RECOVERING WITHIN-COUNTRY INEQUALITY FROM TRADE DATA

Dorothee Hillrichs
ifo Institute, LMU Munich & CESifo

Gonzague Vannoorenberghe
LIDAM/IRES, Université catholique de Louvain

Abstract
This paper develops a novel method to estimate inequality within a country based on what the country imports. If preferences are non-homothetic, rich and poor individuals of a country have different consumption profiles. Imports can thus inform about the income distribution in a country. The global availability of trade data allows us to estimate inequality using a single transparent and comparable method for a large sample of countries over time. Compared to conventional data, we feature an especially good coverage of developing countries. We provide a number of robustness checks and cross-validation exercises that show the good performance of our method. (JEL: F14, D63)

Keywords: International trade, inequality, non-homothetic preferences.

1. Introduction

Measuring income inequality within countries is a notoriously challenging task. Inequality data are typically based on surveys, fiscal data or both. Surveys are, however, costly to conduct and fiscal data is available only where income is taxed. As a consequence, missing values are pervasive particularly among developing countries, where budget constraints and poor institutions make such data collection efforts daunting[1]. Moreover, survey questionnaires and fiscal rules are often country-specific which makes cross-country comparisons difficult even when data is available.

In this paper, we present a novel method for estimating inequality within a country that circumvents these concerns by using import data. Import data have a number of advantages

Acknowledgments: We would like to thank Manuel Oechslin for stimulating discussions that sparked our interest in this project. We thank Moritz Bode, Philip Bodenschatz and Jonas Böschemeier for excellent research assistance. Our thanks also go to Lucas Avezum, Kristian Behrens, Antoine Berthou, Micael Castanheira, Malik Curuk, Léo Czajka, Katharina Erhardt, Mery Ferrando, Harry Huizinga, Udo Kreickemeier, Julien Martin, Florian Mayneris, Yasusada Murata, Gianmarco Ottaviano, Guzman Ourens, Mathieu Parenti, Louis Raes, Sjak Smulders, M. Scott Taylor, Burak Uras, Vincent Vicard as well as the editor, anonymous referees and many seminar participants for their useful comments. This work was supported by the Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) and the Fonds de la Recherche Scientifique – FNRS under EOS Project No. 30784531 "Winners and losers from globalization and market integration" and under the CDR Grant J.0138.18 "Measuring inequality from trade data".

E-mail: hillrichs@ifo.de (Hillrichs); gonzague.vannoorenberghe@uclouvain.be (Vannoorenberghe)

1. For example, for the period 2016-2018, the United Nations' World Income Inequality Database reports income inequality data for only 61 countries. Figure B.2 in the appendix breaks down the data coverage by region and income group of countries.

over survey or fiscal data. They are easier to record than domestic activity, particularly for poor countries (Besley and Persson (2014)). They are double-recorded by the customs authorities of the exporter and of the importer, and thus less sensitive to mismeasurement. They are widely and publicly available, and are recorded by all countries in a similar format, based on the "Harmonized System" of goods classification. We leverage these advantages to generate a comparable measure of inequality for virtually all countries.

Our method builds on the non-homotheticity of preferences, i.e. the fact that an individual's income matters for her consumption basket. For example, consumers may switch from Greek, to Italian, and then to French wine when growing richer. Consider two countries with the same average income. The first country imports wine mostly from Italy while the second imports from France and Greece. Controlling for other confounding factors (e.g. the distance from those countries, etc.), we take this as evidence of a higher inequality in the second country than in the first. Applying this logic over many imported products allows us to build a country-specific measure of inequality based on imports.

To guide our empirical analysis, we embed non-homothetic preferences in an otherwise standard model of international trade à la Armington, in which goods are differentiated by country of origin (e.g. French and Italian wines are two "varieties" of wine). This setup generates a non-homothetic gravity equation: alongside standard gravity determinants such as bilateral trade costs and multilateral resistance terms, the imports of a variety by a country depend on its average income and on the Gini coefficient of its income distribution.[2]

Based on this structure, our empirical approach follows two steps. In the first step, we use a subsample of destinations with high-quality data on inequality and estimate for each variety a parameter capturing the extent to which it is imported by richer or more unequal countries. This parameter captures the relative income elasticity of a variety compared to the other varieties of a product. In terms of our example, this first stage allows us to rank the varieties of wines to which consumers turn as they grow richer. In the second step, we reverse the logic and estimate, for all countries, the Gini coefficient that best fits the observed import patterns given the variety-specific income elasticities estimated in the first step, along with other trade determinants like distance to the exporter.

Applying our method to 8 3-year periods from 1995 to 2018, we generate a Gini coefficient for income inequality for more than 1,000 country-period observations. These include many countries in Sub-Sahara Africa, the MENA region and South Asia, for which survey data are often missing or of poor quality. We provide bootstrapped standard errors to assess the precision of our Gini estimates. Unsurprisingly given our method, countries that import only few varieties or products have much wider confidence intervals and we exclude them from our dataset.

We conduct two main validity checks to assess the performance of our model: a 10-fold cross validation exercise and a comparison with other existing data. Both point to a good fit of our Gini estimates, which fall well within the range of estimates from different sources. Two of the datasets to which we compare our results have a similar coverage to ours: the

2. Recent research in international trade shows that the within-country income distribution plays an important role in explaining trade flows (e.g. Fieler (2011)).

World Inequality Database (WID) and Solt (2009)'s Standardized World Income Inequality Database (SWIID). We show that our Gini estimates for poor countries are on average very close to the WID data. As we describe in detail in section 2.1, both WID and SWIID rely, as we do, on a set of country-specific data and on a set of assumptions that map these data into measures of inequality. We make progress in filling the gaps using high-quality standardized country-specific information (customs data) in combination with theory-based, transparent assumptions on the mapping to inequality.

Where comparison is possible, we tend to find higher levels of income inequality than what is captured by survey data. For some countries, we diverge more strongly from existing sources. For example, we find only modest reductions in inequality in many Latin American countries since the early 2000s compared to survey data, which typically point to very strong reductions in inequality. For the US, we find a relatively stable inequality over the period, a few points higher than most survey-based data (depending on the comparison data source), but lower than the levels reported in the World Inequality Database.

The validity of our method relies on two key assumptions beyond the non-homotheticity of preferences. First, it requires some degree of comparability of preferences across countries. The ranking of income elasticities of varieties within products needs to extend from countries in the first step-sample, on which we estimate the preference parameters, to countries for which we infer inequality in the second step. We do not, however, require that preferences for products are the same across countries: some countries may have a strong preference for wine while others do not. Second, since we rely on imports of consumer goods for identification, our logic requires that imports of goods provide enough information to reliably infer income inequality. If, for example, subsistence farmers consume few imported products or if rich individuals fly to Paris to drink their wine, our inequality measure will miss some relevant variation.

We discuss these issues at length and provide refinements to our method. We also examine the robustness of our method to a number of model assumptions and find a high correlation of the inequality estimates across specifications. As we show in this paper, relying on import data provides valuable information to fill in the gaps in inequality measurement. Our method complements the many forensic country-specific analyses by extending available data on inequality in a simple and transparent way when survey and fiscal data are scarce. We encourage researchers to use our estimates in combination with other sources of information on inequality or other imputation methods, in the spirit of Henderson et al. (2012) in the context of economic growth. Our discussion on the precision of our estimates and on the factors that could influence our prediction error in sections 6 and 7 should help spot countries on which our estimates can carry more weight.

The remainder of the paper is structured as follows. Section 2 relates our method to other approaches taken in the literature on recovering inequality information, and gives an overview of the international trade literature our method is based on. Section 3 then outlines the theoretical structure we use to develop our method. Section 4 details the estimation strategy and discusses caveats to the method. Section 5 describes the trade and inequality survey data. Section 6 presents the estimated inequality data, section 7 validates our estimates with out-of-sample analyses and section 8 discusses the robustness of our method. Section 9 concludes.

2. Related literature

We contribute to the research on measuring inequality and add to the literature on non-homothetic preferences in international trade.

2.1. Literature on inequality data

The sparsity of inequality data based on surveys has gained the attention of researchers from various fields. McGregor et al. (2019) provide a comprehensive overview of challenges and solutions to measuring inequality. Ferreira et al. (2015) introduces a special issue of the *Journal of Economic Inequality* (Volume 13, Issue 4) devoted to assessing and comparing the most widely used inequality data sets available at the time. Different studies share the aim of extending the data coverage on inequality while ensuring consistency. Galbraith and Kum (2005) predict household income inequality from a linear model of the Deininger and Squire (1996) Gini coefficient and the industrial payment inequality, controlling for the manufacturing share in employment as well as the Gini underlying welfare definition. They however have difficulties in updating their data due to changes in the primary data used (Galbraith et al., 2014). Solt (2009)'s SWIID database collects all possible sources on inequality for all countries, regardless of the method, and uses an algorithm to harmonize these diverse sources into a comparable measure of net income inequality comparable to the Luxembourg Income Studies (LIS). Another source of information on inequality is the World Inequality Database (WID) led by Facundo Alvaredo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. Their initial focus lies on the upper tail of the income distribution, but the WID data has recently been updated to reach a large coverage of Gini coefficients. WID combines information from survey and fiscal data to derive a detailed account of a country's income distribution. While their approach may lead to more accurate and nuanced inequality measures, data requirements are high and hard to meet in developing countries.

SWIID, WID and our our method use information about the country: all possible Gini data in SWIID, survey and fiscal data in WID, and customs data in our case. All three methods also use a set of assumptions to map these different data into a consistent and comparable measure. SWIID assumes that the rule to convert an inequality measure to a standard measure is the same across countries of a large world region. When applying the Distributional National Accounts framework, WID needs assumptions to fill in patchy and inconsistent survey and fiscal data, as well as on how different transfers and taxes should be distributed across individuals. The advantage of our method is that the country-specific information, the customs data, is more standardized and reliable compared to most microdata on which other methods rely. This comes at the cost of relying on imports as a proxy for inequality rather than on actual measures of inequality. We therefore see our approach, which is applicable to nearly all countries uniformly under comparatively low data requirements, as complementing their work. As we show in section 6, our estimates are close to WID for many countries.

Our approach is similar in spirit to recent advances in the field that use secondary data to uncover changes in the income distribution. Blumenstock et al. (2015), for example, map usage of mobile phones to the distribution of wealth in Rwanda. Alesina et al. (2016) and

Lessmann and Seidel (2017) use satellite nighttime light imagery to generate within-country regional inequality data.

Our method is closely connected to Aguiar and Bils (2015). Aguiar and Bils (2015) measure consumption inequality for the US based on the relative allocation of spending of rich and poor households across luxuries and necessities. They employ a similar two-step estimation procedure to us. Their purpose, however, is different from ours. The goal of Aguiar and Bils (2015) is to address a systematic measurement error in the US Consumer Expenditure Survey. Instead, this study aims to generate a cross-country dataset on inequality. The methodology of Almås (2012) is also similar in spirit to our own. Almås (2012) estimates PPP corrected incomes across countries from estimating Engel curves for food. Similarly, our inequality measure exploits the variation in spending allocation across goods at different income levels. Our estimation, though, includes a larger variety of goods and is applied to a larger set of countries. The outcome variable of interest also differs between our study and Almås (2012).

2.2. Non-homothetic preferences in the trade literature

Our method builds on insights from the international trade literature. Inspired by the seminal dissertation of Linder (1961), a growing number of studies draw attention to the role of the income distribution as a determinant of trade patterns. Hallak (2006, 2010) provides evidence that similarity in the per capita income increases trade flows between countries. Choi et al. (2009) map income similarity to import price similarity suggesting a role for a country's income distribution in explaining the quality of imported goods. Caron et al. (2014) show how import demand varies differentially with per capita income across goods. The authors estimate sector-level income elasticities which they use to assess the role of per capita income in explaining several trade "puzzles". Jaimovich and Merella (2015) argue that non-homothetic preferences for quality matters for the strength of specialization in international trade. In this study, we provide further evidence on the significance of per capita income as a determinant of bilateral trade flows. To our knowledge, we are the first to exploit the variation of income elasticities across goods to derive a measure of a country's income distribution.

Another strand of the literature connects non-homothetic preferences to the differential impact of trade on consumers of different income. If consumers buy different baskets of goods, they will benefit differently from the relative price changes brought about by trade (see Porto (2006), Nigai (2016), Fajgelbaum and Khandelwal (2016), or Artuc et al. (2019) among others). The question whether richer households tend to consume more or less imported products is often at the core of the unequal effects of trade on the consumption side. In a recent paper, Borusyak and Jaravel (2021) argue that US households across the income distribution spend similar shares on imported product (both direct and indirect through input-output linkages). Importantly, they show that the origin of imported goods in their sample differs across the income distribution, which is the type of variation that our analysis relies on, albeit with a different product definition.

From a theory perspective, non-homothetic preferences provide the link that connects income and imports. The trade literature suggests various microfoundations generating non-unitary income elasticities across goods. We introduce non-homotheticities similar to Comin

et al. (2021) and Matsuyama (2019) who work with an implicit additively separable CES function. Alternatives such as Faber and Fally (2017) and Handbury (2019) or Feenstra and Romalis (2014) would result in very similar Engel curves. Fieler (2011) introduces non-homothetic preferences into a CRRA structure, but the connection that it generates between the elasticity of substitution and the degree of non-homotheticity seems less natural for non-homotheticities at the level of varieties. Relating to the consumer choice literature, Fajgelbaum et al. (2011) propose a nested logit demand system in which the expenditure share on quality varies with per capita income. Fajgelbaum and Khandelwal (2016) estimate heterogeneous income elasticities from an Almost Ideal Demand System.

Given the established high prediction power in log-linear models for trade flows, and the higher sensitivity to outliers in an estimation with import shares as dependent variable (instead of log imports, see Hillrichs and Vannoorenberghe (2022)), we opt for the class of non-homothetic CES preferences.[3] Compared to AIDS, our approach does not constrain the Engel curves to be linear in log expenditure, and is closer in spirit to Banks et al. (1997) , who allow for more flexibility in the curvature of the Engel curve.

3. Theory

We first present the approach in a general framework. To derive our estimating equation, we then put a parametric structure on the general model. For ease of exposition, we abstract from the time dimension and integrate it explicitly when describing our estimation strategy in section 4.

3.1. The general idea

The economy consists of C countries, indexed by $d \in \{1, .., C\}$ and J goods, indexed by $j \in \{1, .., J\}$. We think of goods as being differentiated by country of origin à la Armington, and assume that a variety is a good produced by a particular origin country, indexed by $o \in \{1, .., C\}$.

Within each country, we classify individuals into income categories, indexed by $h \in 1, ..., H$, such that individuals in h have an income I_h. In country d, there are N_{dh} individuals in income category h, and a total of N_d individuals. We denote the average income of country d as $I_d = \sum_h N_{dh} I_h / N_d$ and define the expenditure that an individual in income category h in destination country d spends on the variety of good j coming from o as $x_{jo}(\mathcal{B}, \mathbf{A}_d, I_h)$, where $\mathcal{B}$ is a matrix of unobserved variety-specific coefficients. $\mathbf{A}_d$ is a matrix of destination-specific variables such as the vector of prices or of some taste parameters that can be observed or proxied. Aggregating over all individuals in country d gives:

$$X_{jod} = \sum_h N_{dh} x_{jo}(\mathcal{B}, \mathbf{A}_d, I_h). \tag{1}$$

3. The structural change literature typically invoke Stone-Geary preferences (Kongsamut et al., 2001). However, under Stone-Geary preferences the income effect vanishes at high levels of income, which is undesirable for our purposes.

Using a second order Taylor approximation of x_{jo} around the average income I_d and simplifying, the imports of variety jo in country d become (see appendix A.1 for derivation):

$$Log(X_{jod}) \approx Log\left(N_d x_{jo}(\mathcal{B}, \mathbf{A}_d, I_d)\right) + \frac{1}{2}\frac{\partial^2 x_{jo}(\mathcal{B}, \mathbf{A}_d, I_h)}{\partial^2 I_h}\bigg|_{I_h = I_d} \frac{I_d^2}{x_{jo}(\mathcal{B}, \mathbf{A}_d, I_d)} G_d, \tag{2}$$

where G_d is the squared coefficient of variation of income in country d, $G_d \equiv \left(\sum_h \frac{N_{dh}}{N_d}(I_h - I_d)^2\right)/I_{dt}^2$. The first term captures the consumption of jo that would prevail in d if all individuals had the average income I_d. The second term captures the effect of income inequality on the consumption of jo and is non-zero as long as the the Engel curves are not linear.

We assume that data on the distribution of income are observable with great accuracy for some countries ($d \in \mathbb{O}$) while they are not for others ($d \in \mathbb{U}$). Our aim is to recover the distribution of income (G_d) for countries where it is unobserved, i.e. for $d \in \mathbb{U}$. Thanks to the wide availability of high-quality international trade data, we can observe the left hand side of equation (2) for virtually all imported varieties and destinations d. We thus propose a mapping from trade flows to a measure of inequality for countries $d \in \mathbb{U}$.

Our method has two steps. The first is to recover $\mathcal{B}$ by estimating (2) for $d \in \mathbb{O}$. In the second step, we use the estimated matrix $\hat{\mathcal{B}}$ to compute the predicted consumption of each variety if all households had the country's average income. We then obtain an estimate of G_d for $d \in \mathbb{U}$ by regressing the difference between actual imports and predicted imports without inequality on the degree of convexity of the Engel curve.[4]

To illustrate the role of the convexity of the Engel curve, consider reallocating one dollar from the poor to the rich, an increase in inequality. The poor decrease their spending on the variety while the rich increase theirs (we think of a "normal" variety to simplify exposition). If the Engel curve is convex, the decrease in consumption by the poor is small compared to the increase in consumption by the rich and the total consumption of the variety rises. Keeping average income constant, relatively large imports of a good with a convex Engel curve are a sign of high inequality. The opposite holds if the Engel curve is concave.

The validity of our method relies on a few key assumptions. First, it requires *non-homotheticity of preferences* for at least some varieties such that Engel curves are non-linear. Second, it relies on some degree of *comparability of preferences* across varieties within products ($\mathcal{B}$ is not specific to d whereas $\mathbf{A_d}$ can contain preference parameters). Section 3.2 makes our assumptions on preferences explicit. Third, since we rely on imports of consumer goods for identification, this requires that *imports of goods provide sufficient information* about consumption patterns to extract information about the full income distribution. We discuss these issues and provide a number of extensions to the baseline model to gauge the impact of these assumptions in section 8.

4. Our strategy assumes that average income per capita is measured correctly for all countries. While measurement error in income per capita data certainly exists, the lack of inequality measures seems much more of a concern. We could in principle extend our method to recover both income per capita and income inequality for countries with low-quality data but leave it for future work.

3.2. Baseline model

We now make a number of functional form assumptions to bring equation (2) to the data as a non-homothetic gravity equation.

Consumers in country d derive utility from a bundle of differentiated goods, indexed by $j \in \{1, ..., J\}$. Utility is defined implicitly by a CES aggregator over the consumption of all goods:

$$U = \left(\sum_j \mu_{jd}(U) C_j^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}$$

where $\mu_{jd}(U)$ is a demand shifter that depends on the utility level of the consumer, such that $\sum_j \mu_{jd}(U) = 1$. $\mu_{jd}(U)$ governs non-homotheticities across goods: the relative taste for good j increases with utility if $\mu'_{jd}(U) > 0$, in line with Comin et al. (2021).[5]

Each of the differentiated goods consists of varieties, differentiated by country of origin in an Armington fashion[6], such that:

$$C_j = \left[\sum_o \varphi_{jod}(U) c_{jod}^{\frac{\gamma_j}{\gamma_j - 1}} \right]^{\frac{\gamma_j - 1}{\gamma_j}}$$

where $\gamma_j < 0$ is equal to one minus the good-specific substitution elasticity between varieties and:

$$\varphi_{jod}(U) = \frac{\alpha_{jod}^{\frac{1}{1-\gamma_j}} U^{\frac{\beta_{jo}}{1-\gamma_j}}}{\sum_{o'} \alpha_{jo'd}^{\frac{1}{1-\gamma_j}} U^{\frac{\beta_{jo'}}{1-\gamma_j}}}.$$

In a similar manner as across goods, we introduce non-homotheticities in preferences by making the demand shifter across varieties depend on the utility level of a consumer.[7] Consumers with a higher utility have stronger preferences for varieties with a high β_{jo}. α_{jod} is a taste parameter shared by all individuals in country d. We define a consumer's share of spending on good j as S_{jd} and the share of spending on j that goes to variety jo as s_{jod}. As we show in the appendix, minimizing expenditure for a given level of utility gives:

$$s_{jod}(U) = \frac{\alpha_{jod} U^{\beta_{jo}} p_{jod}^{\gamma_j}}{\sum_{o'} \alpha_{jo'd} U^{\beta_{jo'}} p_{jo'd}^{\gamma_j}} \quad \text{and} \quad S_{jd}(U) = \frac{\mu_{jd}(U)^\rho P_{jd}(U)^{1-\rho}}{\sum \mu_{j'd}(U)^\rho P_{j'd}(U)^{1-\rho}}$$

with

$$P_{jd}(U) = \left(\sum_{o'} \varphi_{jo'd}(U)^{1-\gamma_j} p_{jo'd}^{\gamma_j} \right)^{\frac{1}{\gamma_j}}.$$

5. Our utility is equivalent to that in Comin et al. (2021) if our $\mu_{jd}(U)/U$ equals $g(U)^{\varepsilon_j \frac{\rho-1}{\rho}}$ in Comin et al. (2021) .

6. We discuss implications of the Armington assumption and provide an extension to the baseline model that allows for differentiation within country-product pairs in Section 8.

7. Alternative specifications of non-homotheticities could make the taste preferences a function of the consumption of an outside good, as in Faber and Fally (2017) or Handbury (2019), without affecting the results.

Since our underlying demand shifters depend on utility, recovering inequality from expenditure patterns means that we would recover an inequality of utility. We show in the appendix that, if a consumer's expenditure is equal to its income, utility is proportional to income under mild conditions and we can replace utility in the above equations by income. We explore alternative interpretations in section 8. Spending on variety jo by a consumer with income I in country d becomes:

$$x_{jod}(I) = s_{jod}(I)S_{jd}(I)I = \frac{\alpha_{jod} I^{\beta_{jo}} p_{jod}^{\gamma_j}}{\sum_{o'} \alpha_{jo'd} I^{\beta_{jo'}} p_{jo'd}^{\gamma_j}} S_{jd}(I)I. \tag{3}$$

The income elasticity of demand for variety jo is given by:

$$\frac{\partial x_{jod}(I)}{\partial I}\frac{I}{x_{jod}} = 1 + \beta_{jo} - \bar{\beta}_{jd}(I) + \varepsilon_{jd}^S(I) \tag{4}$$

where:

$$\bar{\beta}_{jd}(I) = \sum_{o'} s_{jo'd}(I)\beta_{jo'} \quad \text{and} \quad \varepsilon_{jd}^S(I) = \frac{\partial S_{jd}(I)}{\partial I}\frac{I}{S_{jd}(I)}.$$

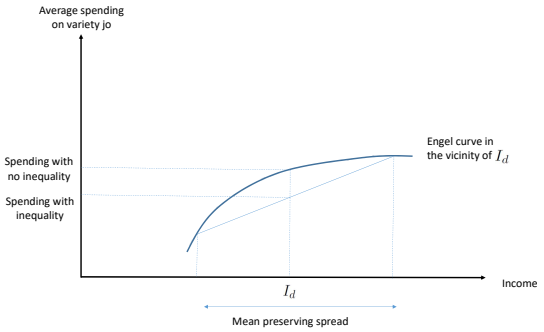
The term $\beta_{jo} - \bar{\beta}_{jd}(I)$ shows whether variety jo has a higher parameter β than the other varieties of j consumed by an individual with income I. $\varepsilon_{jd}^S(I)$, the product share income elasticity, indicates whether richer individuals spend relatively more on good j. Combining these effects determines the income elasticity of variety jo. It can easily be seen that $\partial\bar{\beta}_{jd}/\partial I > 0$, i.e. richer individuals consume varieties with a higher average β. Some varieties jo may be inferior for some relatively rich individuals ($\beta_{jo} - \bar{\beta}_{jd}(I) + \varepsilon_{jd}^S(I) < -1$), but can be normal ($-1 < \beta_{jo} - \bar{\beta}_{jd}(I) + \varepsilon_{jd}^S(I) \leq 0$) or even a luxury ($\beta_{jo} - \bar{\beta}_{jd}(I) + \varepsilon_{jd}^S(I) > 0$) from the perspective of poorer individuals. The second derivative is:

$$\begin{aligned}\frac{\partial x_{jod}(I)}{\partial^2 I}\frac{I^2}{x_{jod}(I)}\bigg|_{I=I_d} = & \left(1 + \beta_{jo} - \bar{\beta}_{jd}(I_d) + \varepsilon_{jd}^S(I_d)\right)\left(\beta_{jo} - \bar{\beta}_{jd}(I_d) + \varepsilon_{jd}^S(I_d)\right) \\ & - \sum_{o'} s_{jo'}\left(\beta_{jo'} - \bar{\beta}_{jd}(I_d)\right)^2 + \frac{\partial\varepsilon_{jd}^S(I_d)}{\partial I_d} I_d \end{aligned} \tag{5}$$

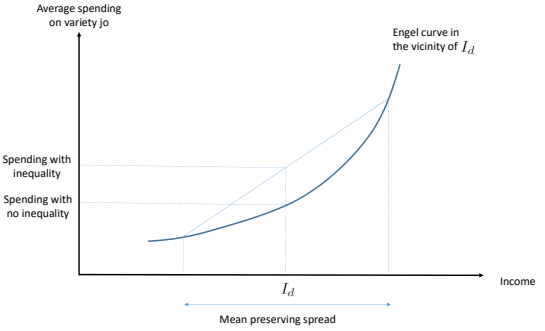
β_{jo} is closely related to the curvature of the Engel curve but in a non-monotonic manner. Consider first the case where $S_{jd}(I_d)$, i.e. the share of spending on each good, is independent of income. For intermediate β_{jo} (including $\beta_{jo} = \bar{\beta}_{jd}$), the Engel curve is concave, implying that a mean-preserving rise in income inequality reduces the consumption of the variety with the average β within a good. For varieties with a sufficiently high β_{jo} (luxuries) or a sufficiently low β_{jo} (inferior goods), the Engel curve is convex and a mean-preserving rise in inequality raises the consumption of these varieties. We illustrate these different cases graphically in Figure 1.[8]

8. For high β_{jo} varieties, the Engel curve is upward sloping and convex: a mean-preserving spread in inequality means a slightly lower consumption by the poor and a large increase in consumption by the rich, raising the total consumption of the variety. From equation (5), a necessary condition for the Engel curve to be convex for low β_{jo} is that the good be inferior, i.e. $\beta_{jo} - \bar{\beta}_{jd} + \varepsilon_{jd}^S < -1$. A mean-preserving increase in inequality implies in this case a large increase in consumption by the poor and a moderate decrease in consumption by the rich.

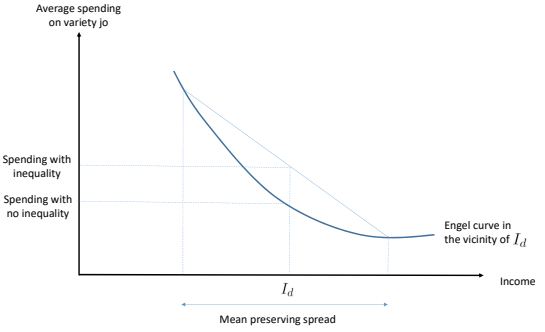
FIGURE 1. Theory mechanisms: spending and curvature of Engel curves



(A) Intermediate β_{jo}



(B) High β_{jo}



(C) Low β_{jo}

Notes: Engel curves for different levels of β_{jo}, assuming a constant S_{jd} over the income range.

Allowing for non-homotheticities across goods brings in two additional effects on the curvature. First, the curvature at the good's level (captured by $\frac{\partial \varepsilon^S_{jd}(I_d)}{\partial I_d} I_d$) is added to the curvature of a variety's Engel curve. This effect is the same for all varieties of a good. Second, it changes the thresholds for being an intermediate, relatively high or low β variety in the discussion above. A variety with an intermediate β_{jo} now becomes a variety with a β_{jo} close to $\bar{\beta}_{jd} - \varepsilon^S_{jd}$.

Each variety jo is produced under perfect competition, with constant marginal costs w_{jo}. Exporters of good j from country o have to ship $\tau_{jod} \geq 1$ units for one unit to arrive at destination d. These assumptions ensure that there is no pricing to market across countries ($p_{jod} = \tau_{jod} w_{jo}$), an assumption we come back to in section 8.

In sum, the expression for spending on variety jo by country d if all consumers had the average income - the first part of (2) - resembles a standard gravity equation, except for a coefficient on income per capita that is variety-specific:

$$Log(N_d x_{jod}(I_d)) = F_{jo} + F^1_{jd} + \gamma_j Log(\tau_{od}) + \beta_{jo} Log(I_d) + Log(\alpha_{jod}) \tag{6}$$

where

$$\begin{aligned} F_{jo} \equiv & \gamma_j Log(w_{jo}) \\ F^1_{jd} \equiv & Log(S_{jd}(I_d)) - Log\left(\sum_{o'} \alpha_{jo'd} I_d^{\beta_{jo'}} p_{jo'd}^{\gamma_j}\right) + Log(I_d). \end{aligned}$$

Replacing equation (5) in equation (2), the gravity equation can thus be rewritten as:

$$\begin{aligned} Log(X_{jod}) = & F_{jo} + F_{jd} + \gamma_j Log(\tau_{od}) \\ & + \beta_{jo} Log(I_d) + \frac{1}{2}\beta_{jo}\left(1 + \beta_{jo} - 2\left(\bar{\beta}_{jd}(I_d) - \varepsilon^S_{jd}(I_d)\right)\right) G_d \\ & + Log(\alpha_{jod}) \end{aligned} \tag{7}$$

where:

$$F_{jd} \equiv F^1_{jd} + \left[\left(\bar{\beta}_{jd} - \varepsilon^S_{jd}\right)^2 - \left(\bar{\beta}_{jd}(I_d) - \varepsilon^S_{jd}(I_d)\right) - \sum_{o'} s_{jo'd}(I_d)(\beta_{jo} - \bar{\beta}_{jd}(I_d))^2\right] G_d. \tag{8}$$

4. Two-step estimation strategy

We recover within-country inequality from import patterns in two steps. Step 1 estimates the non-homothetic gravity equation derived in section 3.2 on the subsample of countries for which we observe high-quality inequality data. Using the parameter estimates, we then predict counterfactual imports for all countries, whether or not they are part of the destination sample in step 1, imposing perfect income equality. We calculate the difference between observed and such predicted imports and regress this difference on the inequality coefficients estimated in step 1. A country's level of inequality is pinned down by the within-product variation in imports across varieties with different sensitivities to inequality.

4.1. First stage

To move from theory to empirics, we need to make two adjustments to equation (7). First, a difficulty in estimating (7) is that the coefficient on inequality, $b_{jod} \equiv \frac{1}{2}\beta_{jo}\left(1+\beta_{jo}-2\left(\bar{\beta}_{jd}(I_d)-\varepsilon^S_{jd}(I_d)\right)\right)$, is destination-specific as it contains the expenditure share-weighted average income elasticity of imports in destination d ($\bar{\beta}_{jd}$) and the product-level income elasticity ($\varepsilon^S_{jd}(I_d)$). To arrive at an estimable version of equation (7), we group countries by world region (Americas, Europe or other) and income group (high income country or other). According to equation (4), countries with similar import patterns should have a similar value of $\bar{\beta}_{jd}$. Both the country's location and its income affect imports according to our theory so we can reasonably assume $\bar{\beta}_{jd} = \bar{\beta}_{jg(d)}$ for all destinations in group g. We expect countries in the same group to have similar levels of ε_{jd} so we assume $\varepsilon^S_{jd}(I_d) \approx \varepsilon^S_{jd'}(I_{d'})$ for $d, d' \in g$. We return to the definition of groups in chapter 8. The grouping strategy imposes that imports of a variety jo respond in the same way to a change in inequality in Guatemala and El Salvador (both in group Americas, low income) but may well respond differently to a change in US inequality (Americas, high income).

Second, we need to impose a structure on the unobserved bilateral trade costs τ_{od}. We make the standard assumption that bilateral trade costs τ_{od} are a linear combination of log geographical distance, dummies for a shared border, a common official language, a common colonial past and for participation in the same regional trade agreement, $\ln(\tau_{od}) = \sum_k \kappa_k k_{od}$ with κ_k the trade cost elasticity to variable k.

The estimable version of (7) is the twoway-fixed effects model

$$\begin{aligned}\ln(X_{jodt}) = {} & F_{jot} + F_{jdt} + \sum_k \delta_{jk} k_{od} \\ & + \sum_{o'} \beta_{jo'} \mathbb{1}_{o'=o} \ln(I_{dt}) + \sum_{o'}\sum_{g'} b_{jo'g'} \mathbb{1}_{o'=o} \mathbb{1}_{d\in g'} G_{dt} \\ & + \varepsilon_{jodt}, \end{aligned} \tag{9}$$

which we estimate by product on the subsample of country-periods for which we have reliable inequality data ($dt \in \mathbb{O}$). To stress the time dimension of our data, we here introduce the time subscript explicitly. F_{jot} and F_{jdt} denote exporter-product-time and destination-product-time fixed effects. In line with our theory, the exporter fixed effects absorb production costs on the exporter side. Importantly, the destination fixed effect absorbs average consumption of product j, average consumption of the domestic variety as well as non-homotheticities at the product level. The coefficients on the trade cost proxies are the product of the trade cost elasticty γ_j and the elasticity of trade costs to the respective proxy variable κ_k. The variety-specific income elasticities are captured by the coefficients on the interaction between destination logarithmic GNI per capita and an indicator $\mathbb{1}_{o'=o}$ that takes value one for exports from origin o and zero for all other varieties. The identification relies on the variation in income per capita across destination-periods of a given exporter-product pair. Income elasticites are identified relative to a reference exporter for each product. For brevity, we refer to β_{jo} as income elasticity from now on. The inequality elasticities b_{jog} are estimated on the interaction of a group indicator, an exporter (variety) indicator and the

destination coefficient of variation squared. The indicator function $\mathbb{1}_{d\in g}$ takes value one for all countries importing variety jo that are in group g. Hence inequality elasticities are identified relative to a reference exporter for each group and product. The variety-specific inequality coefficients are identified from the cross-sectional and time-series variation in inequality across an exporter's destinations in the same destination group. Since we control for the average income of the importers, variation in inequality is mean-preserving, in line with theory. Finally, the unobservable taste shocks $\ln(\alpha_{jodt})$ enter the error term ε_{odjt}. We allow for heteroscedastic shocks to be correlated over time by clustering standard errors at the (product-)exporter-destination level.

4.2. Second stage

Armed with the parameter estimates of step 1, we exploit the variation of inequality elasticities across varieties within products to infer within-country inequality from import patterns. Define

$$\widetilde{\ln(X_{jodt})} \equiv \ln(X_{jodt}) - \left(\hat{F}_{jot} + \sum_k \hat{\delta}_{jk} k_{od} + \hat{\beta}_{jo} \ln(I_{dt})\right), \tag{10}$$

where $\hat{\delta}_{jk}, \hat{\beta}_{jo}$ and $\hat{F}_{jot}$ are the parameter estimates from the first step. The term in brackets equals the predicted variety imports if inequality in dt was zero.[9] We thus can compute this term for all countries, whether or not we observe inequality data in reality. To recover within-country inequality, we project the deviations of predicted imports from observed imports on the estimated inequality coefficients for varieties imported by dt. We also include product-level fixed effects (by destination-time). These have the same role as in step 1, absorbing the expenditure share at the product level and average expenditure on the domestic variety. The second step regression equation is given by

$$\widetilde{\ln(X_{oj}^{dt})} = f_j^{dt} + \sum_{(dt)'} \mathcal{G}^{(dt)'} \mathbb{1}_{(dt)'=dt} \hat{b}_{jo}^{g(d)} + \nu_{jo}^{dt}, \tag{11}$$

which we estimate with a one-way fixed effects model on the full sample $dt \in \mathbb{O} \cup \mathbb{U}$. We use the superscript notation to highlight that the estimation uses only within destination-period variation. The coefficient $\mathcal{G}^{dt}$ corresponds to the squared coefficient of variation for each country and time period. We estimate it from the interaction of an indicator for destination-time dt and the inequality coefficient estimates of varieties imported by d in t. It is identified from the variation in inequality elasticities across varieties within products. Recall that the inequality elasticities govern the curvature of the Engel curve. As the estimates are variety-destination group specific, the mapping takes into account that curvatures vary across destinations.

9. We use this formualation for ease of exposition. Predicted imports also include the destination-time-product fixed effects. The second step includes country-time pairs that were not part of the first stage. We therefore do not have a first stage estimate for these fixed effects and re-estimate them in step 2.

Moving from step 1 to step 2, we switch what is a variable and what is a parameter in equation (9). As we are relying on estimated regressors in step 2, which are subject to uncertainty, we estimate standard errors for $\hat{\mathcal{G}}^{dt}$ using wild bootstrap (Horowitz, 2019), a parametric bootstrap particularly suited for heteroscedastic data. We create 100 bootstrap samples for our first step where the dependent variable equals the predicted trade flows according to equation (9) plus the predicted error term scaled by a random draw from the Mammen distribution. We estimate step 1 on each bootstrap sample and obtain 100 parameter estimates that we combine with the observed data to create the regression sample for step 2. We then estimate step 2 for each bootstrap sample and compute standard errors and confidence bounds with the percentile method.

4.3. Identification

As explained in section 3.1, our method relies to some extent on comparability of preferences across countries. Our key assumption regarding the comparability of preferences is that the relationship between trade and per capita income is linear, i.e. that β_{jo} is common to all destination countries. Unreported test results show that we cannot reject the linear relationship between trade flows and per capita income for more than 90 % of varieties.[10] However, our method does not require that all preference parameters are equal in all countries. For example, since b_{jog} is common to all destinations *in a group*, the linearity assumption about inequality applies only locally. Moreover, product-level expenditure shares may differ across destinations and are captured in our estimation by destination-good-time fixed effects. Our identification strategy aims precisely to limit the need for preference comparability by using *within-product* variation between varieties.

Moreover, we allow for countries to differ in terms of a destination-variety specific preference shifter, our structural error term. In the first stage, consistent estimates of β_{jo} and b_{jog} for a variety requires that the error term be uncorrelated with the income and inequality of destination countries. We think that this assumption will likely hold for many varieties, but it might sometimes be violated[11]. For example, if Swiss watches are particularly popular in countries that host a Grand Slam tennis tournament and these countries are richer than average, this may drive an upward bias in the β of Swiss watches. A biased first stage estimate would however only pose a threat to our strategy if it carries over to the second stage.

In the second stage, the existence of an idiosyncratic preference shifter biases the estimate of the Gini coefficient if it is correlated with the estimated income elasticity across varieties. In other words, it would require that a country disproportionately likes varieties typically preferred by the world's rich (or poor) for reasons that have nothing to do with income or gravity variables. If the estimated β_{jo} in the first stage is biased for some varieties, another

10. We replicated our first stage allowing for an additional interaction between a variety dummy and the squared GNI per capita. The coefficient on the squared term is insignificant in 90% of cases.

11. In section 8, we provide a robustness exercise where we allow preferences to vary by large world regions. We reestimate our model with exporter-product-region of the destination fixed effects to allow for region-variety specific demand shifters (e.g. Asian consumers may have a different demand shifter for French wine than Northern Europeans). We allow for region-variety specific income elasticities.

issue would be a correlation between those first-stage biases and the country's preference shifters across all varieties. We see such correlations as quite unlikely over the hundreds of products that we consider.

4.4. Refinement

In addition to the wild bootstrap described above, we explore the sensitivity of our estimates to the products a country is importing with a block bootstrap. In this exercise, we draw random products with replacement from a country's import basket and re-estimate step 2 100 times. This gives an alternative estimate for the Gini confidence interval and reveals whether Gini estimates for some countries are driven by high-leverage goods.

While too high sensitivity of the Gini estimates to single products would cast doubt on our method, some products might, however, be more informative about inequality than others. Putting a higher weight on these goods would improve the predictive accuracy of our method. A refinement of our method finds those goods that are most informative about within-country inequality. Inspired by the so-called forward and backward stepwise selection methods, two machine learning methods for sequentially selecting variables from a large set of possible predictors in a model (see Chapter 6 in James et al. (2021)), we select the set of products most informative about within-country inequality.

In our context, forward selection starts with a single product and sequentially adds goods to the estimation sample. Backward selection starts from the full sample and sequentially drops products. The order in which products are added or dropped is such that at each stage the reduction in the root mean squared error, defined here as $\sum_{dt\in\mathbb{O}}\left(\left(\hat{\mathcal{G}}_{dt}-G_{dt}\right)^2\right)^{1/2}$, is maximal or the increase minimal.

Once the order of goods is determined, the ideal "import basket size", i.e. the number of goods to be included, is determined. For the latter, we use the 10-fold cross-validation using parameter estimates where the country for which we are predicting inequality was not part of the first step-sample. The former is based on in-sample estimates. Using out-of-sample estimates to find the ideal import basket size prevents overfitting, i.e. choosing a combination of goods that happens to fit the observed Gini G_{dt} well. When a country does not import a good that is selected in an iteration step, it enters the RMSE calculation with its inequality estimate from the previous iteration. The backward selection procedure works analoguous. Starting from the full sample of goods, products are subsequently dropped such that the decrease in RMSE is largest from one iteration to the next.

5. Data

5.1. Trade data

We choose the BACI trade database compiled by CEPII, which is based on the UN Comtrade database, to obtain bilateral trade data at the product level. The BACI trade database reconciles trade flow reports by the importer and the exporter country (for details of the

methodology see Gaulier and Zignago (2010)). This is important for our purposes. Exploiting the double-recording nature of trade flows lessens the dependency on the statistical capacity of a single country regarding the data quality. Products are classified by the Harmonized System (HS). We aggregate trade flows to the HS 4-digit product level to limit the incidence of zero trade flows and maintain a sufficient number of varieties by product. We retain only consumption goods, as classified by UNCTAD, since expenditure on these goods should be relatively more affected by per capita income and inequality than raw materials or intermediate goods.

We observe trade flows for the period 1995 to 2018. To smooth out annual shocks, we aggregate the trade flows to three-year periods and take the three-year average value for yearly varying control variables. To increase the reliability of our data further, we exclude trade flows to or from countries with fewer than 500 000 inhabitants. We keep only the largest exporters, the top half of the export value distribution within a product over the entire sample period. While this step cuts the number of observations from more than 11.4 Million to about 9 Million, the observed total export value falls by less than 10% for all product-period pairs due to the skewness in trade values. In the first stage estimation, we restrict the sample to exporters with more than 15 destinations per good and period as well as 15 destination-periods in a given group g per product. The final data set has 8,8 million observations, while the first stage data has about 3.8 million observations.

Table 1 presents a summary of our trade data. Our final sample consists of 360 HS 4-digit codes, 123 exporting countries, and 152 importing countries. In the first stage, variety-specific income elasticities are identified from the variation of per capita income and net income inequality across on average 264 country-periods of our first step sample. The number of country-periods per variety in the first step sample varies between 28 (Poland, Mate) and 441 (India, Medicaments for retail sale). The identification of within-country inequality in the second stage makes use of variation across on average more than 6300 varieties imported per period. The number of varieties imported per product and period ranges from 5 to 77.

TABLE 1. Summary statistics of trade data

	Mean	Sd	Median	Min	Max
Importer-time per variety (1st sample)	265	99	260	28	441
Varieties per importer & time & HS4	23	15	19	5	77
Varieties per importer & time	6,340	3,693	5,927	27	14,629

Notes: The table presents summary statistics on the number of observations along key dimensions in the trade data. In total, our sample contains 15667 exporter-hs4 combinations () in the first stage sample and 1182 importer-time pairs.

5.2. Within-country inequality data

The second key data used in our estimation is data on inequality. The measure of inequality that enters our estimating equation is the squared coefficient of variation of a country's income distribution (G), which is not usually available in cross-country databases of

inequality. To connect theory to data, we follow Kleiber (2008) and Bandourian et al. (2002) and assume a Dagum distribution (Dagum, 2008) for the mapping of the Gini to the coefficient of variation, see appendix B.2 for details.

We obtain the within-country Gini data from the World Income Inequality Database (WIID Version 2020) administered by UNU-WIDER. The WIID is itself a collection of Gini data sets coming from various institutional surveys and research studies. Each data point is described by the underlying welfare concept (net/gross income, consumption), the reference population (urban/rural/all) and equivalence unit of analysis (person/household) and is categorized by a quality level (low/average/high), which refers to the transparency of the primary data collection. We set these categories such that we extract the largest data series with a uniform interpretation of the Gini coefficient.

We restrict our sample $dt \in \mathbb{O}$ from those country-year pairs for which the Gini in terms of disposable income is recorded. The choice of net income as welfare concept is consistent with our theory and furthermore motivated by the higher number of high quality Gini data in the WIID (Version 2020) compared to Gini data in terms of consumption, which would be a natural alternative welfare concept.[12] This choice also dictates the interpretation of our trade-based Gini estimates as capturing net income inequality. We impose that the data source be ranked high or average quality, which is the case for most entries of net income Gini.[13] In addition, we set the unit of analysis to "person" and the reference population to "all". Where the WIID still has duplicate entries for a country-period after applying our selection criteria we give preference to data from specific sources in the following order: the Luxembourg Income Study (LIS) 2019, OECD 2019, OECD 2018, ECLAC 2019, Eurostat 2019. Table B.1 in the appendix gives a detailed list of all original Gini data sources.

In total, we include 443 destination-periods for our first stage estimation. A comparison of columns 2 and 3 in Table 2 shows that our selection lowers the number of country-period pairs with net income Gini data available in the WIID by 77. The map in Figure B.1 in the Appendix presents countries for which we observe the Gini for some period, i.e. the destinations in our first step sample. While Europe and America are mostly well covered, data is scarce for African and many Asian countries.

The first column in Table 2 gives the number of countries for which we will be able to estimate a Gini index with our approach. Our method achieves a substantially larger coverage than the WIID. This is true even if one counts all entries in the WIID irrespective of the Gini definition and the quality rating for the same period (974 unique country-period entries). The final four columns present the coverage of countries in four data sets that we do not use in our estimation. These data sets serve as benchmarks to evaluate our Gini estimates in section 7. The first is an updated version of the LIS (version 2020 whereas WIID contains version 2019), the second is the World Bank's Povcal (mostly rated average quality by WIID and therefore not part of our first stage sample), the third data is Solt's Standardized World Income Inequality Data (SWIID), and the last is the World Inequality Database (WID).

12. We apply our methodology also to the case of consumption inequality as a robustness exercise.

13. Only for the few country-periods for which WIID records no high quality entries, we include data from average quality sources if all other selection criteria are met.

Neither SWIID nor WID are featured in the WIID data. SWIID provides Gini estimates combining a wide range of Gini data sources by imputation anchored in LIS. WID combines survey and tax data where available and also uses imputation for missing values.[14] Both SWIID and WID cover in particular more small countries with populations of less than 1 million and countries for which trade data is unavailable such as Taiwan.

TABLE 2. Coverage in various Gini datasets

	trade-based estimate	WIID 2020		LIS 2020	WDI 2024	SWIID v 9.6	WID 2024
		all net income	selection				
1995-1997	107	74	47	25	43	145	166
1998-2000	119	68	54	27	43	155	167
2001-2003	125	64	53	13	41	169	163
2004-2006	129	69	57	40	55	182	161
2007-2009	132	60	56	41	56	185	162
2010-2012	137	63	61	46	58	176	165
2013-2015	140	61	60	43	63	166	166
2016-2018	141	55	55	34	62	140	170
Total	1,030	514	443	269	421	1,318	1,320

Notes: Luxembourg Income Study Version March 2021. World Development Indicators: accessed April 2024. World Inequality Database: accessed April 2024. SWIID: accessed April 2024. WIID: Version May 2020. Gini definition: Net income inequality. Other selection criteria see appendix.

5.3. Other data

Our proxies for trade costs (bilateral distance, dummies for shared border, official language and colonial past as well as participation in the same regional trade agreement) come from CEPII and Egger and Larch (2008). We use real GNI per capita data provided by the World Bank. Table 14 in the appendix provides a list of all variables we use including data sources.

6. Results

We first briefly discuss the results of the first step, the income elasticity estimates, before presenting our trade-based inequality measure. In the appendix, we provide our within-country income inequality estimates for all countries and all periods.

6.1. First stage: parameter estimates

In our first stage, we estimate separate gravity coefficients for each of 360 products (HS 4-digits) and the income and inequality elasticities β_{jo} and b_{jog} for each of our more than

14. We discard 71 WID entries for countries with WID Transparency Score (2023) $= 0/10$ (https://wid.world/transparency).

15000 varieties. The bottom part of Table 3 summarizes the distribution of the coefficients for the standard gravity parameters across 360 separate regressions. The coefficient on distance is on average -0.98 across all regressions and is significantly different from zero in 347 out of the 360 products as reported in the last column. All of the standard gravity coefficients are on average well in line with the literature. To assess the variation of income elasticities across varieties of the same good, we compute the interquartile range of income elasticities per good and show its distribution in the first line of Table 3. Similarly, we compute the interquartile range of inequality elasticities between varieties within each group of destinations for each product. The middle part of Table 3 shows the distribution of this statistic across all 360 products. The interquartile range captures the extent to which the imports of different varieties of the same product react differently to inequality even in the same group of destinations. Such a heterogeneity across varieties is a necessity for our method as our second step relies on precisely on the within-product variation of inequality elasticities. The last column provides results of a Wald test, specifying the number of products for which we can reject that the income or inequality elasticities are the same across varieties. The large share of products for which these differences are significant shows that imports of varieties within products provide relevant variation.

To get a sense of the trade patterns underlying our income elasticity estimates, we plot in Figure 2 the weighted average income elasticity of an exporter against its log average per capita income over the period. The standardized[15] income elasticity is weighted by the product's share of the country's total exports. Panel (a) of Figure 2 shows a close to flat relationship when using all products. The high income elasticity we find for some exporters with a low per capita income appears to be driven by South-East Asian economies, which are heavily specialized in textile and leather. As shown in panel (b) of Figure 2, removing those industries makes the relationship weakly positive. South-East Asian countries are highly integrated in global textile value chains. Since firms from high-income markets optimize production costs and outsource labor-intensive, final production stages to these countries, it is not unreasonable to find large exports of final consumption goods from low-income to high-income countries. Unfortunately, value added data are not available at a comparable product disaggregation nor for a comparable number of countries as are data on gross import flows. For our method, integration into global value chains creates a challenge only if it implies that country exports different varieties to different markets depending on the destination's income. We address this concern in section 8.

6.2. Second stage: inequality estimates

Our second stage provides estimates of G_{dt}, which we transform into a Gini index using the mapping explained in the appendix B.2. We interpret the Gini index in terms of net income as dictated by the data choice in the first step.

Table 4 presents summary statistics of our Gini estimates over the full sample period. Our method generates estimates for 1142 country-periods, with an average value of the

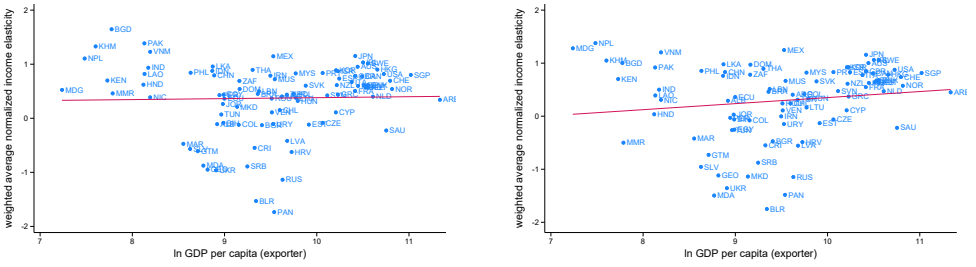
15. Income elasticities are normalized by the product mean and standard deviation.

TABLE 3. First step results, summary statistics.

	Incomegroup	Region	Mean	SD	p25	p50	p75	reject null
$\beta_{jo}^{p75} - \beta_{jo}^{p25}$			1.22	0.64	0.94	1.15	1.35	350
$b_{jog}^{p75} - b_{jog}^{p25}$	high	Americas	2.40	0.93	1.81	2.28	2.86	281
		Europe	3.81	1.34	2.88	3.53	4.41	322
		RoW	2.85	0.89	2.22	2.80	3.37	273
	mid/low	Americas	1.23	0.43	0.99	1.19	1.40	305
		Europe	2.99	1.07	2.35	2.85	3.40	286
		RoW	0.82	0.28	0.64	0.77	0.95	337
colonial ties			0.53	0.41	0.39	0.57	0.73	269
common language			0.50	0.28	0.33	0.49	0.63	280
contiguity			0.75	0.40	0.55	0.69	0.87	337
distance			−0.98	0.45	−1.22	−1.01	−0.79	347
RTA			0.18	0.34	0.05	0.14	0.28	140
Number of obs.			10405.90	6095.61	5237.50	9814.00	14657.50	.

Notes: Summary statistics for within-HS4-interquartile range of preferences parameter estimates, and of trade cost parameter estimates. The last column shows the number of products for which we can reject the null hypothesis of joint nullity of the preference parameter estimates, respectively the number of goods for which the trade cost parameter estimates are significantly different from 0 at the 5% confidence level. Total number of products is 360.

FIGURE 2. Income elasticity estimates across exporters



(A) All products (B) Excluding textile and leather

Notes: The figure shows the relationship between the weighted average of the normalized income elasticity $\frac{\hat{\beta}_{oj} - \bar{\hat{\beta}}_j}{s.d._j(\hat{\beta}_{oj})}$ and the natural logarithm of the exporter's average GDP per capita over the sample period 1995 - 2018. The weights are the product share in total exports over the sample period. Countries exporting less than 10 products are omitted from the graph. Products of HS Chapters 41-43 (leather) and 50-67 (textiles and footwear) are excluded in panel b.

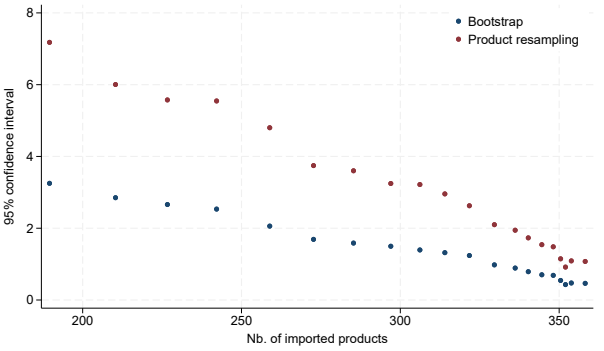
TABLE 4. Baseline Gini estimates: summary statistics.

	Mean	p10	p25	p50	p75	p90	N
Gini, all estimates	43.13	28.87	33.42	45.11	52.57	56.71	1142
Gini, retained set	42.78	28.60	33.20	44.41	52.42	56.84	1030

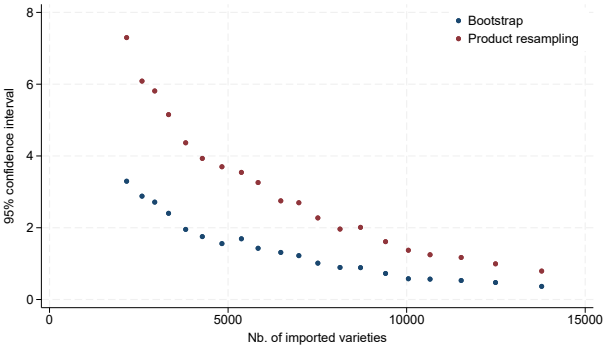
Notes: Summary statistics for net income Gini estimates. Row 1 reports statistics including all estimates. Row 2 reports the statistics after discarding the most noisy estimates.

Gini of 43.1. The full dataset, including bootstrap standard errors of both the wild and the moving-block bootstrap is available in Table 3 of the online appendix. The size of

FIGURE 3. Binscatter plot of confidence intervals



(A) Products



(B) Varieties

Notes: Binscatter plot of the size of the confidence interval for our Gini estimates as a function of the number of imported products (panel a) or imported varieties (panel b). Confidence intervals constructed with wild bootstrap in blue and product resampling (block bootstrap) in red (see section 4 for details).

the 95% confidence intervals generated by both methods are highly correlated (0.91), but the resampling on products generates wider confidence intervals. While most confidence intervals seem reasonably tight, some are extremely large. Figure 3 plots the link between the width of the confidence intervals and the number of imported products or varieties. Our procedure relies on imports to recover inequality and it is unsurprising that if a country imports only a few different products or varieties, our estimate will be very sensitive to random variations in those dimensions.

In light of these arguments, we exclude countries with numbers of varieties or of products in the bottom 10% of the distribution from our subsequent analysis. The bottom decile falls below 175 products and 1847 varieties. The second line of Table 4 reports the distribution of our Gini coefficient excluding those country-time pairs with too few imports. The number of country-time pairs that we consider for the analysis reduces from 1142 to 1030.

The world map of income inequality in Figure 4 depicts our trade-based inequality estimates for 2016-2018. Darker shades indicate higher inequality, countries in white lack data and countries in grey are those with two few imported varieties or products.[16] As evident from the map, our method can shed light on inequality in regions that would otherwise be left blank. Panel (a) of Figure 6 shows the correlation between our Gini estimate and WIID.

16. Data may be missing for two main reasons: no trade data (e.g. North Korea) or no data on GNI per capita (e.g. Venezuela 2016-2018).

FIGURE 4. Net income inequality estimates, 2016-2018.

Notes: The map shows within-country inequality for the period 2016-2018. Darker shades indicate higher income inequality. Within-country income inequality is estimated using equation (11) and transformed to the Gini index using the mapping described in section B.2..

Journal of the European Economic Association
Preprint prepared on November 4, 2024 using jeea.cls *v1.0.*

To illustrate our import-based Gini dataset, we present here some descriptive graphs about the evolution of inequality by income groups of countries (according to the World Bank definition) and by world regions. We show the evolution of the median Gini in a group over time.

Panel (a) of Figure 5 shows a slight decrease in inequality between 1995 and 2018 on average in the world. Inequality is lowest and stable among the most advanced economies while it is higher in all other categories. Upper middle income countries see a surge in inequality, from 36 to 48 Gini points while low income countries see their median income inequality increase from below 55 to 60 over the period.[17]

Panel (b) of Figure 5 shows a similar graph by world region. For consistency over time, we only consider countries within a region for which we have inequality data in all periods. With our import-based method, 20 countries in Sub-Sahara Africa, 12 in MENA and 12 countries in East Asia and Pacific fulfill these criteria. Those that do not either have missing data on income per capita in some period, or do not import enough varieties and products in all periods as specified above. Constructing such a graph based on WIID would not be meaningful, as these world regions would have at most one or two countries with consistent data on inequality over the period. A striking pattern of three inequality levels emerges. In line with survey data, Latin America stands out as one of the most unequal regions but following a decreasing trend. Sub-Sahara Africa and South Asia feature similarly high levels of inequality but are on a rather increasing trend, with Sub-Sahara Africa the region with highest inequality according to our estimates. North America, the East Asian and Pacific countries and the Middle East & North Africa (MENA) region form a middle group. Income inequality within MENA increases markedly since 2002. In our initial period, MENA's median Gini is more than 10 points below Latin America's, but both regions converge to the same level in our last period. Such an evolution for MENA would not be visible in the scarce survey data for countries in this region and as a result the Arab Spring revolts in 2011 caught many observers by surprise.[18] On the other side of the inequality spectrum, Europe and Central Asia have the lowest levels of net income inequality.

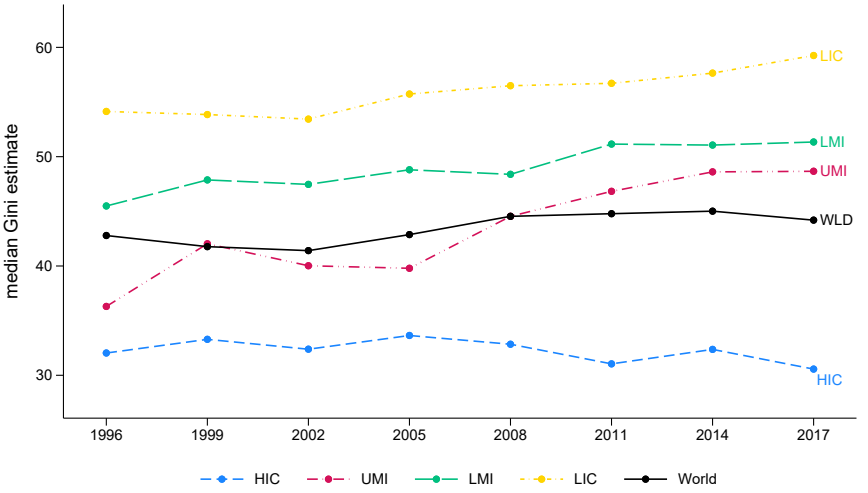
6.3. Refinement results

Table C.1 in the appendix reports the number of goods by sector that come out as the most informative about inequality in the forward and backward subset selection. Among the goods selected are some goods that would typically include luxury varieties, such as perfumes and wrist watches, but also more basic goods are selected. In general, we find that the procedure favors goods with a higher within-product variance of both the income and inequality coefficients and are positively correlated with the quality ladder length from Khandelwal (2010).

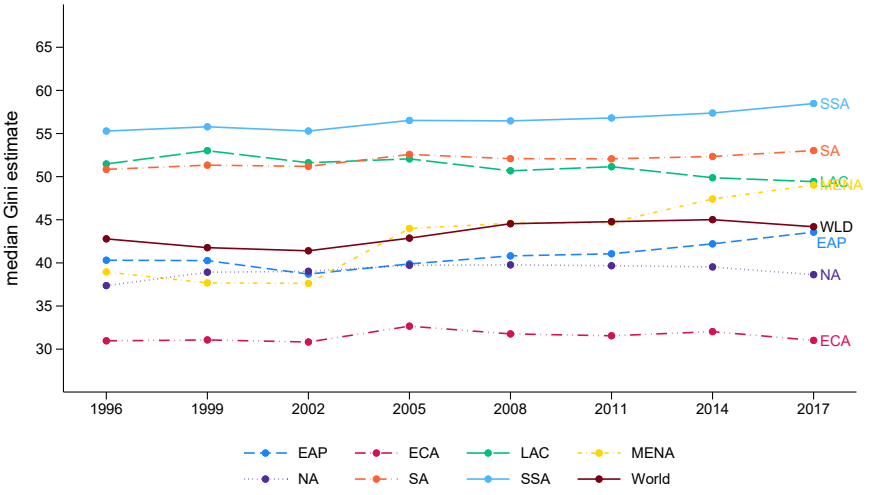
17. Note that the classification of a country into an income group can change over time in line with the World Bank definition. The evolution of the median Gini by group is however not sensitive to these changes.

18. The World Bank refers to this as the 'Arab inequality puzzle' in Ianchovichina et al. (2015).

FIGURE 5. Time series for Gini aggregates



(A) Income groups



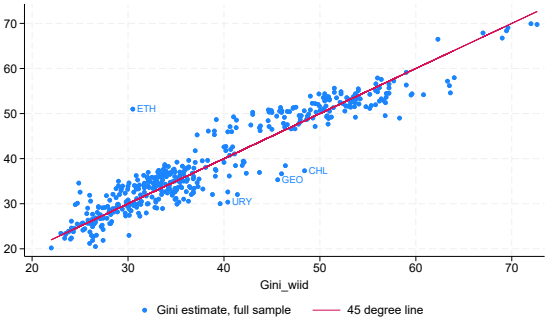
(B) World Region

Notes: The plots show the time series of the Gini coefficient for income inequality in different income groups and world regions. We report the median in each group.

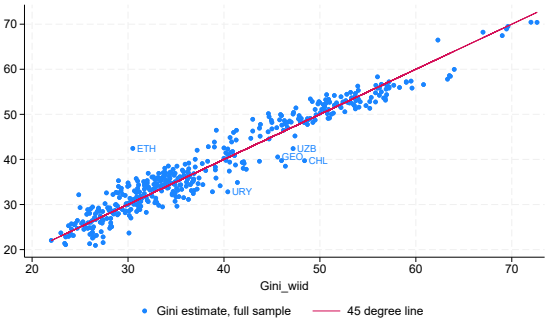
Panels (b) and (c) in Figure 6 present the correlation between the Gini estimates obtained from the selected goods and the WIID data. The plots show that Gini estimates based on

selected goods are closer to the WIID data than those using all products. The fit improves in particular at Gini levels above 40 and outliers are reduced.

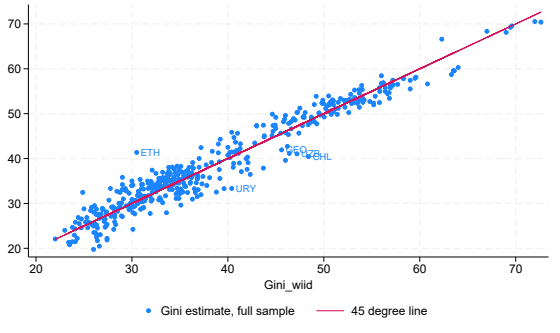
FIGURE 6. Correlation with WIID using subset selection post-estimation



(A) Full sample



(B) Forward selection



(C) Backward selection

Notes: These plots compare the net income inequality per country-year according to different Gini estimates with the values in the WIID data (used in the first estimation step). Countries labeled if difference between Gini estimate and WIID value larger than 10 points in the baseline.

7. Validation of inequality estimates

We now turn to different tests of the validity of our method.

7.1. Comparison with WIID (out-of-sample)

The first validation exercise is a comparison of our predicted measure of inequality to the WIID data. Since we use the WIID data as input in our first-stage, we use a 10-fold cross-validation to avoid a mechanical correlation between the two. We randomly split our first-step sample of country-time pairs with WIID Gini data into ten subsamples. We re-estimate step 1 ten times, dropping one subsample (the "testing data") in each round. We then estimate step 2 for the country-periods in the testing data. We now evaluate the out-of-sample Gini estimates against the WIID data.

Table 5 shows the median absolute deviation of the out-of-sample predicted Gini from the WIID Gini in terms of levels, first differences and long differences. The median absolute deviation is 2.6 points for Gini coefficients in levels. Roughly one half of our estimates exceed the first stage Gini values reported by WIID, while one half are below. We also split the sample into terciles of average income per capita over the period[19] and examine the absolute deviations between our Gini measure and WIID by tercile. The median difference for the bottom tercile is slightly higher, with around 4 Gini point deviation in levels. The top two terciles tend to be closer to WIID. The middle part of Table 5 shows the same statistics computed on first (3-year) differences. The median absolute deviation is 1.7 Gini points, with again a better fit for the richer countries in the sample. The bottom part of Table 5 reports similar statistics for long differences in inequality changes. WIID reports at least two periods of data with high quality for 63 countries. We take the longest available difference available in WIID - 21 years for most countries - and compare it to the corresponding one in our cross-validation exercise. The median absolute deviation in terms of long difference is 2.8, which appears reasonable over a long time span. This number appears substantially larger for the lower income tercile (5.8), a result largely due to observations that are not of the highest quality level in WIID for the early periods. Conditioning only on the highest quality data observations would reduce this number by half.

Table 5 also reports the same statistics for the analysis based on our forward and backward selection of products. Both forward and backward selection slightly improve the out-of-sample performance in levels by reducing the mean deviation from the WIID, but do not generate a very large improvement. They do not have a strong impact on the performance in first or long differences.

It is not straightforward to judge whether such deviations from WIID are large. 2.4 Gini points, our median deviation from WIID in levels, are smaller than the typical difference between Denmark and France, Britain and the US or Brazil and South Africa as reported in WIID. The larger difference we obtain for the bottom tercile could mean that our method

19. These terciles correspond to 941-13.541 US$, 13.5542-27.674 US$ and 27.675 - 49.820 US$ average GNI per capita over the sample period. The bottom tercile based on this sample is just below the world average (14.614 US$).

TABLE 5. Key statistics for absolute deviation from WIID of Gini estimates, baseline and selection procedures

		All		Bottom tercile		Middle tercile		Top tercile	
		Median	Mean	Median	Mean	Median	Mean	Median	Mean
levels	baseline	2.60	3.38	3.96	4.81	2.33	3.01	1.97	2.26
	forward selection	2.12	2.80	3.22	3.65	2.00	2.61	1.78	2.13
	backward selection	2.26	2.88	2.70	3.70	2.22	2.67	1.81	2.24
first differences	baseline	1.73	2.38	2.37	3.12	1.70	2.40	1.40	1.71
	forward selection	1.62	2.28	2.30	2.91	1.60	2.35	1.25	1.66
	backward selection	1.87	2.42	2.24	3.02	1.81	2.50	1.37	1.81
long differences	baseline	2.78	4.56	5.78	6.64	2.48	4.32	2.01	2.05
	forward selection	3.65	4.10	5.36	5.57	3.80	4.08	2.15	2.18
	backward selection	4.10	4.21	4.92	5.12	4.44	4.46	2.01	2.71

Notes: The table presents the median and mean absolute error for out-of-sample Gini estimates and WIID Gini data. The first three rows refer to comparison in levels, the middle three rows show period-on-period differences and the last rows show the error for changes over the maximum time span observed in WIID. The terciles are defined on the average GNI per capita distribution of the first step sample, $dt \in \mathbb{O}$.

performs less well on poorer countries but could equally signal that WIID inequality data are less precise in poorer countries, even if tagged as high quality. The lack of "true" data on inequality in poor countries makes it impossible to conclude but we turn in the next section to other available datasets and some country-specific examples to put these numbers in perspective.

We explore whether our deviations from WIID are correlated with specific characteristics of a country, such as its population size, urbanization rate or spatial inequality among others. For example, if geographically remote individuals consume less imports, our method may perform less well in countries with lower urbanization. In addition, we control for factors that could affect the validity of our import-based approach, such as the number of imported varieties, or the prevalence of re-exports, which could blur the informational content of imports. All details of the exercise are available in the online appendix section 2. The analysis shows that our deviations from WIID are a bit larger for poorer countries (as suggested by Table 5), though this is not robust to the addition of more controls. Deviations are also slightly larger for more populous countries (largely driven by China and India in some periods) and for the few countries that tend to act as trade hubs for reexports. Spatial inequality or urbanization appear largely uncorrelated with our deviation from WIID. The only factor that seems to matter consistently is the quality ranking of the WIID data: we deviate systematically more for data ranked as medium rather than high quality.

7.2. Comparison with other data sets

While it is difficult to obtain comparable data on inequality for a large set of countries, many studies measure inequality at the level of individual countries, for example using surveys. It is beyond the scope of this paper to give an overview of all data sources on inequality (see the sources for Solt (2019)'s data) and we focus on some well-established dataset combining data on net income inequality across countries and over time. We use the Luxembourg Income Study (LIS 2021) and the World Bank's Povcal data as available in March 2024, two widely used sources for survey data. The LIS is generally more focused on advanced economies

whereas Povcal also collects survey data for a large number of developing economies. A third data source is the World Inequality Database (WID). Gini values in the WID are computed with information from both survey and fiscal data and the 2024 version of WID now covers a very large range of countries. Finally, we compare our estimated Gini coefficients to Solt's SWIID database, which uses another methodology to impute inequality across countries and years in a comparable manner.[20]

Before comparing our estimated Ginis to these different measures, it is worth pointing out the very large differences in reported Ginis across different datasets. Figure C.1 in the appendix presents the median absolute deviation between the Ginis of different datasets, both in levels and in first difference, for country-time pairs in our first-stage sample. Some pairs of data produce very similar Ginis. The median difference between LIS and SWIID or between WDI and WIID are around 1 Gini point. Others are much further away, such as WID and WIID or WID and WDI, which have a median difference in Gini estimates above 6. For virtually all pairs, the differences in the bottom tercile are substantially large than in the top tercile. The numbers we obtain in Table 5 thus seem in line with the typical differences obtained across inequality datasets.

Table 6 reports the correlation coefficient (ρ) of our estimates with the inequality values in each of the five established data sets, along with the median (absolute) deviation of our estimates from the values found in the respective data sets. These statistics are shown for levels in the top panel, period-on-period changes in the middle panel and long differences in the bottom panel. Overall, we find a high correlation in levels, ranging between 0.69 for the WID to 0.95 with LIS. Note that our correlation with WID and SWIID are lower than with other datasets since they are computed on a much larger, and on average much poorer, set of countries. We report the median absolute deviation of our Gini estimate to the respective datasets on observations that are part of WIID in the fourth column ("AE, all obs. in WID") and that are not part of WIID in the last column ("AE, all new obs.").[21] The median deviation is typically much smaller on observations that are part of WIID, where countries are richer on average. The three columns bottom, middle and top split observations that are part of WIID by terciles in the same way as in Table 5.[22] We are particularly close to the World Bank Data, even closer than to WIID, on all terciles of the income distribution. On average, our estimated Gini are close to WIID and WDI, around 3 to 5 points higher than SWIID and LIS, and 8 points lower than WID. We report similar statistics for first-differences. The correlation with the databases that mostly cover rich countries (LIS, WIID and WDI) range from 0.26 to 0.44, and are much lower for WID and SWIID on a larger sample. Median absolute deviations vary around 1 for most datasets. Comparing long differences shows a correlation around 0.5

20. Whereas an earlier version of the LIS (from 2019) is among the selected sources from the WIID, we make no use of any version of Povcal or WID data in our first stage estimation.

21. The latter deviations are very large, with a median deviation of around 8 Gini points from SWIID and WID for the sample with no WIID data. Our estimates however fall between these two datasets, which have a median absolute difference of 15 Gini points on this sample.

22. Note that the numbers for the comparison with WIID are now different from Table 5 as we report the in-sample Gini estimates, i.e. not from the k-fold cross-validation.

TABLE 6. Absolute deviation of Gini estimate from various datasets, baseline in-sample.

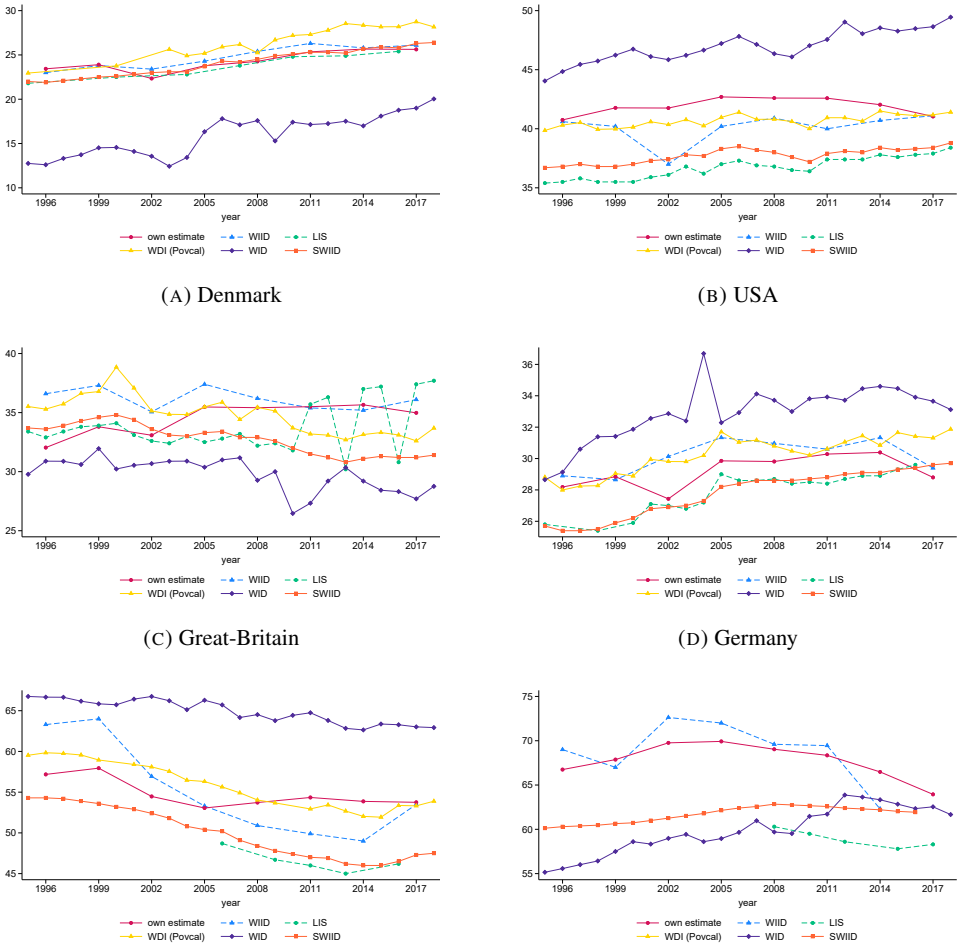
	Data	Stat.	Correlation	AE, all obs. in WIID	AE, bottom	AE, middle	AE, top	AE, all new obs.
Levels	LIS	p50	0.95	3.05	5.93	3.30	2.43	5.65
		N	250	247	44	87	116	3
	SWIID	p50	0.77	2.85	4.07	2.69	2.35	8.31
		N	947	440	148	149	143	507
	WDI	p50	0.95	1.86	2.17	2.01	1.58	4.03
		N	381	353	106	112	135	28
	WID	p50	0.69	6.12	11.06	6.24	2.61	8.76
		N	986	427	135	149	143	559
	WIID	p50	0.95	2.11	2.76	1.97	1.75	.
		N	440	440	148	149	143	0
First differences	LIS	p50	0.28	0.84	0.74	1.14	0.64	3.04
		N	174	173	23	64	86	1
	SWIID	p50	0.13	0.89	1.06	0.93	0.69	1.22
		N	802	391	132	133	126	411
	WDI	p50	0.22	1.04	1.43	1.13	0.84	1.75
		N	302	290	88	87	115	12
	WID	p50	0.02	1.25	1.47	1.55	0.87	1.59
		N	841	377	118	133	126	464
	WIID	p50	0.41	1.34	1.84	1.34	0.96	.
		N	358	358	112	122	124	0
Long-differences	LIS	p50	0.48	2.26	2.23	3.93	1.42	3.09
		N	44	43	10	15	18	1
	SWIID	p50	0.18	2.21	3.35	2.99	1.33	3.86
		N	137	58	19	21	18	79
	WDI	p50	0.46	2.50	2.83	3.25	1.03	8.56
		N	57	54	17	19	18	3
	WID	p50	0.02	2.43	1.46	3.70	1.38	4.56
		N	138	55	16	21	18	83
	WIID	p50	0.53	2.45	4.30	2.69	1.49	.
		N	63	63	24	21	18	0

Notes: The table presents the correlation coefficient and median absolute deviation between our Gini estimates (in-sample) and different Gini datasets. The statistic, correlation coefficient or median, is reported in the first row and the number of observations on which it is computed is reported in the second row per data set. The top part refers to levels and the bottom part to first-differences. The column header for columns 5-9 indicates the subsample for which the statistic is computed. Bottom, middle and top refer to the terciles of the average GNI per cpaita distribution of $dt \in \mathbb{O}$.

with LIS, WIID and WDI, and below for SWIID and WID. Median absolute deviations vary around 2.3 in this case.

To further illustrate our results and assess how they compare to other datasets, we plot in Figure 7 the evolution of the Gini coefficients for 6 countries for which we have high-quality WIID data for almost all periods. We report the evolution of their Gini coefficients for net income inequality as given by the different datasets. Our estimates for Denmark are very tightly in line with most datasets except for WID, showing a low and slightly increasing inequality over the period. In all other cases, the evolution we present is in the (wide) range of estimates from different datsets, with a relatively higher inequality in the US than most datasets except for WID. In the case of Brazil, we show a decrease in inequality over the period, like all other datasets. The decrease we estimate is however less strong than most other data. Finally, for South Africa, we also estimate a very high level of inequality, which exhibits a hump-shaped pattern, tightly linked to the WIID data. This is in contrast to the WID data, which sees inequality raising over the whole period, but closely in line with recent evidence from South Africa using the Distributional National Accounts method (see Chatterjee et al. (2023)).

FIGURE 7. Time series comparison with other data sets, countries in WIID.



(A) Denmark (B) USA

(C) Great-Britain (D) Germany

(E) Brazil (F) South Africa

Notes: These figures compare the net income inequality time series per country according to our Gini estimate with the time series from different survey sources. LIS: Luxembourg Income Study; WDI: World Development Indicators; WIID: World Income Inequality Database; WID: World Inequality Database; SWIID: Standardized World Income Inequality Database. The WIID data is used in the first estimation step. Labels on the horizontal axis indicate the middle year of a three-year period.

We then show the evolution predicted by our estimates for countries that never appear in the WIID data, $dt \in \mathbb{U}$. We select countries from the bottom tercile of the income per capita distribution in that sample over the period (approximately 2750$ on average). We pick three countries with large numbers of imported varieties in that group (Kenya, Tanzania and Bangladesh). Similarly, we show data for the three countries in the bottom decile of income per capita (approximately 1400$ per capita) with the highest number of imported varieties.

We report on average very high income inequality for all these countries, featuring a rather increasing pattern over time except for Bangladesh, which exhibits a slightly decreasing trend.

It is difficult to judge the validity of these estimates as existing data on these countries are very poor. Just for illustrative purposes, we present the data by WID and SWIID, the other two datasets that aim to cover a very large number of countries. We are consistently close to WID and pretty far from SWIID among very poor countries but not necessarily among middle-income or richer ones.

8. Extensions and Discussion

We now discuss how sensitive our results are to a number of assumptions underlying our estimation. Each robustness exercise gives rise to a new vector of parameter estimates $\hat{b}_{jog}$ and import gaps $\widetilde{\ln(X_{jodt})}$, resulting in a new set of estimates for inequality. We report in Table 7 at the end of this section the correlation between the Gini estimates of each robustness exercise with the baseline estimate and the performance of each exercise compared to the WIID data as in the previous section.

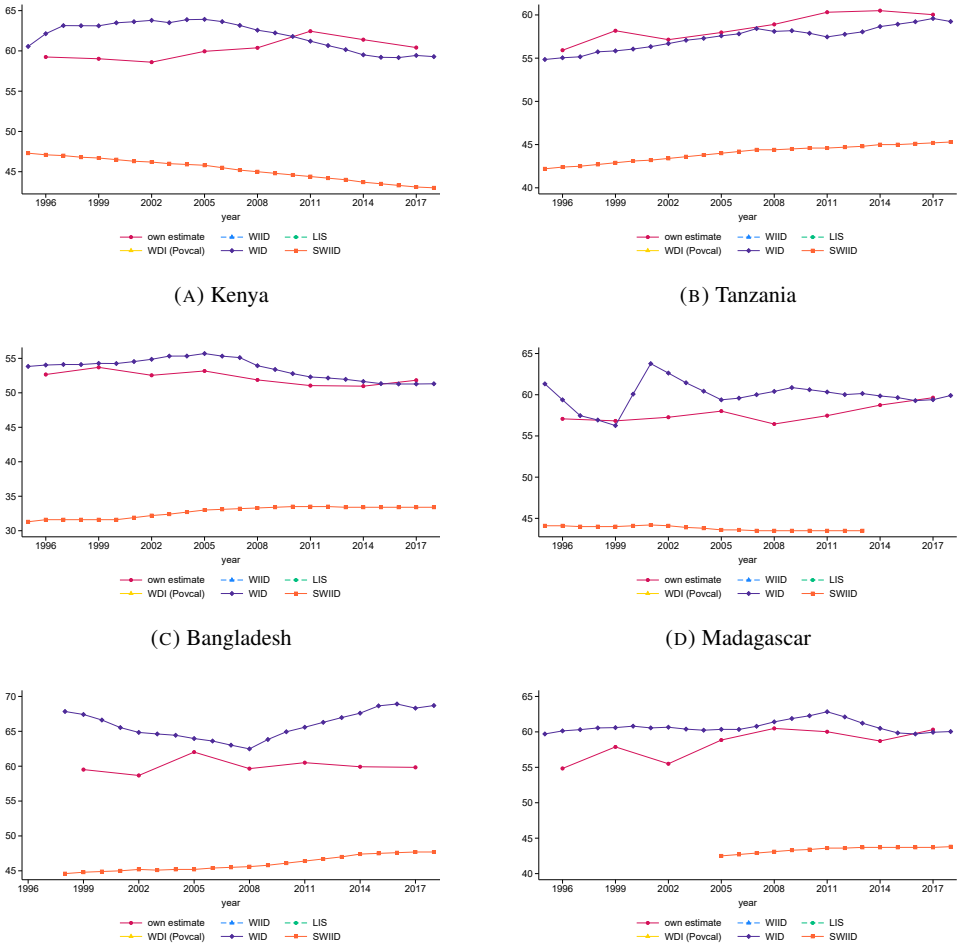
8.1. Modelling assumptions

Armington varieties. Our analysis relies on a definition of varieties in the Armington sense, e.g. German cars are a variety. However, in practice, German cars consist of different brands, with potentially different preference shifters or income elasticities. Since the brand mix will likely be correlated with the destination's income (e.g. Hallak (2010)), this may introduce non-trivial measurement error in our empirical analysis. We show in the online appendix section 1 that this would amount to making the error term depend on the quality differentiation within an exporter-product pair and the income of the destination, potentially biasing our estimate. To limit the bias from differentiation within Armington-varieties, we recompute inequality based on the 182 products (half of the initial number) that are least differentiated. We measure differentiation by the variance of a variety's unit values across destination and select products with the lowest weighted-average variance across varieties.[23] As reported in Table 7, the Gini estimates resulting from this specification are highly correlated with our baseline, and the out-of-sample performance is also comparable to the baseline model.

Pricing to market. A higher income level in a destination may be associated with higher consumption prices if firms price to market (Simonovska, 2015) or if distribution costs depend on income. Even within countries, firms may charge higher prices to richer consumers and the distribution costs may be correlated with income if, for example, poorer households live in more remote places. While our model abstracts from such concerns by assumption (perfect competition and cost structure), such correlations could bias our estimates by making

23. The average is weighted by the world exports of the variety.

FIGURE 8. Time series comparison with other data sets, countries not in WIID.



(A) Kenya (B) Tanzania

(C) Bangladesh (D) Madagascar

(E) Mozambique (F) Togo

Notes: These figures compare the net income inequality time series per country according to our Gini estimate with the time series from different survey sources. LIS: Luxembourg Income Study; WDI: World Development Indicators; WIID: World Income Inequality Database; WID: World Inequality Database; SWIID: Standardized World Income Inequality Database. The WIID data is used in the first estimation step. Labels on the horizontal axis indicate the middle year of a three-year period.

the error term dependent on income.[24] Online appendix section 1.2 develops an extension of our model where each Armington variety is produced by a monopolist. If firms cannot price

24. To the extent that the marginal effect of income on prices is constant across countries, it is part of the variation of interest, and our estimate of β_{jo} would capture the impact of income on imports that goes both through preferences and prices.

discriminate across consumers within countries, their markup can be proxied by the share of their variety in a destination's imports. The second line of Table 7 reports our estimates when controlling for the predicted share[25] of a variety in a destination based on a homothetic gravity equation estimated at the corresponding HS2 level. This has very little effect on the outcome. The test we perform to assess the importance of product differentiation within Armington-varieties can also be informative about price discrimination as it effectively limits the set of products to those for which prices do not differ too much across countries, i.e. for which there may be less price discrimination. As mentioned before, differences with the baseline results are only minor. Online appendix 1.2 discusses an alternative model where firms can price discriminate within countries and how to account for this in our setup.

Time-varying parameters. In our baseline analysis, we allow some preference parameters to vary over time (the product level demand shifter μ_{jdt} or the idiosyncratic taste parameter α_{jodt}), but assume that our main parameters of interest, β_{jo} and b_{jog}, are constant over time. These parameters could however have changed over the period. We replicate our analysis by splitting the estimation in two sub-periods, allowing all coefficients to vary by decade. The Gini estimates from this exercise are positively correlated with our baseline estimates. The out-of-sample performance improves a bit compared to our baseline model.

Regional preferences. Our baseline estimate assumes that preferences for varieties are the same across countries: the income elasticity parameter β_{jo} is independent of the destination and the taste shifter α_{jodt} is drawn from the same distribution for all destinations. While our groups capture that $\bar{\beta}_{jd}$ and ε^S_{jd} differ by groups, they do not allow for a real heterogeneity in preferences across groups. We now allow for the coefficient on income per capita (β_{jo}) to depend on the region of the destination country. This means that, for example, Chinese varieties can now have a different income elasticity in Europe, Asia/Africa, or America. Beyond the income elasticity, we also introduce a variety-world region fixed effect to allow for varying demand shifter. This would capture for example that the preference for a French variety differs across world regions, where we define these regions based on UN subregions. The results, reported in the fourth line of Table 7 are very correlated to the baseline and show a slight improvement in the out-of-sample performance.

The CES structure. Another concern is that CES preferences imply that the elasticity of substitution is the same between all varieties of a good. The ratio of consumption of any two varieties of a good is thus independent of the consumption of other varieties. We relax this assumption by allowing for a nested CES utility with two nests per product. The elasticity of substitution between any two varieties of the same nest is then equal, but the elasticity across nests is allowed to differ from the one within nests. Such a structure allows for a different substitution elasticity between varieties of different quality levels, for example. In practice, it requires that we replace the product-destination-time fixed effect by a nest-product-destination-time fixed effect. We replicate our analysis defining varieties from OECD and non-OECD countries as two separate nests. The fifth line in Table 7 shows that this additional flexibility in our model does not affect our results much. The Gini estimates are highly correlated with those of our baseline model and the out-of-sample performance

25. We use the predicted share rather than the actual share to mitigate endogeneity concerns.

similar to the baseline model. This extension can also help us address the lack of data on domestic varieties. With our CES structure, the relative imports of foreign varieties do not depend on the consumption of domestic varieties. If domestic varieties were not equally substitutable with all foreign varieties, however, a price change of domestic varieties would affect relative imports, which we may misinterpret as a result on inequality. The different CES nests capture such heterogeneity in substitution elasticities.

Savings. Our model assumes that households spend their whole income on consumption. In practice, some households save a fraction of their income and the propensity to save may be correlated with income. This would break the link between expenditure and income, and begs the question whether expenditure is a better proxy for the level of utility, and hence import patterns, than income. We do not explicitly model how savings (whether positive or negative) would enter utility in such a setup, but believe that they would matter, creating a tighter link between income and utility than expenditure and utility. Still, we show in the online appendix section 1 that our baseline estimates remain valid even if expenditure governs the demand shifters as long as long as the elasticity of expenditure to income is constant. If this is not so, our method yields a measure that is conceptually closer to consumption inequality. We therefore replicate our whole analysis using an inequality measure based on consumption rather than income, but show in Table 7 that it differs more from WIID in the cross-validation exercise.

8.2. Informativeness of imports

In our model, all households consume all varieties. Under this assumption, imports contain information about individuals across the whole income distribution. One may be concerned, however, that our model predicts that spending on varieties with a high income elasticity approaches zero - but is not zero - among the very poor while subsistence farmers in Sub-Saharan Africa may not buy any Swiss watches in reality. This could be an issue for our strategy if the very poor do not consume imported but only domestic varieties of most products as then imports do not convey any information about the poor's income. To address these concerns, we first run our analysis limiting the set of goods to those that are most likely consumed and imported by households throughout the income distribution.

Quality ladder. First, we limit our analysis to goods with a "long quality ladder" according to Khandelwal (2010) and where "long" means to have a quality ladder length above the median.[26] Goods with long quality ladders are those for which very high and very low quality is exported worldwide. Thus, these goods are potentially affordable to and consumed by very diverse groups of consumers. Restricting to goods with a longer quality ladder does not affect the out-of-sample performance, neither on average nor differentially across the groups.

Goods consumed by the poor. As an alternative approach we build on Banerjee and Duflo (2007), who summarize insights from surveys about spending of households with less than

26. Khandelwal (2010) computes quality as a residual from regressing market shares on prices and a set of fixed effects.

1$, respectively less than 2$, per day. We select those HS 4 codes that correspond to products mentioned in Banerjee and Duflo (2007). Among these products are food items, alcoholic beverages, tobacco, medication, personal care, textiles, jewelry and watches, radio receivers and TVs, and bicycles.[27] These goods are likely also consumed by households across the entire income distribution and therefore may carry the most relevant information for our purposes. A similar conclusion applies as in the case of long quality ladders.

Non-exported goods. As a third exercise, we constrain the set of products to those where nearly all expenditure must go to imports. Because disaggregated production data is scarce for developing countries, we identify those goods from trade data. We restrict the sample of goods for which exports of a country are less than 5% of its imports. While we cannot rule out that a domestic variety is available, low exports and large imports suggest that the majority of households consumes a foreign variety and thus expenditure shifts due to changes in the income distribution will be reflected in imports. The results in Table 7 show that the out-of-sample performance for the bottom tercile is very stable relative to the baseline. It is somewhat weaker for richer countries, which export a large number of products and for which the constraint is more binding.

8.3. Alternative empirical strategies

Our first stage estimates use the standard within-transformation for twoway-fixed effects models. This estimator could be biased for two reasons: first, since we are more likely to observe inequality in high-income countries, the estimator could suffer from sample selection bias. Second, due to the log transformation of trade values, we drop zero trade flows from the regression sample.

Reweighting. To address potential sample selection bias, we re-estimate step 1 using weighted least squares with inverse probability weights. We compute the probability weights by splitting the destinations in quintiles of the per capita income distribution of the full sample. As expected, most countries in the step 1-regression sample are in the top quintile. We thus reweight all destinations in the lower quintiles to contribute the same number of observations as the top quintile to the regression sample, $w_q = N_{q5}/N_q$. Table 7 shows that the out-of-sample performance barely changes overall but modestly improves the results for the poorer countries.

Zeros trade flows. Our baseline estimation equation is log linear and therefore drops all trade flows equal to zero. This can downward bias our first stage estimates when a country exports only to rich or unequal countries and not at all to poor, equal economies. In addition, as trade flows are bounded from below, error terms will exhibit heteroscedasticity. To reduce these concerns, we re-estimate the first stage with PPML (Silva and Tenreyro, 2006) and continue estimating the second stage with OLS[28]. We include a zero for imports of a variety if the country ever imported any variety of a product. About 25% of our PPML-first stage

27. The list of HS 4 codes is provided in Table B.3.

28. The second stage dependent variable is not bounded by zero and so the motivation to use PPML does not apply here.

sample are zero flows. The correlation with the baseline estimate and the out-of-sample performance drop somewhat.

Group definitions. Next, we experiment with different group definitions. As stated in section 4, we form groups based on world regions and income groups and assume a constant curvature parameter $b_{jog(d)}$ of the Engel curve. This is due to the fact that countries within groups should have a similar composition of imported varieties, and therefore a similar average income elasticity of imports and product-level curvature ($\bar{\beta}_{jd}$ and ε^S_{jd}). To match this logic more closely, we group countries by the similarity of their import share vectors, with the requirement that groups span both countries with and without inequality data in WIID.[29] We compute the vectors of import shares, on which we base the clustering algorithm, in two distinct ways. "Group definition 1" computes import shares based on actual import flows. "Group definition 2" uses predicted shares from a gravity equation that includes the same gravity controls as in our baseline (listed at the bottom of Table 3), as well as a dummy for countries on the same continent and controls for the stock of bilateral migrants (from the World Bank's Global Bilateral Migrants Database). We also control for destination-time and origin-time fixed effects as well as the product of income per capita of the origin and destination. Both of these alternative grouping strategies yield very similar results to our baseline.

Unit values. Finally, we test whether controlling explicitly for the log unit value of the flow in our specification matters for our results. From a theory perspective, prices are a key determinant of the share of spending on each variety (see 3), and our baseline controls for prices through a combination of gravity terms and origin-time fixed effects. Controlling for unit values may also address problems linked to pricing to market, but comes at the risk of a simultaneity bias, since unit values are computed based on expenditure, our left-hand side variable. We show in Table 7 that our results almost identical to our baseline.

29. The grouping is executed per HS Chapter (in total 15) and based, for each importer, on the share of expenditure on each variety over 24 year aggregate expenditure. Next, we calculate the cosine similarity of importers in terms of these shares. We group the countries $dt \in \mathbb{O}$ into three groups using Ward's linkage method. For each country $dt \in \mathbb{U}$, we calculate the average similarity across countries $dt \in \mathbb{O}$ by group and assign the country $dt \in \mathbb{U}$ to the group with the highest average similarity across members.

TABLE 7. Absolute deviation of Gini estimate from WIID, other than baseline out-of-sample

	Extension	Correlation	AE, all obs. in WIID	AE, bottom	AE, middle	AE, top
Theory assumptions	Homogeneous varieties/Armington	1.00	2.60	3.71	2.46	1.87
	Pricing to market	1.00	2.60	3.97	2.33	1.97
	Time-varying income elasticities	0.98	2.13	2.50	2.20	1.84
	Regional preferences	0.95	2.41	3.37	2.25	1.97
	CES/substitutability	1.00	2.51	3.57	2.10	1.97
	Savings	0.92	3.06	7.12	3.55	2.97
Informativeness of imports	Long ladder goods	1.00	2.58	4.11	2.46	1.86
	Banerjee, Duflo-goods	0.99	2.70	3.84	2.41	1.98
	Foreign goods	0.97	3.23	3.98	2.60	3.11
Empirical implementation	WLS	0.98	2.46	3.41	2.26	1.90
	PPML/Zero trade flows	0.79	4.10	9.26	4.31	1.83
	groups definition 1	0.94	3.11	4.55	2.50	3.02
	groups definition 2	0.95	2.86	4.17	2.63	2.64
	log unit values	1.00	2.62	3.93	2.27	1.97

Notes: The table presents correlation coefficients between the baseline and robustness exercise Gini estimates along with the median absolute error with WIID, in total and by tercile. Terciles are defined on the average GNI per capita distribution among $dt \in \mathbb{O}$.

9. Conclusion

In this paper, we propose a novel method for measuring inequality in a country based on its imports of consumer goods. We rely on a model that embeds non-homothetic preferences in an otherwise standard model of international trade à la Armington, in which goods are differentiated by country of origin. This setup generates a non-homothetic gravity equation, where the imports of a variety by a destination country depend on its average income and on the Gini coefficient of its income distribution.

Our empirical approach follows two steps. In the first step, we use a subsample of destinations with high-quality data on inequality and estimate for each variety the extent to which it is imported by richer or more unequal countries. In the second step, we reverse the logic and estimate, for all countries, the Gini coefficient that best fits the observed import patterns given the variety-specific elasticities estimated in the first step.

The key advantage of our approach is that it relies on international trade data, which are publicly and consistently recorded for virtually all countries. Our method provides a transparent mapping of trade into inequality for all countries, offering an alternative to imputation methods based on survey and fiscal data. Different cross-validation exercises suggest that our method compares well to existing high-quality data on inequality. Our approach relies on a set of assumptions, which play a key role in mapping import patterns to inequality. We experiment with a number of robustness checks, modifying some of those assumptions, and show that our results are very stable.

Authors' affiliations: ifo Institute, Ludwig-Maximilians-Universität München, CESifo, and Université catholique de Louvain

Appendix A: Theory

A.1. Derivation of equation (2)

We use a second-order Taylor approximation of $x_{jo}(\mathcal{B}, \mathbf{A_d}, I_h)$ around $I_h = I_d$:

$$x_{jo}(\mathcal{B}, \mathbf{A}_d, I_h) \approx x_{jo}(\mathcal{B}, \mathbf{A}_d, I_d) + (I_h - I_d) \left.\frac{\partial x_{jo}(\mathcal{B}, \mathbf{A}_d, I_h)}{\partial I_h}\right|_{I_h = I_d} + \frac{1}{2}(I_h - I_d)^2 \left.\frac{\partial^2 x_{jo}(\mathcal{B}, \mathbf{A}_d, I_h)}{\partial^2 I_h}\right|_{I_h = I_d} \tag{A.1}$$

Summing up over all individuals gives:

$$X_{jod} \approx N_d x_{jo}(\mathcal{B}, \mathbf{A}_d, I_d) \left[1 + \frac{1}{2} \left.\frac{\partial^2 x_{jo}(\mathcal{B}, \mathbf{A}_d, I_h)}{\partial^2 I_h}\right|_{I_h = I_d} \frac{I_d^2}{x_{jo}(\mathcal{B}, \mathbf{A}_d, I_d)} G_d\right] \tag{A.2}$$

where $G_d \equiv \left(\sum_h \frac{N_{dh}}{N_d}(I_h - I_d)^2\right)/I_{dt}^2$ is the squared coefficient of variation of income in country d. Taking logs and using the approximation $Log(1+y) \approx y$ on the square bracket gives (2).

A.2. Utility Maximization

We derive the Hicksian demand by minimizing expenditure given a utility level. The first-order condition with respect to the consumption of variety jo is:

$$p_{jod} - \lambda \varphi_{jod}(U) C_{jd}^{\frac{1}{1-\gamma_j}} c_{jod}^{\frac{1}{\gamma_j - 1}} \frac{\partial U}{\partial C_j} = 0, \tag{A.3}$$

where λ is the Lagrange multiplier. Rearranging shows that:

$$c_{jod} = \varphi_{jod}(U)^{1-\gamma_j} p_{jod}^{\gamma_j - 1} C_{jd} \left(\lambda \frac{\partial U}{\partial C_j}\right)^{1-\gamma_j}. \tag{A.4}$$

Raising equation (A.3) to the power γ_j, multiplying by α_{jod} and adding over all varieties gives:

$$\sum_o \varphi_{jod}(U)^{1-\gamma_j} p_{jod}^{\gamma_j} = \left(\frac{\partial U}{\partial C_j}\lambda\right)^{\gamma_j} \tag{A.5}$$

and replacing $\frac{\partial U}{\partial C_j}\lambda$ in (A.4) by its value from (A.5):

$$p_{jod} c_{jod} = \frac{\varphi_{jod}(U)^{1-\gamma_j} p_{jod}^{\gamma_j}}{\sum_{o'} \varphi_{jo'd}(U)^{1-\gamma_j} p_{jo'd}^{\gamma_j}} P_{jd} C_{jd} = \frac{\alpha_{jod} U^{\beta_{jo}} p_{jod}^{\gamma_j}}{\sum_{o'} \alpha_{jo'd} U^{\beta_{jo'}} p_{jo'd}^{\gamma_j}} P_{jd} C_{jd}$$

where P_{jd} is the CES price index corresponding to an individual with utility level U:

$$P_{jd} = \left(\sum_{o'} \varphi_{jo'd}(U)^{1-\gamma_j} p_{jo'd}^{\gamma_j}\right)^{\frac{1}{\gamma_j}}$$

Using that $P_{jd} = \frac{\partial U}{\partial C_{jd}}\lambda$ gives the optimal consumption of good j:

$$C_{jd} = P_{jd}^{-\rho}\lambda^{\rho}\mu_{jd}(U)^{\rho}U_d$$

Total spending is:

$$\sum_j P_{jd}C_{jd} = U\lambda^{\rho}\left(\sum_j \mu_{jd}(U)^{\rho}P_{jd}^{1-\rho}\right)$$

which implies that:

$$S_{jd} = \frac{P_{jd}C_{jd}}{\sum_m P_m C_m} = \frac{\mu_{jd}(U)^{\rho}P_{jd}^{1-\rho}}{\sum_m \mu_{md}(U)^{\rho}P_{md}^{1-\rho}}$$

Taking C_{jd} to the power $(\rho-1)/\rho$ and summing over all j gives λ:

$$\lambda = \left[\sum_j \mu(U)^{\rho}P_{jd}^{1-\rho}\right]^{\frac{1}{1-\rho}}.$$

λ, the aggregate price index, provides the link between expenditure and utility as $E = \lambda U$. λ is itself a function of U due to the non-homothetic preference parameters. Under the assumption that consumers with a higher utility do not systematically prefer varieties that have a higher or lower (quality) adjusted price (the inverse of which is captured by $\alpha_{jod}p_j^{\gamma_j}$), utility is equal to expenditure up to a constant, and we can replace utility by income in our setup.

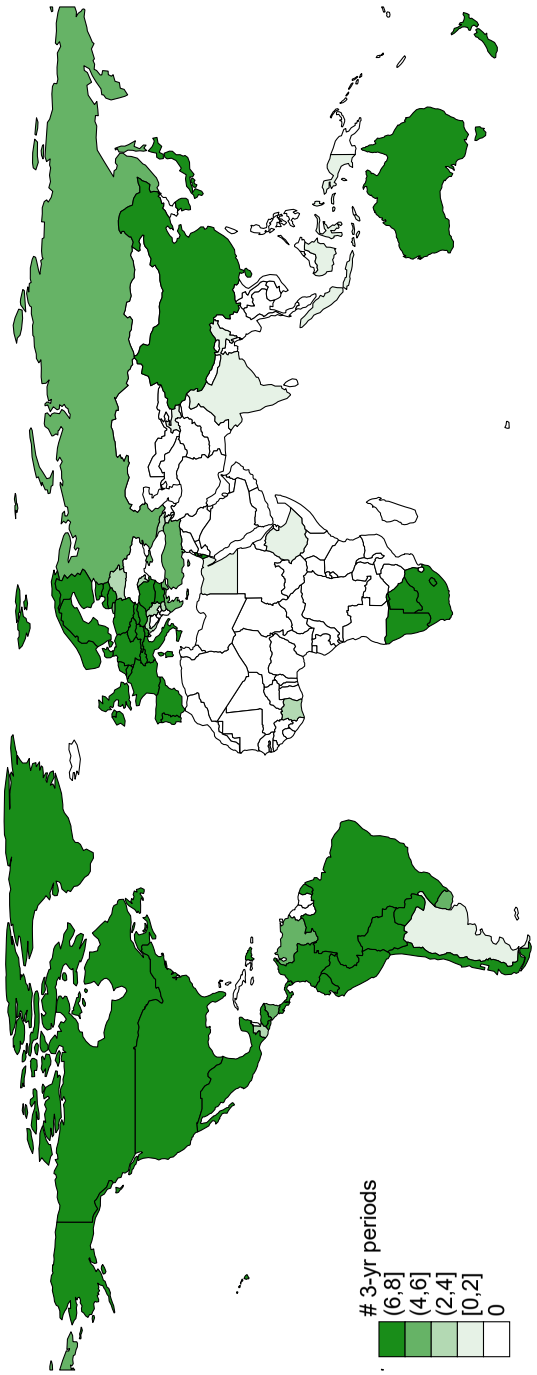
Appendix B: Data

B.1. Data sources and coverage

TABLE B.1. Data sources

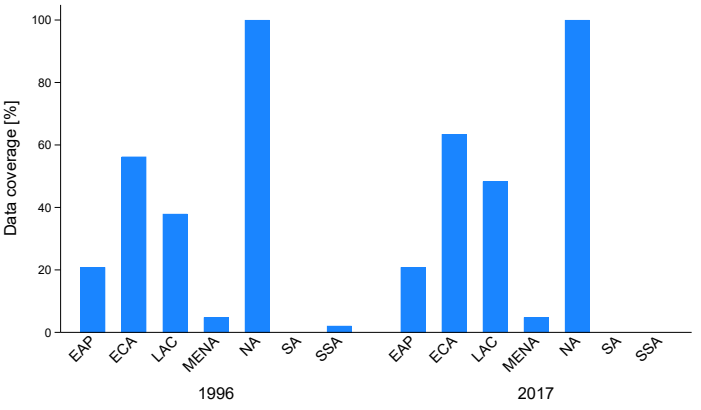
Source	Variable	Notes
CEPII/BACI	trade flows	HS 6 digit, aggregated to 4 digit. Consumption goods classified by UNCTAD. Further excluded are HS chapters 26-27 "Mineral Products", chapters 31, 34, 35, 36, 38 "Products of Chemical or Allied Industries" as well as chapters 72-81 and chapters 83, 84 of "Base Metals and Articles of Base Metal" and finally products listed under subheading 93 "Arms and Ammunition".
CEPII/Gravity	distance, colonial dummy, common official language dummy, shared border dummy	
M. Larch's Regional Trade Agreements Database from Egger and Larch (2008)	RTA dummy	
World Development Indicators	GNI per capita (PPP)	
UNU-WIDER: WIID	Gini	used in first stage. Welfare definition: "Income, disposable"; quality: "Average" & "High"; age coverage: "All"; area coverage: "All"; income unit: "Household"; unit of analysis: "Person". Further selection among duplicate entries is based on sources: LIS 2019, OECD, SEDLAC/ECLAC, Eurostat. Additionally, non-duplicate entries fulfilling the unit selection criteria listed above are included.
LIS	Gini	for comparison, accessed March 2021
Povcal	Gini	for comparison, accessed March 2024
World Inequality Database	Gini	for comparison, accessed March 2024
SWIID	Gini	for comparison, version 9.6

FIGURE B.1. Gini availability: First step sample.

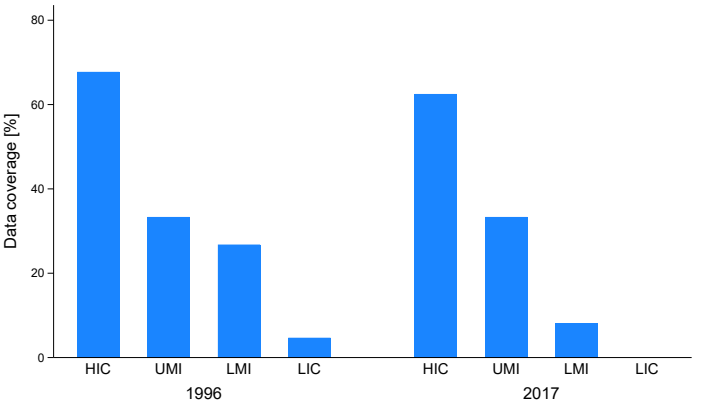


Notes: The map shows the number of periods for which within-country inequality is recorded in the WIID and fulfills our selection criteria (see text).

FIGURE B.2. Availability of net income Gini in WIID



(A) by region



(B) by income group

Notes: The chart shows share of countries in a world region/income group for which high quality net income inequality data is available in the World Income Inequality Database. Regions: EAP - East Asia Pacific, ECA - Europe & Central Asia, LAC - Latin America & Caribbean, MENA - Middle East & North Africa, NA - North America, SA - South Asia, SSA - Sub-Sahara Africa. Income groups: H - high income, UM - upper middle income, LM - lower middle income, L - low income.

TABLE B.2. HS Chapter descriptions.

HS Chapter	Title	Examples
1	Animal & Animal Products	Meat, fish, dairy products
2	Vegetable Products	Vegetables, fruit, plants
3	Foodstuffs	Edible preparations, drinks, tobacco
4	Mineral Products	Salt, chalk
5	Chemicals and Allied Industries	Soap, perfumes, fertilizers
6	Plastics/Rubber	Tableware, kitchenware, apparel of rubber
7	Raw Hides, Skins, Leather & Furs	Apparel and clothing accessoires, saddlery, cases
8	Wood & Wood Products	Paper, paper board, books, printed matter
9	Textiles	Clothes
10	Footwear/Headgear	Shoes
11	Stone/Glass	Pearls, Diamonds, Ceramic Products, Glass
12	Metals	Tools, cutlery
13	Machinery/Electrical	Computer, vacuum cleaner, dish washer
14	Transportation	Cars, bikes
15	Miscellaneous	Sports equipment, music instruments, toys

TABLE B.3. Products consumed by the poor.

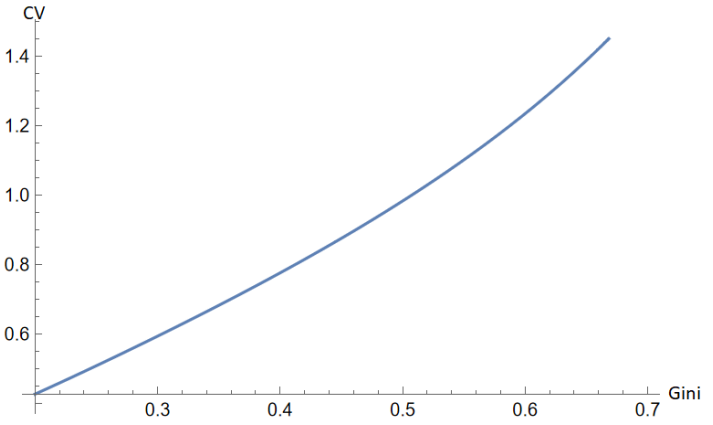
Product group	HS 4-digit code					
Vegetables	0901	0902	0903	1006	1512	1514
Foodstuffs	1704	1806	1902	2101	2203	2204
	2205	2206	2402	2403		
Medication, personal care	3004	3304	3306	3401		
Plastic tableware	3924					
Cigarette paper	4813					
Textiles	6109	6110	6205	6301	6302	6309
Jewelery	7018	7113	7114	7117		
Radio/TV	8519	8527	8528			
Bicycles	8712					
Festivity articles, misc.	9404	9505	9608	9609	9613	

Notes: The table lists HS 4 codes corresponding to products mentioned in Banerjee and Duflo (2007) to which we restrict the data for a robustness exercise (section 8).

B.2. Mapping using the Dagum distribution

We follow a common approach in the literature measuring inequality and assume that the distribution of income in a country follows a Dagum distribution (Kleiber (2008), Dagum (1977)). Using a large sample of countries, Bandourian et al. (2002) show that the Dagum distribution provides the best fit for income distribution among distributions with 3 parameters, and comes very close to less parsimonious distributions in terms of fit. The Dagum distribution has three parameters, two of which (a and p) matter for both the coefficient of variation of income and for the Gini coefficient, resulting in a non-unique mapping between Gini and coefficient of variation. To reduce the dimension of the problem, we set the parameter a at the average value in Bandourian et al. (2002), equal to 4.22. Varying p gives a unique mapping that connects the coefficient of variation and the Gini coefficient,

FIGURE B.3. Correspondence between Gini (x-axis) and coefficient of variation (y-axis) with Dagum distribution for $a = 4.22$



which we illustrate in Figure B.3. This is the relationship we use to convert the Gini from our data to the coefficient of variation that enters our model. An obvious alternative would be to fix p at its average value in Bandourian et al. (2002) and let a vary. This method would generate a more convex mapping than in Figure B.3 but would not match the highest observed Gini coefficients in WIID as they would violate the parameter restrictions of the Dagum distribution (a needs to be larger than 2 for the variance to exist). We have also experimented with other strategies to connect the coefficient of variation and the Gini coefficient varying both a and p but these less tractable methods generate similar results to our baseline.

	All	Forward selection	Backward selection
Food, Drinks, Tobacco	62	18	12
Chemical & allied industries	25	6	5
Plastics/rubber	16	3	2
Leather & furs	8	3	1
Wood products	34	12	8
Textiles	75	19	14
Footwear	13	1	0
Stone/glass	43	10	6
Metal products	13	0	0
Electronic products	12	1	0
Transportation	10	4	2
Miscellaneous	49	10	5
Total	360	87	55

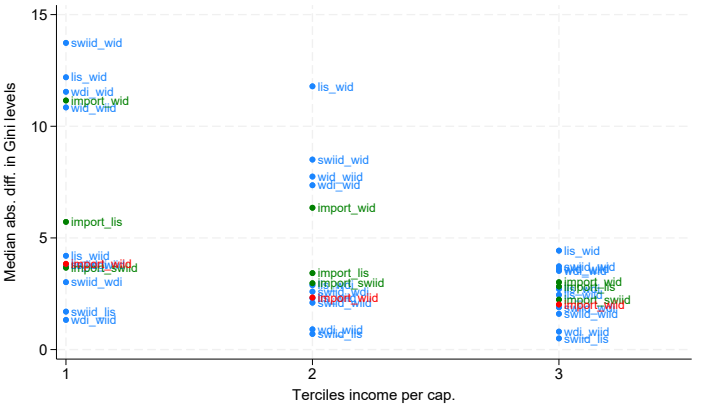
TABLE C.1. Number of selected products by sector in forward and backward selection

Appendix C: Evaluation

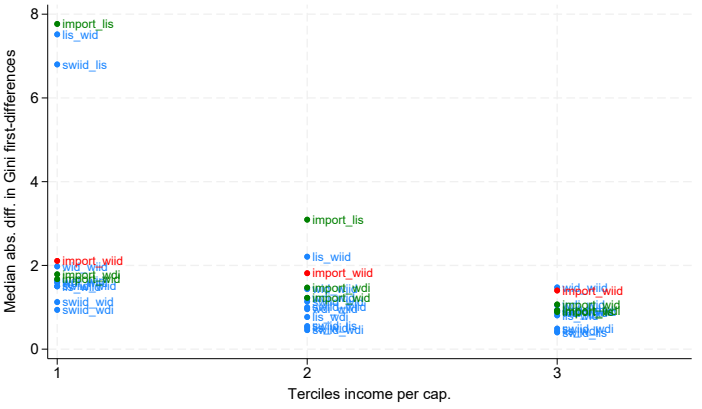
C.1. Backward/forward product selection

C.2. Cross-data comparison

FIGURE C.1. Cross data comparison



(A) Levels



(B) First-differences

Notes: Both graphs depict the median absolute difference between pairs of inequality datasets considered in section 7.1. In red: median absolute difference between our Gini estimates ("import") and WIID (k-fold cross validation) as reported in Table 5. In green: median absolute difference between our Gini estimates and other datasets (in-sample estimation) as reported in Table 6.

References

Aguiar, M. and M. Bils (2015). Has consumption inequality mirrored income inequality? *The American Economic Review 105*(9), 2725–2756.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016). Ethnic inequality. *American Economic Review 102*(2), 428–488.

Almås, I. (2012). International income inequality: Measuring PPP bias by estimating Engel curves for food. *American Economic Review 102*(2), 1093–1117.

Artuc, E., G. Porto, and B. Rijkers (2019). Trading off the income gains and the inequality costs of trade policy. *Journal of International Economics 120*, 1–45.

Bandourian, R., J. McDonald, and R. Turley (2002). A comparison of parametric models of income distribution across countries and over time. LIS Working paper 305.

Banerjee, A. V. and E. Duflo (2007). The economic lives of the poor. *Journal of economic perspectives 21*(1), 141–168.

Banks, J., R. Blundell, and A. Lewbel (1997). Quadratic engel curves and consumer demand. *Review of Economics and Statistics 79*(4), 527–539.

Besley, T. and T. Persson (2014). Why do developing countries tax so little? *Journal of Economic Perspectives 28*(4), 99–120.

Blumenstock, J., G. Cadamuro, and R. On (2015). Predicting poverty and wealth from mobile phone metadata. *Science 350*(6264), 1073–1076.

Borusyak, K. and X. Jaravel (2021). The distributional effects of trade: theory and evidence from the united states. NBER Working Paper 28957.

Caron, J., T. Fally, and J. R. Markusen (2014). International trade puzzles: a solution linking production and preferences. *The Quarterly Journal of Economics 129*(3), 1501–1552.

Chatterjee, A., L. Czajka, and A. Gethin (2023). Redistribution without inclusion? inequality in south africa since the end of apartheid. WID Working paper.

Choi, Y. C., D. Hummels, and C. Xiang (2009). Explaining import quality: The role of the income distribution. *Journal of International Economics 77*, 265–275.

Comin, D., D. Lashkari, and M. Mestieri (2021). Structural change with long-run income and price effects. *Econometrica 89*(1), 311–374.

Dagum, C. (1977). A new model of personal income distribution: specification and estimation. *Economie Appliquée 30*, 413–437.

Dagum, C. (2008). A new model of personal income distribution: specification and estimation. In *Modeling income distributions and Lorenz curves*, pp. 3–25. Springer.

Deininger, K. and L. Squire (1996). A new data set measuring income inequality. *The World Bank Economic Review 10*(3), 565–591.

Egger, P. and M. Larch (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics 76*(2), 384–399.

Faber, B. and T. Fally (2017). Firm heterogeneity in consumption baskets: Evidence from home and store scanner data. Technical report, National Bureau of Economic Research.

Fajgelbaum, P. D., G. M. Grossman, and E. Helpman (2011). Income distribution, product quality, and international trade. *Journal of Political Economy 119*(4), 721–765.

Fajgelbaum, P. D. and A. K. Khandelwal (2016). Measuring unequal gains from trade. *Quarterly Journal of Economics 131*(3), 1113–1180.

Feenstra, R. C. and J. Romalis (2014). International prices and endogenous quality. *Quarterly Journal of Economics 129*(2), 477–527.

Ferreira, F. H., N. Lustig, and D. Teles (2015). *Appraising cross-national income inequality databases: An introduction*. The World Bank.

Fieler, A. C. (2011). Nonhomotheticity and bilateral trade: Evidence and a quantitaive explanation. *Econometrica 79*(4), 1069–1101.

Galbraith, J. K., B. Halbach, A. Malinowska, A. Shams, and W. Zhang (2014). Utip global inequality data sets 1963-2008: updates, revisions and quality checks.

Galbraith, J. K. and H. Kum (2005). Estimating the inequality of household incomes: a statistical approach to the creation of a dense and consistent global data set. *Review of Income and Wealth 51*(1), 115–143.

Gaulier, G. and S. Zignago (2010). Baci: international trade database at the product-level (the 1994-2007 version).

Hallak, J. C. (2006). Product quality and the direction of trade. *Journal of International Economics 68*, 238–265.

Hallak, J. C. (2010). A product-quality view of the linder hypothesis. *The Review of Economics and Statistics 92*(3), 453–466.

Handbury, J. (2019). Are poor cities cheap for everyone? non-homotheticity and the cost of living across U.S. cities. mimeo.

Henderson, V., A. Storeygard, and D. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Hillrichs, D. and G. Vannoorenberghe (2022). Trade costs, home bias and the unequal gains from trade. *Journal of International Ecoomics 139*, 103884.

Horowitz, J. L. (2019). Bootstrap methods in econometrics. *Annual Review of Economics 11*, 193–224.

Ianchovichina, E., L. Mottaghi, and S. Devarajan (2015). Inequality, uprisings, and conflict in the arab world. *MENA Economic Monitor*.

Jaimovich, E. and V. Merella (2015). Love for quality, comparative advantage and trade. *Journal of International Economics 97*(2), 376–391.

James, G., D. Witten, T. Hastie, R. Tibshirani, et al. (2021). *An Introduction to Statistical Learning*. Springer.

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

Kleiber, C. (2008). A guide to the dagum distributions. In *Modeling income distributions and Lorenz curves*, pp. 97–117. Springer.

Kongsamut, P., S. Rebelo, and D. Xie (2001). Beyond balanced growth. *The Review of Economic Studies 68*(4), 869–882.

Lessmann, C. and A. Seidel (2017). Regional inequality, convergence, and its determinants–a view from outer space. *European Economic Review 92*, 110–132.

Linder, S. B. (1961). *An Essay on Trade and Transformation*. Ph. D. thesis, Uppsala.

Matsuyama, K. (2019). Engel's law in the global economy: Demand-induced patterns of structural change, innovation, and trade. *Econometrica 87*(2), 497–528.

McGregor, T., B. Smith, and S. Wills (2019, 07). Measuring inequality. *Oxford Review of Economic Policy 35*(3), 368–395.

Nigai, S. (2016). On measuring the welfare gains from trade under consumer heterogeneity. *Economic Journal 126*, 1193–1237.

Porto, G. (2006). Using survey data to assess the distributional effects of trade policy. *Journal of International Economics 70*(1), 140–160.

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

Simonovska, I. (2015). Income differences and prices of tradables: Insights from an online retailer. *The Review of Economic Studies 82*(4), 1612–1656.

Solt, F. (2009). Standardizing the world income inequality database. *Social Science Quarterly 90*(2), 231–242.

Solt, F. (2019, Feb). Measuring income inequality across countries and over time: The standardized world income inequality database. OSF. https://osf.io/3djtq. SWIID Version 8.0, February 2019.