor productivity units s_i by dividing the GDP per capita with the estimated values of w_i. This is displayed in Fig. 1(b). Per-capita endowments of labor productivity units rise from about $2,500$ units in Mexico and Turkey to about $60,000$ units in Norway. This large range corresponds to the large heterogeneity in per-capita GDP across countries: from GBP $1,000$ to GBP $26,000$. Nevertheless, the price of labor productivity units absorbs a fraction of this heterogeneity. As expected, a higher household endowment in labor productivity units is observed in more industrialized countries.

Gravity Table 5 presents the econometric results from the gravity equation eq. (28) that breaks down per-capita trade values with respect to distance with their trade partners and their remoteness. The first row reports the estimation of the theoretical parameter c_0. This coefficient is statistically significant and positive in accord with the theory. It is however smaller than its expected value, which reflects the omission of the rest of the world on which a share of expenditure is spent. The second and third rows report the coefficients

Table 5: Estimation of gravity equation

	(1)	(2)	(3)	(4)	(5)
$R_{ijt} - \frac{N_{ijt}}{N^I_{it}} \times R_{it}$	-187.4***	-202.4***	-102.4***	-101.3***	-100.1***
	(10.05)	(9.603)	(9.357)	(9.361)	(9.341)
Common language		249.5***	144.6***	140.7***	135.8***
		(7.004)	(7.067)	(7.183)	(7.185)
Common border			282.3***	280.9***	275.8***
			(6.865)	(6.879)	(6.882)
Colonial links				28.90**	12.49
				(9.438)	(9.593)
Year fixed effects					Yes
Observations	13050	13050	13050	13050	13050
Adjusted R^2	0.0259	0.112	0.214	0.214	0.219

Notes: This table reports estimations of the gravity equation. The dependent variable is the per-capita trade flow sourced from BACI. Robust standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.

for geographic distance and remoteness, c_1 and c_2. These coefficients have the expected signs and very similar amplitudes, which confirms our theoretical prediction. This suggests a good model fit. The next three rows introduce typical gravity controls.[7] These controls have expected signs as colonial links and common language and borders increase trade. The control for common language does not change estimated values of c_1 and c_2 whereas the introduction of common border dummies slightly reduces them, reflecting the negative correlation between distance.

The introduction of colonial links has a limited impact. Finally, the last column introduces time-fixed effects to control for general economic fluctuations that would not be reflected in the expenditure and prices of labor productivity units (e.g., the 2008 crisis). The results do not differ from the previous specifications. Note that the fit of this gravity equation ($R^2 = 0.395$) is similar to the ones found in the literature, which does not allow us to reject the present model with vertical differentiation compared to other models. To sum up, our results from this gravity regression do not reject the properties of our vertical differentiation model according to which trade flows have similar sensitivity to geographic distance and

[7] See e.g. Gómez-Herrera (2013) for a review of methodology of gravity estimations.

Table 6: Estimation of gravity equation (constrained)

	(1)	(2)	(3)	(4)	(5)
$R_{ijt} - \frac{N_{ijt}}{N^I_{it}} \times R_{it}$	-187.4***	-202.4***	-102.4***	-101.3***	-100.1***
	(10.05)	(9.603)	(9.357)	(9.361)	(9.341)
Common language		-249.5***	-144.6***	-140.7***	-135.8***
		(7.004)	(7.067)	(7.183)	(7.185)
Common border			-282.3***	-280.9***	-275.8***
			(6.865)	(6.879)	(6.882)
Colonial links				-28.90**	-12.49
				(9.438)	(9.593)
Year fixed effects					Yes
Observations	13,050	13,050	13,050	13,050	13,050
Adjusted R^2	0.026	0.112	0.214	0.214	0.219

Notes: This table reports estimations of the constrained gravity equation. We assume that $\alpha = 1$ and $c_1 = -c_2$. The dependent variable is the per-capita trade flow sourced from BACI. Robust standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.

remoteness. This conclusion allows us to estimate more precisely the model parameter by putting the appropriate structure on estimated coefficients.[8]

We further proceed with the estimation of a constrained gravity specification that fits our theoretical model, i.e. where $\alpha = 1$ and $c_1 = -c_2$. The results are reported in Table 6 which has the same column structure as Table 5. The first row shows the estimated value for $-c_1$ (and therefore c_2). This coefficient keeps the same magnitude as in a non-constrained regression. We retain the estimate of the fourth column, $r = -\widehat{c}_2 = 102.4$, for further analysis. To give an idea of this number, one can check that a consumer will purchase 111 USD less of imports from the USA if she lives 10,000km farther away from it, keeps the same remoteness level, and import $n_{ij} = 2.41$ HS6 varieties in each HS4 good (=$102.4 * (0.45 * 1 * 2.41)$). This is to be compared with the average income of $1,062$ USD spent on imports for each country ($31,869$ USD / 30 countries).

Quality upgrade cost profile We finally turn to the estimation of the labor input parameters for the production of high- and low-quality versions of HS4 goods. We use the procedure

[8] In the appendix, we check the properties of the gravity equation in the absence of extensive margins. Towards this aim, we replicate the above analysis by setting N_{ij} to the same value equal to the average number of varieties across country pairs. Table B1 reports the same properties of regression coefficients: β and γ have opposite signs and same magnitudes. Because it does fit as well the trade flows, this regression model produces a lower R^2. The estimated β is also slightly lower.

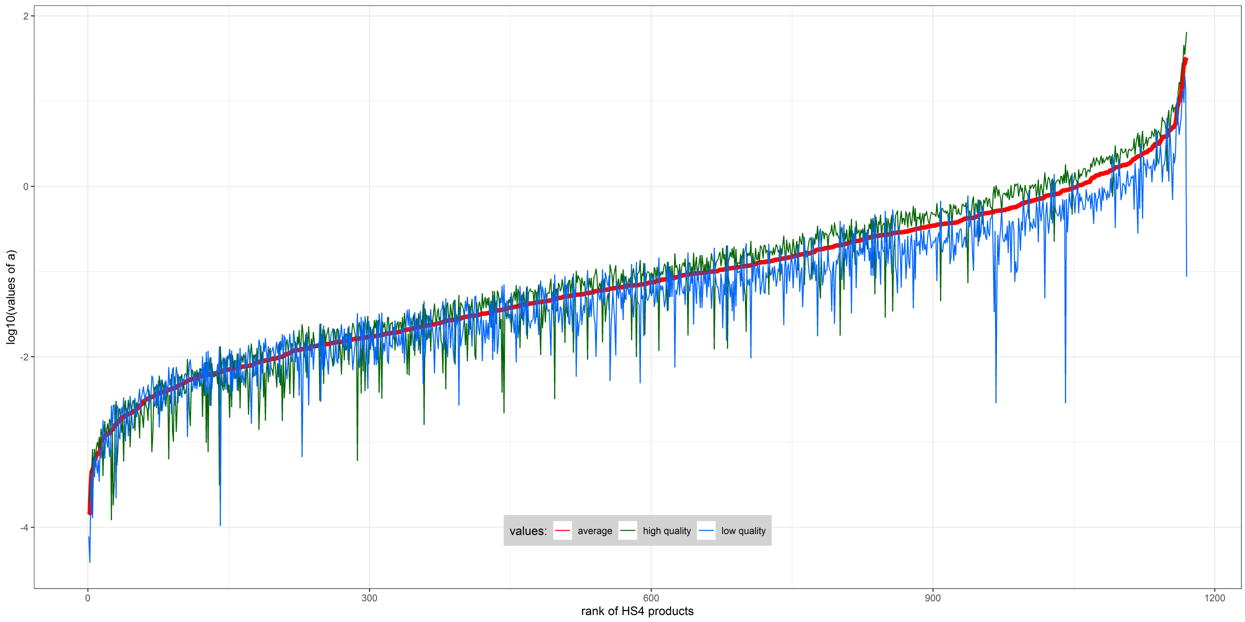


Figure 2: Distribution of values of a.

explained in the empirical specification section. Fig. 2 displays the logarithm (base 10) of those parameter values. The blue and green lines show the high- and low-quality input for each HS4 good while the red curve represents the average input. Goods are ranked by average input values. We make two remarks. On the one hand, the difference between input for high- and low-quality goods is congruent with the above finding of $\gamma = 2.97$ as the average differences between quality versions lie about 0.47 $(= \log_{10} 2.97)$. The input heterogeneity between high and low quality across goods is shown by the various spikes. This heterogeneity is depicted in Fig. 3 that presents the input of high- and low-quality versions with respect to the average input on each good. The low-quality versions require about half of the average input while high-quality ones are about half more, which implies a ratio of about 3 $(= 1.5/0.5)$. Fig. 3 also highlights the dispersion of the input needed to produce high and low-quality goods, which is not neglected in our approach.

On the other hand, Fig. 2 shows that the logarithm of average input is linear in most of its graph, reflecting a power distribution. The two first columns of Table 7 report the linear regression of the logarithm of average input with respect to the index of the good ($z =$rank$/1170$). This shows a very good fit with the data and a power of 2.96. By contrast, the two last columns consider the data match with a Pareto distribution, which should be given by the same regression with respect to $\log_{10}(1 - z)$. Since the fit is less good, we prefer the former specification.

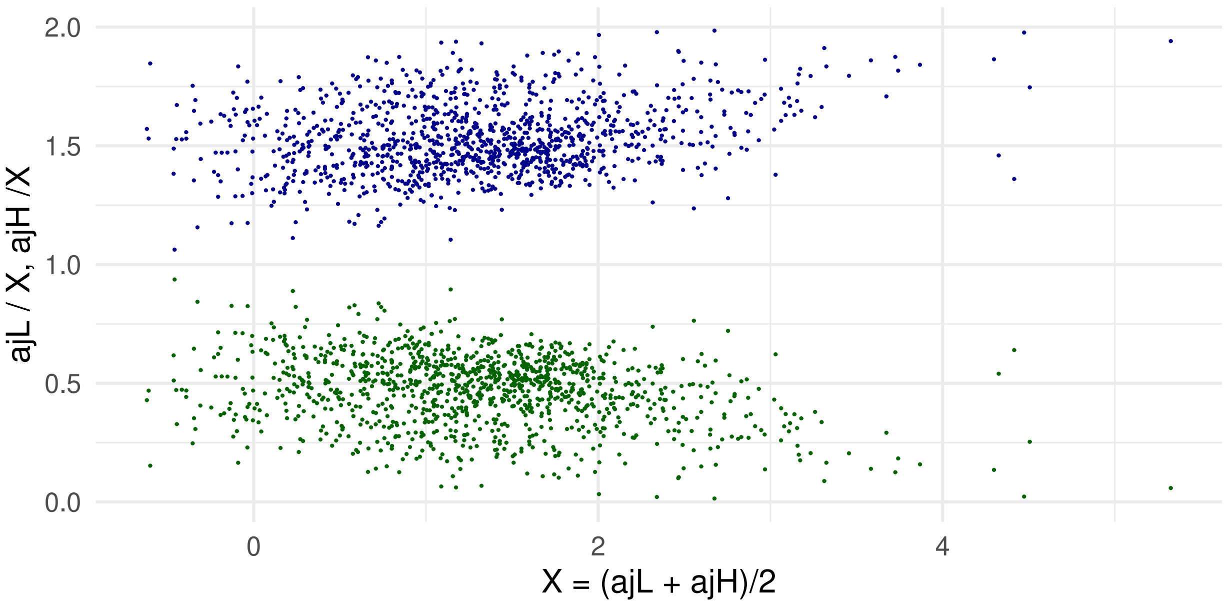


Figure 3: Difference between unit values of high-quality and low-quality products.

Table 7: Estimations of labor inputs a

	(1)	(2)	(3)	(4)
ζ	2.967***	2.966***		
	(0.004)	(0.004)		
$log_{10}(1-\zeta)$			-1.827***	-1.825***
			(0.006)	(0.006)
Constant	-2.696***	-2.697***	-2.006***	-2.022***
	(0.002)	(0.005)	(0.004)	(0.011)
Fixed effects:				
years		Yes		Yes
Observations	17,619	17,619	17,619	17,619
Adjusted R^2	0.964	0.964	0.825	0.825

Notes: This table reports estimations of the number of labor units a. The dependent variable is the base 10 logarithm of average a. Columns 2 and 4 include a fixed effect for years. Robust standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.

6 Quantification

In this section, we quantify our model on the set of countries investigated in the empirical section. We first use our empirical estimations to calibrate parameters and cost profiles of the model. We then run a benchmark simulation and check the correspondence with actual data. We then proceed to the study of a series of economic shocks and trade policy reforms.

6.1 Calibration

Given the above empirical analysis, we set the values of trade costs to the predictions obtained in the regression model reported in Table 2 (column 6) and using actual distances, contiguities, languages, and colonial ties. We also set the value of the quality-cost ratio to $\alpha = 1.5$, which rounds to the value estimated in Table 3 (column 3). We set the country average endowments of labor productivity units s_i to the values underlying Fig. 1 (sub-figure b) for the year 2014. We set $r = 102.4$ to the value estimated in the gravity equation.

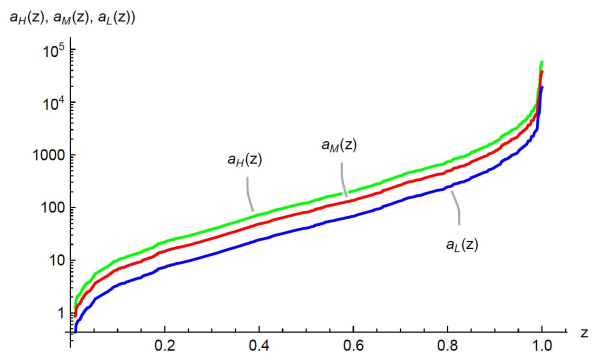
We also recover the cost schedule of high and low-quality varieties and the schedule of per-quality-unit labor input needed to upgrade each variety z from the sorted profile of estimated average costs $a_M = (a_H + a_L)/2$ presented in Fig. 2 (red curve). We apply the value of quality-cost ratio to $\alpha = 1.5$ and use the relationships $a_H = \alpha a$, $a_M = a(2\alpha - 1)/2$ and $a_L = (\alpha - 1)a$ to get $a_L = a_M * 2(\alpha - 1)/(2\alpha - 1)$ and $a_H = a_M * 2\alpha/(2\alpha - 1)$. Those profiles are shown in the left panel of Fig. 4. To recover the per-quality-unit labor input schedule, we use the value of r and identity eq. (9) to get:[9]

$$\ell(z) = r + \frac{2(\alpha - 1)}{2\alpha - 1}\int_0^1 a_M(\xi)\,\mathrm{d}\xi + \frac{2}{2\alpha - 1}\int_0^z a_M(\xi)\mathrm{d}\xi. \tag{29}$$

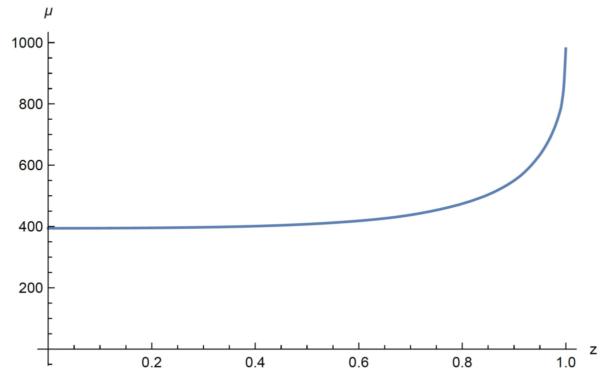
This expresses the quality input schedule as a function of the average input cost profile and the basic parameters α and r and the profile of average costs $a_M(z)$. This schedule is displayed in the right panel of Fig. 4. As can be seen, the schedule is bounded below and above. An importer with $\mu_i/(w_j\tau_{ij})$ lying between $\ell(0) = 382.9$ and $\ell(1) = 946.7$ will consume a mix of high- and low-quality varieties It will import no high-quality varieties for values below that range and only high-quality varieties for values above.

Finally, we set the number of varieties n_{ij} to their average number across all country pairs times the producer population shares in 2014. That is $n_{ij} = (1/N^2)\sum_i\sum_j N_{ij,2014} * m_{j,2014}$. We cannot indeed use the estimated number of varieties of $N_{ij,2014}$ as suggested by our model. This is because $N_{ij,2014}$ has a too strong variance that blurs the simulated equilibrium outcomes. $N_{ij,2014}$ is not enough correlated with population sizes and implies too much distortion between each country's labor supply and the level of the global demand for the country's production. Imposing a number of varieties correlated with population size

[9] Indeed, note that using $\ell'(z) = a(z)$, we get $r = \alpha\ell(0) - (\alpha - 1)\ell(1) = \ell(0) - (\alpha - 1)[\ell(1) - \ell(0)] = \ell(0) - (\alpha - 1)\int_0^1 \ell'(\xi)\,\mathrm{d}\xi = \ell(0) - (\alpha - 1)\int_0^1 a(\xi)\,\mathrm{d}\xi$. Solving for $\ell(0)$ and plugging into eq. (9) we have $\ell(z) = r + (\alpha - 1)\int_0^1 a(\xi)\,\mathrm{d}\xi + \int_0^z a(\xi)\mathrm{d}\xi$. Finally, using $a = 2a_M/(2\alpha - 1)$, this gives the result.



(a) Calibration of high- and low- quality input-cost schedule (USD)



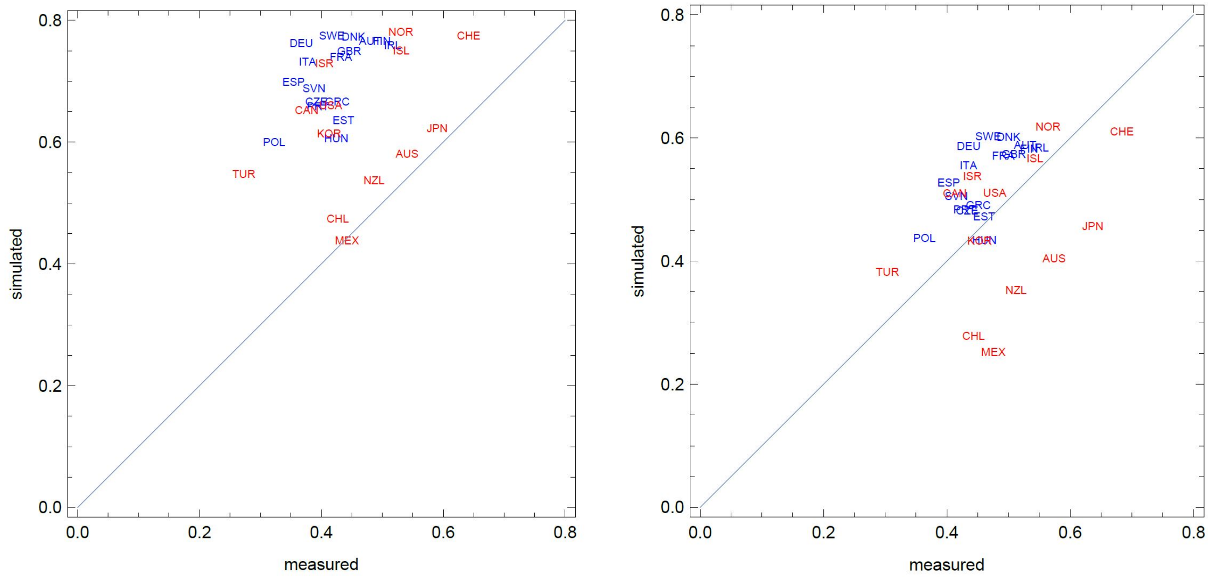
(b) Input-per quality schedule

Figure 4: Costs of high and low varieties and per-quality-unit labor input schedule.

follows Krugman (1993)'s spirit according to which larger countries stimulate the entry of more numerous firms and varieties.

6.2 Benchmark

Using those values, we find the equilibrium labor productivity prices by solving the system of linear equations eq. (22). We then recover the inverse of the country marginal utilities of income μ_i and the values of $\mu_i/(\tau_{ij}w_j)$ that determine the shares of high-quality goods in imports for each country pair. Those simulated shares are calculated in quantity (volume) and value according to eq. (16) and eq. (18). To validate the model, we compare those simulated shares with those of the measured share by country. Towards this aim, we average the country-pair shares by importing country. Fig. 5 plots those simulated shares as a function of the measured ones. The blue color represents the EU countries. Whereas the simulated and measured shares do not match perfectly (on the 45° line), they display a clear positive correlation, which validates the model.



(a) Shares in quantity (b) Shares in value

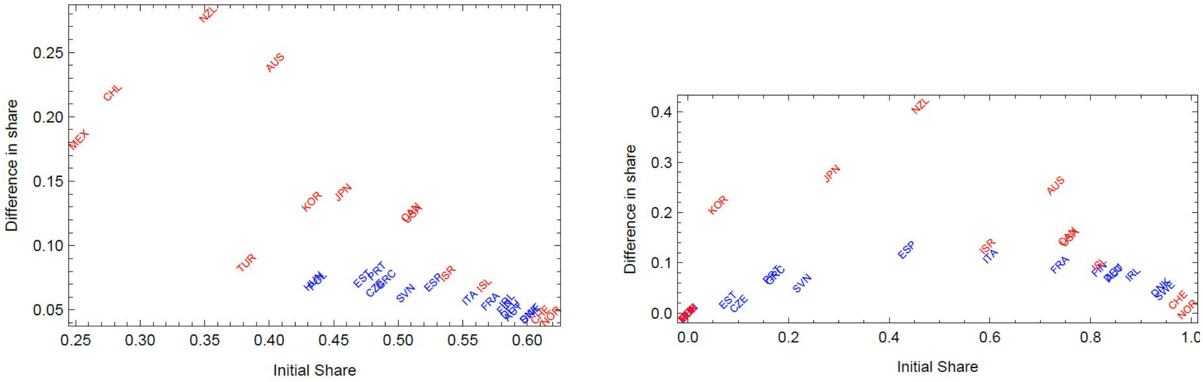
Figure 5: Simulated and estimated shares of high-quality goods across countries.

6.3 Policy reforms

We now apply our framework to study the effects of economic shocks and policy reforms. Since the focus of this paper is on vertical differentiation, we present their effect on the share of high-quality goods in values. We also focus on the UK because we include an analysis of Brexit.

6.3.1 Zero trade cost

The trade literature emphasizes the benefit of economic integration whereby countries remove the barriers to trade. In this counterfactual, we assume that all barriers to trade, including the transport costs. That is, $\tau_{ij} = 1$. In Fig. 6, we present the impact on the share of high-quality goods for goods that are imported (left hand) and exported (right hand). As expected, all countries increase their import and export values of high-quality goods. The increase in the share of imported high-quality varieties lies above 15% in the most isolated countries and/or less productive countries. This share does not increase more than 10% in the European countries (blue color), the latest entrants having higher gains. This is explained by their proximity and level of economic integration. The share of exported high-quality varieties increases everywhere, but more for non-EU countries (blue color). The highest gain in export value is attributed to New Zealand, which is explained by the elimination of its remote geographical position. Countries with low shares high quality goods gain more if they are more remote and productive. Countries with high shares gain less as they already produce at the top of the quality ladder (this model with two types does not allow to discuss much about quality ladder).



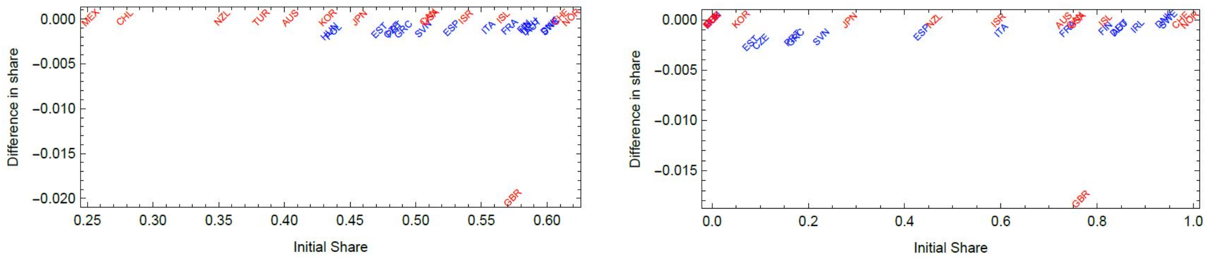
(a) Import value (b) Export value

Figure 6: Towards zero trade Costs: country share of high-quality varieties.

6.3.2 Soft and hard Brexit

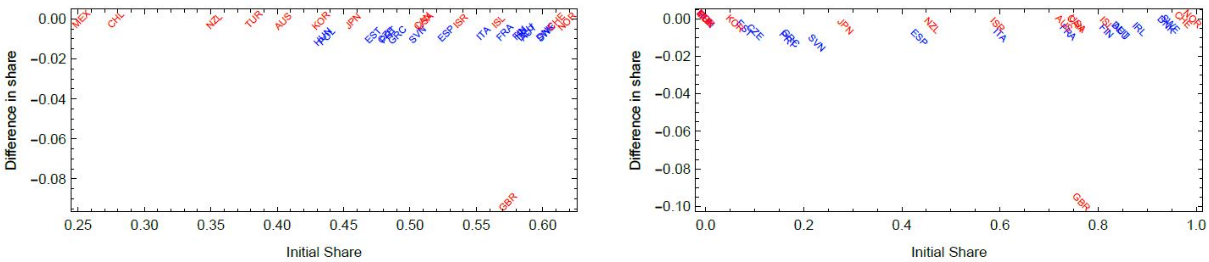
In January 2020, the UK began the implementation of its withdrawal from the EU. This followed its 2016 referendum on the continuation of its EU membership, which showed a (slight) majority in favor of the British exit (”Brexit”). Between and after those two dates, much economic research has assessed the economic effects of this political decision (Dhingra et al. (2017) amongst others). We here offer a new perspective of quantification of the Brexit effect in light of our quantitative model of vertical differentation. We first present the case of a ”soft” Brexit, wherein the UK keeps its economic ties. We have in mind the actual outcome of the 2020 negotiations that concluded a Trade and Cooperation Agreement. The latter promoted tariff-free and quota-free trade in all products. Meanwhile, the UK is no longer part of the EU Single Market, which implies a rise in the non-tariff barriers. As a result, trade suffers from trade costs caused by the installation of new customs checks, paperwork, congestion, delays, product norm compatibility issues, etc. In this counterfactual, we suppose that the former trade costs between the UK and EU partners are proportionally increased by 5%; that is, τ_{ij} becomes $\tau_{ij} * 1.05$ for all $i = UK$ and $j =$EU countries and vice-versa.

Fig. 7 shows those changes for the country’s average shares of high-quality imports and exports. The UK average shares of high-quality varieties drop by about 2% both at import and export. The increase in trade costs of the soft Brexit obliges the UK consumers to consume lower quality goods and UK firms to sell varieties of lower quality to match competitors in the EU Non-EU countries (red color) are unaffected whereas EU countries (blue color) experience a deterioration of their imports and exports of high-quality goods by



(a) Import value (b) Export value

Figure 7: Towards a soft Brexit: share of high-quality in value and across countries.



(a) Import value (b) Export value

Figure 8: Towards a hard Brexit: share of high-quality in value and across countries.

less than 0.25 percent. This deterioration is larger amongst the EU countries with initially small shares in their exports of high-quality varieties (e.g. Estonia, Czech Republic). Yet, such a small deterioration of the average shares of high-quality goods in the trade flows is unlikely to be observable in all those countries, except in the UK

The tense negotiations between the UK and EU representatives have many times put forward the possibility of a hard Brexit in which the UK would have been considered as a third country submitted to higher tariffs and stronger customs control. We suppose that this outcome would raise the trade cost between the UK and the EU by 10% more than the soft Brexit. That is, τ_{ij} becomes $\tau_{ij} * 1.15$ for all $i = UK$ and $j =$EU countries and vice-versa. Its effect is presented in Fig. 8. This time, the share of high-quality varieties drops by about 10%. The fall in high-quality imports by EU countries lies between 1 and 2%.

This exercise is related to the numerous assessments of the Brexit shock summarized by Dhingra and Sampson (2022). It is pointed out that the leave decision has hurt the UK GDP, even before the Brexit date. The total GDP loss was estimated at 2-3% at the end of 2019. Freeman et al. (2022) estimate that the new trade agreement caused a 25% fall in imports from the EU, and a smaller and temporary decline in the UK exports to the EU. They further point to the asymmetry in the adjustment to integration and disintegration shocks. Yet, those studies do not describe the impact of product quality. The present exercise makes clear that the trade costs implied by Brexit should be high to be observed in average product baskets.

7 Conclusion

In this paper, we study the quantitative properties of the trade equilibrium of a model with vertical differentiation and heterogeneous goods, many countries, and within-country heterogeneous income. We use the class of costs and preferences which renders expenditures linear in income and therefore makes the general equilibrium model highly tractable. We estimate the model parameters with OECD countries' trade data. Toward this aim, we estimate a gravity equation and the unit prices of high and low-quality goods. This procedure allows us to quantify the effect of trade shocks and reforms such as Brexit. It also permits the assessment of the welfare impact of many counterfactual exercises. Finally, the paper offers a remarkable methodological contribution as it builds its quantification exclusively only trade data, without any import from other sources from other data or literature bits.

References

Alchian, A. A. and W. R. Allen, *University Economics*, Belmont, California: Wadsworth, 1964.

Armington, Paul S, "A Theory of Demand for Products Distinguished by Place of Production," *Staff Papers-International Monetary Fund*, 1969, pp. 159–178.

Baldwin, Richard and James Harrigan, "Zeros, quality, and space: Trade theory and trade evidence," *American Economic Journal: Microeconomics*, 2011, *3* (2), 60–88.

Bernasconi, Claudia and Tobias Wuergler, "Per Capita Income and the Quality and Variety of Imports," Mimeo, University of Zurich 2012.

Berthou, Antoine and Charlotte Emlinger, "The trade unit values database," *Economie internationale*, 2011, *128* (4), 97–117.

Choi, Yo Chul, David Hummels, and Chong Xiang, "Explaining import quality: The role of the income distribution," *Journal of International Economics*, 2009, *77* (2), 265–275.

Conte, Maddalena, Pierre Cotterlaz, Thierry Mayer et al., *The CEPII gravity database*, CEPII, 2022.

Dhingra, Swati and Thomas Sampson, "Expecting brexit," *Annual Review of Economics*, 2022, *14*, 495–519.

__ , **Hanwei Huang, Gianmarco Ottaviano, João Paulo Pessoa, Thomas Sampson, and John Van Reenen**, "The costs and benefits of leaving the EU: trade effects," *Economic Policy*, 2017, *32* (92), 651–705.

Dingel, Jonathan I, "The determinants of quality specialization," *The Review of Economic Studies*, 2017, *84* (4), 1551–1582.

Eaton, Jonathan and Ana Cecilia Fieler, "The Margins of Trade," *NBER Working Paper No. 26124*, 2019.

Fajgelbaum, Pablo D and Amit K Khandelwal, "Measuring the unequal gains from trade," *The Quarterly Journal of Economics*, 2016, *131* (3), 1113–1180.

Fajgelbaum, Pablo, Gene M Grossman, and Elhanan Helpman, "Income distribution, product quality, and international trade," *Journal of political Economy*, 2011, *119* (4), 721–765.

__ , __ , **and** __ , "A Linder hypothesis for foreign direct investment," *The Review of Economic Studies*, 2015, *82* (1), 83–121.

Feenstra, Robert C and John Romalis, "International prices and endogenous quality," *The Quarterly Journal of Economics*, 2014, *129* (2), 477–527.

Fieler, Ana Cecilia, "Nonhomotheticity and bilateral trade: Evidence and a quantitative explanation," *Econometrica*, 2011, *79* (4), 1069–1101.

Flam, Harry and Elhanan Helpman, "Vertical product differentiation and North-South trade," *The American Economic Review*, 1987, pp. 810–822.

Fouquin, Michel, Jules Hugot et al., "Two centuries of bilateral trade and gravity data: 1827-2014," Technical Report, Universidad Javeriana-Bogotá 2016.

Freeman, Rebecca, Kalina Manova, Thomas Prayer, Thomas Sampson et al., *Unravelling deep integration: UK trade in the wake of Brexit*, Centre for Economic Performance, London School of Economics, 2022.

Gaulier, Guillaume and Soledad Zignago, "Baci: international trade database at the product-level (the 1994-2007 version)," Working Paper 2010-23, CEPII 2010.

Gervais, Antoine, "Product quality and firm heterogeneity in international trade," *Canadian Journal of Economics/Revue canadienne d'économique*, 2015, *48* (3), 1152–1174.

Gómez-Herrera, Estrella, "Comparing alternative methods to estimate gravity models of bilateral trade," *Empirical economics*, 2013, *44* (3), 1087–1111.

Gorman, William M, "Separable utility and aggregation," *Econometrica: Journal of the Econometric Society*, 1959, pp. 469–481.

Hallak, Juan Carlos, "Product quality and the direction of trade," *Journal of international Economics*, 2006, *68* (1), 238–265.

_ **and Peter K Schott**, "Estimating cross-country differences in product quality," *The Quarterly journal of economics*, 2011, *126* (1), 417–474.

Jaimovich, Esteban and Vincenzo Merella, "Quality ladders in a Ricardian model of trade with nonhomothetic preferences," *Journal of the European Economic Association*, 2012, *10* (4), 908–937.

_ **and** _ , "Love for quality, comparative advantage, and trade," *Journal of International Economics*, 2015, *97* (2), 376–391.

Khandelwal, Amit, "The long and short (of) quality ladders," *The Review of Economic Studies*, 2010, *77* (4), 1450–1476.

Khandelwal, Amit K, Peter K Schott, and Shang-Jin Wei, "Trade liberalization and embedded institutional reform: evidence from Chinese exporters," *American Economic Review*, 2013, *103* (6), 2169–95.

Krugman, Paul, "Scale economies, product differentiation, and the pattern of trade," *The American Economic Review*, 1980, *70* (5), 950–959.

Krugman, Paul R, "On the relationship between trade theory and location theory," *Review of international economics*, 1993, *1* (2), 110–122.

Mussa, Michael and Sherwin Rosen, "Monopoly and product quality," *Journal of Economic theory*, 1978, *18* (2), 301–317.

Picard, Pierre M and Alessandro Tampieri, "Quality and Trade with Many Countries and Industries," Mimeo, University of Luxembourg 2021.

_ **and** _ , "Vertical differentiation and trade among symmetric countries," *Economic Theory*, 2021, *71*, 1319–1355.

Piveteau, Paul and Gabriel Smagghue, "Estimating firm product quality using trade data," *Journal of International Economics*, 2019.

Pollak, Robert A, "Conditional demand functions and consumption theory," *The Quarterly Journal of Economics*, 1969, *83* (1), 60–78.

Verhoogen, Eric A, "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," *The Quarterly Journal of Economics*, 2008, *123* (2), 489–530.

Appendix A. Theoretical model

Linear real expenditure

Let us first remind the definitions of the additional cost and benefit of a quality upgrade: $a(z) = a_H(z) - a_L(z)$ and $b(z) = b_H(z) - b_L(z)$. So, the per-quality input schedule is written as $\ell(z) = a(z)/b(z)$. Then, the real expenditure successively writes as

$$\begin{aligned} E(y) &= \int_{\mathcal{H}(y)} a_H(\zeta)\mathrm{d}z + \int_{\mathcal{L}(y)} a_L(\zeta)\mathrm{d}\zeta \\ &= \int_0^{\ell^{-1}(y)} a(\zeta)\mathrm{d}\zeta + \int_0^1 a_L(\zeta)\mathrm{d}\zeta \\ &= \int_{\ell(0)}^{y} \frac{a(\ell^{-1}(y))}{\ell'(\ell^{-1}(y))}\mathrm{d}y + \int_{\ell(0)}^{\ell(1)} \frac{a_L(\ell^{-1}(y))}{\ell'(\ell^{-1}(y))}\mathrm{d}y. \end{aligned} \tag{A1}$$

where we substitute z by $\ell^{-1}(y)$. The assumption of linear real expenditure writes as $E'(y) = 1$ and imposes that the term within the first integral of the above expression is equal to 1. Therefore $\ell'(\ell^{-1}(y)) = a(\ell^{-1}(y))$, which gives

$$\ell'(z) = a(z). \tag{A2}$$

Using eq. (A2) and proportionate upgrade cost the real expenditure successively writes as

$$\begin{aligned} E(y) &= \int_{\ell(0)}^{y} \mathrm{d}y + \int_{\ell(0)}^{\ell(1)} \frac{a_L(\ell^{-1}(y))}{a(\ell^{-1}(y))}\mathrm{d}z \\ &= \int_{\ell(0)}^{y} \mathrm{d}y + \int_{\ell(0)}^{\ell(1)} (\alpha - 1)\mathrm{d}z \\ &= (y - \ell(0)) + (\alpha - 1)(\ell(1) - \ell(0)) \\ &= y - r \end{aligned} \tag{A3}$$

where

$$r = \alpha\ell(0) - (\alpha - 1)\,\ell(1).$$

Integrating eq. (A2) and using $\ell(0) = a(0)/b(0)$, we obtain the following per-quality input schedule:

$$\ell(z) = \frac{a(0)}{b(0)} + \int_0^z a'(\zeta)\mathrm{d}\zeta \tag{A4}$$

Finally, differentiating $\ell(z) = a(z)/b(z)$ and plugging eq. (A2) in this expression gives the differential equation

$$\frac{a'(z)}{a(z)} = \frac{b'(z)}{b(z)} + b(z). \tag{A5}$$

which accepts the solution

$$b(z) = \frac{a(z)}{a(0)/b(0) + \int_0^z a(\zeta)d\zeta}. \tag{A6}$$

Equilibrium utility

The equilibrium utility is given by

$$\begin{aligned} V_i &= \sum_{j=1}^{N} n_{ij} \int_{\mathcal{H}\left(\frac{\mu_i}{\tau_{ij} w_j}\right)} b_H(\zeta)\mathrm{d}\zeta + n_{ij} \int_{\mathcal{L}\left(\frac{\mu_i}{\tau_{ij} w_j}\right)} b_L(z)\mathrm{d}\zeta \\ &= \sum_{j=1}^{N} n_{ij} V\left(\frac{\mu_i}{\tau_{ij} w_j}\right) \end{aligned} \tag{A7}$$

where

$$V(y) = \int_0^{\ell^{-1}(y)} b(\zeta)\mathrm{d}\zeta + \int_0^1 b_L(\zeta)\mathrm{d}\zeta$$

We successively have

$$
\begin{aligned}
V(y) &= \int_0^{\ell^{-1}(y)} b(\zeta)\mathrm{d}\zeta + \int_0^1 b_L(\zeta)\mathrm{d}\zeta \qquad \text{(A8)}\\
&= \int_{\ell(0)}^{y} \frac{b(\ell^{-1}(y))}{\ell'(\ell^{-1}(y))}\mathrm{d}y + \int_{\ell(0)}^{\ell(1)} \frac{b_L(\ell^{-1}(y))}{\ell'(\ell^{-1}(y))}\mathrm{d}z\\
&= \int_{\ell(0)}^{y} \frac{b(\ell^{-1}(y))}{a(\ell^{-1}(y))}\mathrm{d}y + \int_{\ell(0)}^{\ell(1)} \frac{b_L(\ell^{-1}(y))}{a(\ell^{-1}(y))}\mathrm{d}z\\
&= \int_{\ell(0)}^{y} \frac{1}{\ell(\ell^{-1}(y))}\mathrm{d}y + \int_{\ell(0)}^{\ell(1)} \frac{b_L(\ell^{-1}(y))}{b(\ell^{-1}(y))}\frac{1}{\ell(\ell^{-1}(y))}\mathrm{d}z\\
&= \int_{\ell(0)}^{y} \frac{1}{y}\mathrm{d}y + \int_{\ell(0)}^{\ell(1)} \frac{b_L(\ell^{-1}(y))}{b(\ell^{-1}(y))}\frac{1}{y}\mathrm{d}z\\
&= \int_{\ell(0)}^{y} \frac{1}{y}\mathrm{d}y + \int_{\ell(0)}^{\ell(1)} (\beta-1)\frac{1}{y}\mathrm{d}z\\
&= (\ln y - \ln \ell(0)) + (\beta-1)(\ln \ell(1) - \ln \ell(0))\\
&= \ln y - \beta \ln \ell(0) + (\beta-1)\ln \ell(1)
\end{aligned}
$$

where we substitute z by $\ell^{-1}(y)$ in the second equality, substitute $\ell'(z)$ by $a(z)$ in the third one, use $\ell(\ell^{-1}(y)) = y$ in the fifth one, and use $b_L = (\beta-1)b$ in the sixth one.

Trade balance

The trade balance writes as

$$
\sum_{j\neq i} m_i \tau_{ij} w_j E\left(\frac{\mu_i}{\tau_{ij}w_j}\right) n_{ij} = \sum_{j\neq i} m_j \tau_{ji} w_i E\left(\frac{\mu_j}{\tau_{ji}w_i}\right) n_{ji},\ \forall i. \qquad \text{(A9)}
$$

Inserting $m_i \tau_{ii} w_i E\left(\frac{\mu_i}{\tau_{ii}w_i}\right) n_{ii}$ on both sides, this gives

$$
\sum_{j} m_i \tau_{ij} w_j E\left(\frac{\mu_i}{\tau_{ij}w_j}\right) n_{ij} = \sum_{j} m_j \tau_{ji} w_i E\left(\frac{\mu_j}{\tau_{ji}w_i}\right) n_{ji},\ \forall i. \qquad \text{(A10)}
$$

Replacing $E(y)$ by $y - r$, we obtain

$$\sum_j m_i \left(\mu_i - r\tau_{ij} w_j\right) n_{ij} = \sum_j m_j \left(\mu_j - r\tau_{ji} w_i\right) n_{ji}, \; \forall i. \tag{A11}$$

Taking into account aggregation $\mu_i = \int \mu_{ih} \mathrm{d}G$, the identity eq. (13) is equivalent to

$$\mu_i = \frac{w_i s_i}{n_i} + \frac{r}{n_i} \sum_{l=1}^{N} \tau_{il} w_l n_{il} \tag{A12}$$

So, we successively replace in eq. (A11) and get

$$\sum_j m_i n_{ij} \left(\frac{w_i s_i}{n_i} + \frac{r}{n_i} \sum_{l=1}^{N} \tau_{il} w_l n_{il} - r\tau_{ij} w_j \right) = \sum_j m_j n_{ji} \left(\frac{w_j s_j}{n_j} + \frac{r}{n_j} \sum_{l=1}^{N} \tau_{jl} w_l n_{jl} - r\tau_{ji} w_i \right), \; \forall i. \tag{A13}$$

After simplification, we get

$$w_i \left(m_i s_i + r \sum_j m_j n_{ji} \tau_{ji} \right) = \sum_j \frac{m_j n_{ji}}{n_j} \left(w_j s_j + r \sum_{l=1}^{N} \tau_{jl} w_l n_{jl} \right) \tag{A14}$$

Per-quality input schedule

We get a more general insight by using

$$\begin{aligned} \ell(z) &= \frac{a_H(0) - a_L(0)}{b_0} + \int_0^z \left(a_H(z) - a_L(z)\right) \mathrm{d}z \\ r &= \alpha \ell(0) - (\alpha - 1)\ell(1) \end{aligned} \tag{A15}$$

So, we get

$$\begin{aligned} \ell(z) &= r + (\alpha - 1) \int_0^1 \left(a_H(z) - a_L(z)\right) \mathrm{d}z + \int_0^z \left(a_H(z) - a_L(z)\right) \mathrm{d}z \\ &= r + \alpha \int_0^1 \left(a_H(z) - a_L(z)\right) \mathrm{d}z - \int_z^1 \left(a_H(z) - a_L(z)\right) \mathrm{d}z \\ &= r + \int_0^1 a_H(z) \mathrm{d}z - \frac{1}{\alpha} \int_z^1 a_H(z) \mathrm{d}z \\ &= r + \frac{2\alpha}{2\alpha - 1} \int_0^1 a_M(z) \mathrm{d}z - \frac{2}{2\alpha - 1} \int_z^1 a_M(z) \mathrm{d}z \end{aligned} \tag{A16}$$

where the last line hold for proportional cost profiles $a_H/a_L = \alpha/(\alpha-1) \iff a_H(z) - a_L(z) = \frac{1}{\alpha}a_H(z) \iff a_M(z) = (a_H(z) + a_L(z))/2 = a_H(z)\frac{(2\alpha-1)}{2\alpha} \iff a_H(z) - a_L(z) = \frac{2}{2\alpha-1}a_M(z)$. Therefore the profile $a_H(z)/r$ and $a_M(z)/r$ should not be too low in order to have a quality schedule $\ell(z)$ that is not too flat.

Share of high quality goods

The share of high-quality goods in the cif value of trade flow from i to j is

$$SHV_{ij} = \frac{1}{E_{ij}} \int n_{ij} \left(\int_0^{\widehat{z}_{ijh}} \tau_{ij} w_j a_H(z) \mathrm{d}z \right) \mathrm{d}F_i(s_{ih}), \tag{A17}$$

where $\widehat{z}_{ijh}$ solves $\ell(\widehat{z}_{ijh}) = \mu_{ih}/(\tau_{ij}w_j)$. This sums up every household's expenditure on each high quality good and divides the results by the total expenditure. Substituting z by $\ell^{-1}(y)$ so that $\mathrm{d}z = \frac{1}{\ell'(\ell^{-1}(y))}\mathrm{d}y$ and using the facts that $\ell'(z) = a(z)$ and $a_H(z)/a(z) = \alpha$ where $a(z) = a_H(z) - a_L(z)$, the above expression successively reduces to

$$\begin{aligned} SHV_{ij} &= \frac{n_{ij}\tau_{ij}w_j}{E_{ij}} \int \left(\int_{\ell(0)}^{\ell(\widehat{z}_{ijh})} a_H(\ell^{-1}(y)) \frac{1}{\ell'(\ell^{-1}(y))} \mathrm{d}y \right) \mathrm{d}F_i(s_{ih}) \\ &= \frac{n_{ij}\tau_{ij}w_j}{E_{ij}} \int \left(\int_{\ell(0)}^{\ell(\widehat{z}_{ijh})} \alpha \frac{a(\ell^{-1}(y))}{\ell'(\ell^{-1}(y))} \mathrm{d}y \right) \mathrm{d}F_i(s_{ih}) \\ &= \frac{n_{ij}\tau_{ij}w_j}{E_{ij}} \int \left(\int_{\ell(0)}^{\mu_{ih}/(\tau_{ij}w_j)} \alpha \mathrm{d}y \right) \mathrm{d}F_i(s_{ih}) \end{aligned} \tag{A18}$$

This gives

$$\begin{aligned} SHV_{ij} &= \frac{\alpha n_{ij}\tau_{ij}w_j}{E_{ij}} \int \left(\int_{\ell(0)}^{\frac{\mu_{ih}}{\tau_{ij}w_j}} \mathrm{d}y \right) \mathrm{d}F_i(s_{ih}) \\ &= \frac{\alpha w_j \tau_{ij} n_{ij}}{E_{ij}} \int \max\left\{0, \min\left[\frac{\mu_{ih}}{\tau_{ij}w_j} - \ell(0), \ell(1) - \ell(0)\right]\right\} \mathrm{d}F_i(s_{ih}) \end{aligned} \tag{A19}$$

where the last identity imposes that $\mu_{ih}/(\tau_{ij}w_j)$ belongs to the interval $[\ell(0), \ell(1)]$. In particular, when all individuals consume a mix of high and low-quality varieties (i.e.

$\ell(0) < \mu_{ih}/\left(\tau_{ij}w_j\right) < \ell(1)$), this expression simplifies to

$$
\begin{aligned}
SHV_{ij} &= \frac{\alpha w_j \tau_{ij} n_{ij}}{E_{ij}} \int \left(\frac{\mu_i}{\tau_{ij}w_j} - \ell(0)\right) \mathrm{d}F_i(s_{ih}) \\
&= \frac{\alpha n_{ij} w_j \tau_{ij}}{E_{ij}} \left(\frac{\mu_i}{\tau_{ij}w_j} - \ell(0)\right) \\
&= \alpha \frac{\frac{\mu_i}{\tau_{ij}w_j} - \ell(0)}{E\left(\frac{\mu_i}{\tau_{ij}w_j}\right)} \\
&= \alpha \frac{\frac{\mu_i}{\tau_{ij}w_j} - \ell(0)}{\frac{\mu_i}{\tau_{ij}w_j} - r}
\end{aligned}
\tag{A20}
$$

This requires in statistics on μ_i, not the household statistics μ_{ih}. This is increasing in $\frac{\mu_i}{\tau_{ij}w_j}$ if $\ell(0) > r$, which holds true. This lies between zero and one if $\ell(0) \leq \mu_{ih}/\left(\tau_{ij}w_j\right) \leq \frac{\alpha\ell(0)-r}{(\alpha-1)} = \ell(1)$, which holds true under the above assumption.

Appendix B: Additional empirical estimations

In this appendix, we report the estimations of the gravity equation in the absence of extensive margins.

Table B1: Estimation of gravity equation – no extensive margin

	(1)	(2)	(3)	(4)	(5)
$GDP\ per\ capita = I/N_i$	0.571***	0.575***	0.661***	0.662***	0.611***
	(0.0229)	(0.0220)	(0.0209)	(0.0209)	(0.0224)
$Pair\ distance = R_{ijt}$	-193.8***	-210.7***	-131.8***	-130.6***	-129.5***
	(10.72)	(10.33)	(9.948)	(9.955)	(9.940)
$Remoteness = R_{it}/N_i$	210.2***	233.8***	186.4***	187.9***	150.6***
	(11.28)	(10.88)	(10.34)	(10.35)	(11.68)
Common language		277.7***	187.0***	182.8***	181.8***
		(8.583)	(8.417)	(8.569)	(8.556)
Common border			351.8***	350.0***	353.3***
			(8.821)	(8.846)	(8.850)
Colonial links				32.24**	35.04**
				(12.34)	(12.32)
Year fixed effects					Yes
Observations	13050	13050	13050	13050	13050
Adjusted R^2	0.187	0.247	0.329	0.329	0.332

Notes: This table reports estimations of the gravity equation without extensive margin. The dependent variable is the per-capita trade flow sourced from BACI. Robust standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.

Table B2: Estimation of gravity equation – no extensive margin, constrained

	(1)	(2)	(3)	(4)	(5)
$R_{ijt} - \frac{N_{ijt}}{N_{it}^{I}} \times R_{it}$	-199.7***	-216.0***	-133.1***	-133.2***	-125.2***
	(10.99)	(10.55)	(10.05)	(10.05)	(10.05)
Common language		294.7***	186.8***	188.8***	178.7***
		(8.710)	(8.515)	(8.610)	(8.649)
Common border			326.7***	331.3***	367.3***
			(7.533)	(8.049)	(8.886)
Colonial links				-17.84	34.96**
				(11.03)	(12.34)
Year fixed effects					Yes
Observations	13050	13050	13050	13050	13050
Adjusted R^2	0.0246	0.103	0.216	0.216	0.221

Notes: This table reports constrained estimations of the gravity equation without extensive margin. We assume that $\alpha = 1$ and $c_1 = -c_2$. The dependent variable is the per-capita trade flow sourced from BACI. Robust standard errors are in parentheses. ***, ** and * denote significance at the 0.1%, 1% and 5% level, respectively.