Journal of International Economics 108 (2017) 1–19

Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

From micro to macro: Demand, supply, and heterogeneity in the trade elasticity☆

Maria Bas[a], Thierry Mayer[b,*], Mathias Thoenig[c]

[a] University Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne (CES), France
[b] Sciences Po, Banque de France, CEPII and CEPR 28, rue des Saints-Peres, 75007 Paris, France
[c] Faculty of Business and Economics, University of Lausanne and CEPR, Switzerland

ARTICLE INFO

Article history:
Received 19 February 2016
Received in revised form 5 May 2017
Accepted 5 May 2017
Available online xxxx

JEL classification:
F1

Keywords:
Trade elasticity
Firm-level data
Heterogeneity
Gravity
Pareto
Log-normal

ABSTRACT

Models of heterogeneous firms with selection into export market participation generically exhibit aggregate trade elasticities that vary across country-pairs. Only when heterogeneity is assumed Pareto-distributed do all elasticities collapse into a unique elasticity, estimable with a gravity equation. This paper provides a theory-consistent methodology for quantifying country-pair specific aggregate elasticities when moving away from Pareto, i.e. when gravity does not hold. Combining two firm-level customs datasets for which we observe French and Chinese individual sales on the same destination market over the 2000–2006 period, we are able to estimate all the components of the bilateral aggregate elasticity: i) the demand-side parameter that governs the intensive margin and ii) the supply side parameters that drive the extensive margin. These components are then used to calculate theoretical predictions of bilateral aggregate elasticities over the whole set of destinations, and how those elasticities decompose into different margins. Our predictions fit well with econometric estimates, supporting our view that micro-data is a key element in the quantification of aggregate trade elasticities.

1. Introduction

The response of trade flows to a change in trade costs, the aggregate trade elasticity, is a central element in any evaluation of the welfare impacts of trade liberalization. Arkolakis et al. (2012) recently showed that this parameter, let us call it ε for the rest of the paper, is actually one of the (only) two sufficient statistics needed to calculate Gains From Trade (GFT) under a surprisingly large set of alternative modeling assumptions. Measuring those elasticities has therefore been the topic of a long-standing literature in international economics. The most common practice (and the one recommended by Arkolakis et al., 2012) is to estimate this elasticity in a *macro-level* bilateral trade equation referred to as structural gravity in the literature following the initial impulse by Anderson and van Wincoop (2003). In order for this estimate of ε to be relevant for a particular experiment of trade liberalization, it is crucial for this bilateral trade equation to be correctly specified as a structural gravity model with, in particular, a *unique* elasticity to be estimated across country pairs.

Our starting point is that the model of heterogeneous firms with selection into export market participation (Melitz, 2003) will in general exhibit a *bilateral-specific* aggregate trade elasticity, i.e. an ε_{ni}, which applies to each country pair, where i denotes the origin and n the destination of the flow. Only when heterogeneity is assumed Pareto-distributed[1] do all ε_{ni} collapse to a single ε. Under any other (commonly-used) distributional assumption, obtaining an estimate of the aggregate trade elasticity from a macro-level bilateral trade

☆ This research has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013) Grant Agreement No. 313522. We thank the Editor and two anonymous referees for very useful remarks. The paper also benefited greatly at an early stage from comments by David Atkin, Dave Donaldson, Swati Dhingra, Ben Faber, Pablo Fagelbaum, Jean Imbs, Oleg Itskhoki, Jan de Loecker, Philippe Martin, Peter Morrow, Steve Redding, Andres Rodriguez Clare, Esteban Rossi-Hansberg, Katheryn Russ, Nico Voigtlander, and Yoto Yotov, during presentations in UC Berkeley, UCLA, Penn State, Banque de France, WTO, CEPII, ISGEP in Stockholm, University of Nottingham, Banca d'Italia, Princeton University and CEMFI.

* Corresponding author.
E-mail address: thierry.mayer@sciencespo.fr (T. Mayer).

[1] Unless otherwise specified, Pareto is understood here as the *unbounded version* used by most of the literature. See Helpman et al. (2008) and Melitz and Redding (2015) for results with the bounded version, where the trade elasticity recovers a bilateral dimension.

http://dx.doi.org/10.1016/j.jinteco.2017.05.001
0022-1996/

equation becomes problematic: first because a whole set of ε_{ni} has to be estimated, and second because structural gravity does not hold anymore. We argue that in this case quantifying trade elasticities at the aggregate level makes it necessary to use micro-level information. To this purpose, we combine sales of French and Chinese exporters in many destination-product combinations for which we also observe the relevant tariff applied. We propose a theory-consistent methodology using this firm-level export data for quantifying all the components of the bilateral aggregate trade elasticity: i) the demand-side parameter that governs the intensive margin and ii) the supply side parameters that drive the extensive margin. These components are then assembled under theoretical guidance to calculate the bilateral aggregate elasticities over the whole set of destinations.

Taking into account country pair heterogeneity in aggregate trade elasticities is crucial for quantifying the expected impact of various trade policy experiments.[2] Consider the example of envisioned Transatlantic or Transpacific trade agreements (TTIP or TPP). Under the simplifying assumption of a unique elasticity, whether the trade liberalization takes place with a proximate vs distant, large vs small economy, is irrelevant in terms of trade-promoting effect or welfare gains calculations. By contrast, our results suggest that the relevant ε_{ni} should be smaller (in absolute value) when trade liberalization concerns country-pairs where the volume of bilateral trade is already large. Regarding welfare, Head et al. (2014) and Melitz and Redding (2015) have shown theoretically that the GFT can be substantially mis-estimated if one assumes a constant trade elasticity when the "true" elasticity is variable (the margin of error can exceed 100% in both papers). The expected changes in trade patterns and welfare effects of agreements such as TTIP or TPP will therefore be different compared to the unique elasticity case. One of the main objectives of our paper is to quantify how wrong can one be when making predictions based on a constant trade elasticity assumption. Naturally, this point also applies to the case of potential breakups of existing agreements such as the EU or NAFTA.

Our approach maintains the traditional CES (σ) demand system combined with monopolistic competition. It features several steps that are structured around the following decomposition of the aggregate trade elasticity into the sum of the intensive margin and the (weighted) extensive margin:

$$\varepsilon_{ni} = \underbrace{1-\sigma}_{\text{intensive margin}} + \underbrace{\frac{1}{\bar{x}_{ni}/x_{ni}^{\mathtt{MIN}}}}_{\text{mean-to-min}} \times \underbrace{\frac{d\ln N_{ni}}{d\ln \tau_{ni}}}_{\text{extensive margin}} . \tag{1}$$

The weight is the inverse of the *mean-to-min ratio*, our observable measuring the dispersion of firm-level performance, that is defined as the ratio of average to minimum sales across markets. As the market gets easier, the model predicts a larger presence of weak firms, which augments productivity dispersion, captured by $\bar{x}_{ni}/x_{ni}^{\mathtt{MIN}}$. This lowers the weight of the extensive margin in the overall trade elasticity, which is intuitive: in extremely easy markets, all potential exporters should be active and the extensive margin of a small change in trade costs should be close to 0. When assuming Pareto with shape parameter θ, the last part of the elasticity reduces to $\sigma - 1 - \theta$, and the overall elasticity becomes constant and reflects only the parameter controlling dispersion in the distribution of productivity: $\varepsilon_{ni}^{P} = \varepsilon^{P} = -\theta$ (Chaney, 2008). Without the Pareto assumption, one needs to calculate the two components of the aggregate elasticity (Eq. (1)). We do so in two steps.

Our first step aims to estimate the demand side parameter σ using firm-level exports. Since protection is imposed on all firms from a given origin, higher demand and lower protection are not separately identifiable when using only one exporting country. With CES, firms are all faced with the same aggregate demand conditions. Thus, considering a second country of origin enables to isolate the effects of trade policy, if the latter is discriminatory. We therefore combine shipments by French and Chinese exporters to destinations that confront those firms with different levels of tariffs. Our setup yields a *firm-level* gravity equation which raises serious estimation challenges. The main issue is the combination of a selection bias (inherent in any firm-level estimation of the Melitz (2003) model) with a very large set of fixed effects to be included in the regression. We use adapted versions of three estimators that have been proposed in the literature to deal with different aspects of the problem. Those three methods are evaluated with Monte Carlo simulations of our theoretical setup, before being implemented on our data. Our preferred estimates of the firm-level trade elasticity imply an average value of $(1-\hat{\sigma})$ around -4.

Our second and main step applies Eq. (1) and combines the estimate of the firm-level elasticity $(1-\hat{\sigma})$ with the central supply side parameter—reflecting dispersion in the distribution of productivity—to obtain theoretical predictions of the aggregate elasticities of total export, number of exporters and average exports per firm to each destination. Those predictions (one elasticity for each exporter-importer combination) require knowledge of the bilateral export productivity cutoff under which firms find exports to be unprofitable. We make use of the mean-to-min ratio to reveal those cutoffs. A key element of our procedure is the calibration of the productivity distribution. As an alternative to Pareto we consider the log-normal distribution that fits the micro-data on firm-level sales very well.[3]

A related contribution of our paper is to discriminate between Pareto and log-normal as potential distributions for the underlying firm-level heterogeneity, suggesting that log-normal does a better job at matching the non-unique response of exports to changes in trade costs. Two pieces of evidence in that direction are provided. The first provides direct evidence that aggregate trade elasticities are non-constant across country pairs. The second is a strong correlation across industries between firm-level and aggregate elasticities—at odds with the prediction of a null correlation under Pareto. We also find that the heterogeneity in trade elasticities is quantitatively important: Although the average of bilateral elasticities is quite well approximated by a standard gravity model constraining the estimated parameter to be constant, deviations from this average level can be large. We show that under log-normal the ε_{ni} are larger (in absolute value) for pairs with low volumes of trade. Hence the trade-promoting impact of liberalization is expected to be larger for this kind of trade partners. For Chinese exports, assuming a unique elasticity would underestimate the trade impact of a tariff liberalization by about 25% for countries with initially very small trade flows (Somalia, Chad or Azerbaijan for instance). By contrast, the error would be to overestimate by around 20% the exports created when the United States or Japan reduce their trade costs.

The next section relates our paper to the existing literature. Section 3 describes our model and empirical strategy. Section 4 deals with the estimation challenges of the firm-level gravity regressions and reports the estimates of the intensive margin elasticity.

[2] Imbs and Méjean (2015) and Ossa (2015) recently argued that another source of heterogeneity, the cross-sectoral one, raises important aggregation issues that matter for aggregate outcomes of trade liberalization. We abstract from this particular kind of aggregation issue (which would reinforce the importance of heterogeneity for aggregate outcomes) in our paper and omit cross-sectoral variation in ε until Section 6 where we present industry-level estimates and use those to show that both demand and supply side determinants enter aggregate elasticities.

[3] Head et al. (2014) provide evidence and references for several micro-level data sets that individual sales are much better approximated by a log-normal distribution when the entire distribution is considered (without left-tail truncation). Freund and Pierola (2015) is a recent example showing, for all of the 32 countries used, very large deviations from Pareto if the data is not vastly truncated to focus on the very largest firms.

Section 5 computes micro-based theoretical predictions of the bilateral aggregate elasticities and compares them to their gravity estimates obtained with Chinese and French aggregate export data. Section 6 investigates the implications of cross-industry heterogeneity for our analysis and provides an additional piece of evidence in favor of non-constant aggregate trade elasticities. The final section concludes.

2. Related literature

In the empirical literature estimating trade elasticities, different approaches and proxies for trade costs have been used, with an almost exclusive focus on aggregate country or industry-level data. The gravity approach to estimating those elasticities mostly uses tariff data to estimate bilateral responses to variation in applied tariff levels. Most of the time, identification is based on the cross-section of country pairs, with origin and destination determinants being controlled through fixed effects (Hummels, 1999; Baier and Bergstrand, 2001; Head and Ries, 2001; Romalis, 2007; Caliendo and Parro, 2015 for instance). A related approach consists in using the fact that most foundations of gravity predict the same coefficient on trade costs and domestic cost shifters to estimate that elasticity from the effect on bilateral trade of exporter-specific changes in productivity, export prices or exchange rates. Costinot et al. (2012) use industry-level data for OECD countries, and obtains a preferred elasticity of −6.53 relying on producer prices of the exporter as the identifying variable.[4] Our paper has consequences for how to interpret those numbers in terms of underlying structural parameters. With a homogeneous firms model of the Krugman (1980) type in mind, the estimated trade elasticity turns out to reveal a demand-side parameter only, $1-\sigma$ (this is also the case with Armington differentiation and perfect competition as in Anderson and van Wincoop, 2003). When instead considering heterogeneous firms à la Melitz (2003), the literature has proposed that the aggregate trade elasticity is driven solely by a supply-side parameter describing the dispersion of the underlying distribution of firm productivity. This result has been shown with several demand systems (CES by Chaney (2008), linear by Melitz and Ottaviano (2008), translog by Arkolakis et al. (2010) for instance), but relies critically on the maintained assumption of a Pareto distribution. The trade elasticity then provides an estimate of the dispersion parameter of the Pareto distribution for firm productivity, θ.[5] We show here that both existing interpretations of the estimated elasticities are too extreme: When the Pareto assumption is relaxed, the aggregate trade elasticity is a mix of demand and supply parameters.

A small set of papers estimate the intensive margin elasticity at the exporter level. Berman et al. (2012) present estimates of the trade elasticity with respect to real exchange rate variations across countries and over time using firm-level data from France. Fitzgerald and Haller (2015) use firm-level data from Ireland, real exchange rate and weighted average firm-level applied tariffs as price shifters to estimate the trade elasticity. The results for the impact of real exchange rate on firms' export sales are of a similar magnitude, around 0.8 to 1. Regarding tariffs, Fitzgerald and Haller (2015) construct a firm-level destination-year tariff as the weighted average of the applied tariffs at the product-destination-year level imposed on the firm's products, using as weights the share of a product in firm total production. They find a tariff elasticity ranging widely from −1.7 to −24 in their baseline table. The preferred estimate of Berthou and Fontagné (2016), who use the response of the largest French exporters in the United States to the levels of applied tariffs is −2.5. We depart from those papers by using an alternative methodology to identify the trade elasticity with respect to applied tariffs; i.e. the differential treatment of exporters from two distinct countries (France and China) in a set of product-destination markets. We also describe thoroughly the estimation challenges involved in firm-level gravity regressions and provide the first rigorous evaluation of the alternative estimators available with Monte Carlo simulations using the canonical Melitz (2003) model as a Data Generating Process (DGP).

Our paper also relates to several recent papers studying patterns and consequences of heterogeneity in trade elasticities. Berman et al. (2012) and Gopinath and Neiman (2014) find that in order to predict correctly the aggregate patterns of trade adjustments to price shocks, one has to take into account firm-level heterogeneity with the use of micro-data. In both papers, heterogeneity matters because firms have different individual responses in export and/or import behavior. In particular, both papers find that the firm-level elasticity depends negatively on the size of the firm (because of variable markups). Our paper also finds that measuring aggregate trade responses requires usage of firm-level data. It is however for a different reason: In our case, heterogeneity in aggregate trade elasticities simply originates in a departure from the common assumption that productive efficiency is Pareto-distributed.[6] While we do recognize that trade elasticities might differ across firms because of variable markups, our paper shows that this is not required to ensure that heterogeneity matters for the aggregate economy and investigates a different, complementary, channel.[7]

We also contribute to the literature studying the importance of the distributional assumption for firm heterogeneity for trade patterns, trade elasticities and welfare. Head et al. (2014), Yang (2014), Melitz and Redding (2015) and Feenstra (2013) have recently argued that the simple gains from trade formula proposed by Arkolakis et al. (2012) rely crucially on the Pareto assumption, which mutes important channels of gains in the heterogenous firms case. Barba Navaretti et al. (2015) present gravity-based evidence that the exporting country fixed effects depends on characteristics of firms' distribution that go beyond the simple mean productivity, a feature incompatible with the usually specified Pareto heterogeneity. Fernandes et al. (2015) use customs data for numerous developing countries to show that a decomposition of total bilateral exports into intensive and extensive margins exhibits an important role for the former, with patterns consistent with log-normally distributed heterogeneity and incompatible with (unbounded) Pareto. The alternatives to Pareto considered to date in welfare gains quantification

[4] Other methodologies (also used for aggregate elasticities) use identification via heteroskedasticity in bilateral flows, and have been developed by Feenstra (1994) and applied widely by Broda and Weinstein (2006) and Imbs and Méjean (2015). Yet, another alternative is to proxy trade costs using retail price gaps and their impact on trade volumes, as proposed by Eaton and Kortum (2002) and extended by Simonovska and Waugh (2011).

[5] This result of a constant trade elasticity reflecting the Pareto shape holds when maintaining the CES demand system but making other improvements to the model such as heterogeneous marketing and/or fixed export costs (Arkolakis, 2010; Eaton et al., 2011). In the Ricardian setup of Eaton and Kortum (2002), the trade elasticity is also a (constant) supply side parameter reflecting heterogeneity, but this heterogeneity takes place at the national level, and reflects the scope for comparative advantage.

[6] Yet another alternative source of bilateral heterogeneity in the trade elasticity could be a composition effect, coming from aggregating products with different underlying elasticities, or comparing pairs with different country characteristics. Our empirical analysis showing heterogeneous elasticities at the aggregate level is based on a ratio approach that conditions on the two exporting countries having the same set of destination-product combinations.

[7] A further interesting result is that heterogeneous firm-level elasticities do not guarantee variable bilateral aggregate trade elasticities. Melitz and Ottaviano (2008) and Berman et al. (2012) are two examples of models with variable markups that yield heterogeneity in firm-level response to trade costs. However, in both cases, when productivity is assumed Pareto-distributed, the bilateral aggregate trade elasticity turns out to be a constant only related to the Pareto shape parameter. Introducing variable markups in the Pareto context therefore is not sufficient to generate the data patterns we uncover here.

exercises are i) the bounded Pareto by Helpman et al. (2008), Melitz and Redding (2015) and Feenstra (2013), ii) the log-normal by Head et al. (2014), Fernandes et al. (2015) and Yang (2014), and iii) a mixture of Pareto and log-normal in Nigai (2017). A key simplifying feature of Pareto is to yield a constant trade elasticity, which is not the case for alternative distributions. Helpman et al. (2008), Novy (2013) and Spearot (2013) have produced empirical evidence showing substantial variation in the trade cost elasticity across country pairs. Our contribution to that literature is to use the estimated demand and supply-side parameters to construct predicted bilateral elasticities for aggregate flows under the log-normal assumption, and compare their first moments to gravity-based estimates. It should be noted that there are other ways to generate bilateral trade elasticities. The most obvious is to depart from the simple CES demand system. Novy (2013) builds on Feenstra (2003), using the translog demand system with homogeneous firms to obtain variable trade elasticities. Spearot (2013) obtains country-pair specific trade elasticities motivated by the Melitz and Ottaviano (2008) model, which combines firm heterogeneity with a linear demand system. Atkeson and Burstein (2008) maintain CES demand, generating heterogeneity in elasticities through oligopoly. We choose here to keep the change with respect to the benchmark Melitz/Chaney framework to a minimal extent, keeping CES and monopolistic competition, while changing only the distributional assumption, comparing Pareto to log-normal.[8]

3. Firm-level and aggregate-level trade elasticities: theory

We use the multi-country one-sector version of the Melitz (2003) theoretical framework. Country i hosts a set of heterogeneous firms facing a constant price elasticity (CES utility combined with iceberg costs) and contemplating exports to several destinations indexed by subscript n. In this setup, firm-level export value x depends upon the firm-specific unit input requirement (α), wages at home (w_i), and real expenditure in n, $A_n \equiv X_n P_n^{\sigma-1}$, with P_n the ideal CES price index relevant for sales in n. A_n is a measure of "attractiveness" of market n (expenditure discounted by the degree of competition in this market). There are trade costs associated with reaching market n, consisting of an observable iceberg-type part (τ_{ni}), and a shock that affects firms differently on each market, $b_{ni}(\alpha)$.[9] Monopolistic competition ensures a complete pass-through of trade costs into delivered prices, such that firm-level sales are

$$x_{ni}(\alpha) = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} [\alpha w_i \tau_{ni} b_{ni}(\alpha)]^{1-\sigma} A_n. \tag{2}$$

The *firm-level trade elasticity*, i.e. the individual reaction of x_{ni} to a change in observable trade costs, is $1-\sigma$.

In order to obtain the aggregate trade elasticity, we start by summing, for each country pair, the sales Eq. (2) across all active firms:

$$X_{ni} = V_{ni} \times \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} (w_i \tau_{ni})^{1-\sigma} A_n M_i^e, \tag{3}$$

where M_i^e is the mass of entrants and V_{ni} is a term which denotes a cost-performance index of exporters located in country i and selling in n. This index, introduced by Helpman et al. (2008), is defined as

$$V_{ni} \equiv \int_0^{a_{ni}^*} a^{1-\sigma} g(a) da, \tag{4}$$

where $a \equiv \alpha \times b(\alpha)$ corresponds to the unitary labor requirement rescaled by the firm-destination shock and $g(.)$ denotes its PDF (with a corresponding CDF denoted by $G(.)$). In Eq. (4), a_{ni}^* is the rescaled labor requirement of the firm that just breaks even and therefore exports to market n. The solution for this cutoff firm is the cost satisfying the zero profit condition, i.e., $x_{ni}(a_{ni}^*) = \sigma w_i f_n$, where f_n is the fixed export cost in each destination n. Using Eq. (2), this cutoff is characterized by

$$a_{ni}^* = \frac{\sigma-1}{\sigma} \frac{1}{\tau_{ni} f_n^{1/(\sigma-1)}} \left(\frac{A_n}{\sigma w_i^\sigma}\right)^{\frac{1}{\sigma-1}}. \tag{5}$$

We are interested in the (partial) elasticity of aggregate trade value with-respect to variable trade costs, τ_{ni}. Partial means here holding constant origin-specific and destination-specific terms (income and price indices) as in Arkolakis et al. (2012) and Melitz and Redding (2015).[10] Using Eq. (3), we obtain the *bilateral aggregate trade elasticity*:

$$\varepsilon_{ni} \equiv \frac{d \ln X_{ni}}{d \ln \tau_{ni}} = 1 - \sigma - \gamma_{ni}, \tag{6}$$

which uses the fact that $d \ln a_{ni}^* / d \ln \tau_{ni} = -1$. The γ_{ni} term, introduced by Arkolakis et al. (2012), describes how V_{ni} varies with an increase in the cutoff cost a_{ni}^*, that is an easier access of market n for firms in i:

$$\gamma_{ni} \equiv \frac{d \ln V_{ni}}{d \ln a_{ni}^*} = \frac{a_{ni}^{*2-\sigma} g\left(a_{ni}^*\right)}{V_{ni}}. \tag{7}$$

Eqs. (6) and (7) show that the aggregate trade elasticity should, in general, not be constant across country pairs. They also make it clear that the aggregate elasticity is a combination of the firm-level trade elasticity, $1-\sigma$, and the contribution to total export changes due to entry and exit of firms into the export market, γ_{ni}.

In order to evaluate ε_{ni}, combining Eq. (7) with Eq. (4) reveals that we need to know the value of bilateral cutoffs a_{ni}^*. In order to obtain those, we define the following expression

$$\mathcal{H}(a_{ni}^*) \equiv \frac{1}{a_{ni}^{*1-\sigma}} \int_0^{a_{ni}^*} a^{1-\sigma} \frac{g(a)}{G\left(a_{ni}^*\right)} da = \frac{V_{ni}}{a_{ni}^{*1-\sigma} G\left(a_{ni}^*\right)}, \tag{8}$$

a monotonic, invertible function with a parametrization that is tightly linked to the distributional assumptions retained for $G(.)$ (see Eq. (20) in Section 5.1). In this model, $\mathcal{H}(.)$ has a straightforward economic interpretation. It is the ratio of average over minimum performance (measured as $a^{*1-\sigma}$) of firms located in i and exporting to n. Using Eqs. (2) and (3) reveals that this ratio also corresponds to the observed mean-to-min ratio of sales:

$$\mathcal{H}(a_{ni}^*) = \frac{\bar{x}_{ni}}{x_{ni}\left(a_{ni}^*\right)} = \frac{\bar{x}_{ni}}{x_{ni}^{\mathrm{MIN}}}. \tag{9}$$

[8] Relying on the same approach (CES and monopolistic competition), Helpman et al. (2008) assume bounded Pareto to obtain bilateral trade elasticities that vary across country pairs. They estimate the distance trade elasticity at the aggregate level, while in this paper, we estimate directly the price elasticity using tariff data and we use firm-level information.

[9] An example of such unobservable term would be the presence of workers from country n in firm α, that would increase the internal knowledge on how to reach consumers in n, and therefore reduce trade costs for that specific company in that particular market (b being a mnemonic for barrier to trade). Note that this type of random trade cost shock is isomorphic to assuming a firm-destination demand shock in this CES-monopolistic competition model.

[10] In practical terms, the use of importer and exporter fixed effects in gravity regressions (the main source of estimates of the aggregate elasticity) holds w_i, M_i^e and A_n constant when estimating Eq. (3).

In firm-level export data sets, the ratio of average to minimum trade flows by firm in each destination country n is an observable. Using Eq. (9), one can reveal $\hat{a}^*_{ni}$, the predicted value of the export cutoff for i firms exporting to n as a function of the mean-to-min ratio of sales in each destination market:

$$\hat{a}^*_{ni} = \mathcal{H}^{-1}\left(\frac{\bar{x}_{ni}}{x^{\mathrm{MIN}}_{ni}}\right). \tag{10}$$

Equipped with the bilateral cutoff, we use Eqs. (6) to (9) to quantify the bilateral aggregate trade elasticity

$$\varepsilon_{ni} = 1 - \hat{\sigma} - \frac{x^{\mathrm{MIN}}_{ni}}{\bar{x}_{n,i}} \times \frac{\hat{a}^*_{ni} g\left(\hat{a}^*_{ni}\right)}{G\left(\hat{a}^*_{ni}\right)}, \tag{11}$$

where $(1 - \hat{\sigma})$ is obtained from the firm-level export equation (see Section 4). We also calculate two other aggregate elasticities: the elasticity of the number of exporters N_{ni} (the so-called extensive margin) and the elasticity of average exports per firm $\bar{x}_{ni}$. The number of active exporters is closely related to the cutoff since $N_{ni} = M^e_i \times G\left(a^*_{ni}\right)$, where M^e_i represents the mass of entrants. Differentiating and using Eq. (11) we can calculate the bilateral extensive margin of trade

$$\frac{d\ln N_{ni}}{d\ln \tau_{ni}} = -\frac{\hat{a}^*_{ni} g\left(\hat{a}^*_{ni}\right)}{G\left(\hat{a}^*_{ni}\right)}. \tag{12}$$

From the accounting identity $X_{ni} \equiv N_{ni} \times \bar{x}_{ni}$, we obtain the (partial) elasticity of average exports per firm to trade simply as the difference between the predicted aggregate elasticity, Eq. (11) and the predicted extensive margins, Eq. (12):

$$\frac{d\ln \bar{x}_{ni}}{d\ln \tau_{ni}} = \varepsilon_{ni} - \frac{d\ln N_{ni}}{d\ln \tau_{ni}} = 1 - \hat{\sigma} - \frac{\hat{a}^*_{ni} g\left(\hat{a}^*_{ni}\right)}{G\left(\hat{a}^*_{ni}\right)}\left(\frac{x^{\mathrm{MIN}}_{ni}}{\bar{x}_{n,i}} - 1\right). \tag{13}$$

Combining Eqs. (11) and (12), we can re-express aggregate elasticities as a function of the intensive and extensive margins and of the mean-to-min ratio:

$$\varepsilon_{ni} = \underbrace{1 - \hat{\sigma}}_{\text{intensive margin}} + \overbrace{\underbrace{\frac{1}{\bar{x}_{ni}/x^{\mathrm{MIN}}_{ni}}}_{\text{mean-to-min}} \times \underbrace{\frac{d\ln N_{ni}}{d\ln \tau_{ni}}}_{\text{extensive margin}}}^{\text{weighted extensive margin}}, \tag{14}$$

which is Eq. (1) presented in the Introduction 1. This decomposition shows that the aggregate trade elasticity is the sum of the intensive margin and of the (weighted) extensive margin. The weight on the extensive margin depends only on the mean-to-min ratio, an observable measuring the dispersion of relative firm performance ($\mathcal{H}(a^*_{ni})$ in the model). Intuitively, the weight of the extensive margin should be decreasing when the market gets easier. Indeed easy markets have higher rates of entry, $G(a^*)$, and therefore increasing presence of weaker firms which augments dispersion measured as $\mathcal{H}\left(a^*_{ni}\right)$. The marginal entrant in an easy market will therefore have less influence on aggregate exports, a smaller impact of the extensive margin. In the limit, the weight of the extensive margin becomes negligible and the whole of the aggregate elasticity is due to the intensive margin/demand parameter. In the (unbounded) Pareto case however, this mechanism is not operational since $\mathcal{H}\left(a^*_{ni}\right)$ and therefore the weight of the extensive margin is constant. In Section 5, we implement our method with both Pareto-distributed a as well as with log-normally-distributed a (which yields a varying dispersion of sales across destinations).

There are two major elements needed for the practical implementation of Eq. (14). First, we need to obtain an estimate of $1 - \hat{\sigma}$, the parameter relevant in the firm-level trade elasticity. This is the topic of Section 4. Second, we need to measure the mean-to-min ratio, $\mathcal{H}(a^*_{ni})$, in the weighted extensive margin that add to the firm-level elasticity to yield the response of total trade to a change in trade costs. This is done in Section 5. Our framework until now has been silent about the product/sector dimension. However, our data (export values and tariffs notably) come with product information, and it is possible that different products (we index those with p) are characterized by different values of the demand elasticity σ and/or of the dispersion of firm performance, $\mathcal{H}\left(a^*_{ni}\right)$. The composition effects coming from such dispersion in sectoral characteristics has been well documented in the recent work by Imbs and Méjean (2015) and Ossa (2015) for instance. We want to first present results that abstract from this dimension and focus on the new source of bilateral heterogeneity we propose: the one coming "purely" from distributional assumptions. In Sections 4 and 5, we therefore abstract from cross-sectoral heterogeneity in σ and in $\mathcal{H}\left(a^*_{ni}\right)$. Those should be understood as averages of the underlying sectoral values, which we obtain by pooling over sectors. This offers the advantage of compactness in the presentation of results. In Section 6, we return to that sectoral issue and let all structural parameters and observables take a different value across sectors in our computation of firm-level and aggregate elasticities.

4. Estimation of the firm-level trade elasticity

4.1. Estimation challenges

Three serious methodological challenges arise when estimating the firm-level response of export values to variation in tariffs while keeping a close link to theory.

4.1.1. The need for multiple origins

The first challenge is to separate the effect of trade costs from destination fixed effects. At this stage, it is useful to account for the product (p) dimension for which we observe both the value exported by the firm, $x^p_{ni}(\alpha)$, and the bilateral tariff rate $t^p_{ni}(\alpha)$. Trade costs include both tariffs and other trade costs (distance D_{ni} for instance), and we assume the standard functional form such that $\tau^p_{ni} = \left(1 + t^p_{ni}\right) D^{\delta}_{ni}$. From now on, we will use the term "market" to designate a product-destination combination. Taking logs of the demand Eq. (2), where $\epsilon^p_{ni}(\alpha) \equiv \left(b^p_{ni}(\alpha)\right)^{1-\sigma}$ is our unobservable firm-market error term, a "firm-level gravity" equation is obtained:

$$\begin{aligned}\ln x^p_{ni}(\alpha) = &(1-\sigma)\ln\left(\frac{\sigma}{\sigma-1}\right) + (1-\sigma)\ln(\alpha w_i) + (1-\sigma)\ln\left(1 + t^p_{ni}\right)\\ &+ (1-\sigma)\delta\ln D_{ni} + \ln A^p_n + \ln \epsilon^p_{ni}(\alpha).\end{aligned} \tag{15}$$

The objective is to estimate $1 - \sigma$ out of the impact of tariffs on firm-level sales. At this stage of the paper, as discussed above, we consider a unique σ, which can be interpreted as an average of elasticities that might vary across products. We will come back to industry-specific elasticities in Section 6. In the gravity literature, it has become common practice to capture A^p_n (a complex construction, that depends non-linearly upon σ) with market fixed effects. This is however not applicable if the data set at hand covers only one origin country, since A^p_n and τ^p_n would then vary across the same dimensions.[11] To

[11] Most if not all papers estimating firm-level gravity rely on only one source of export flows, while still estimating the impact of exchange rate (Berman et al., 2012), tariffs (Berthou and Fontagné, 2016) or both (Fitzgerald and Haller, 2015). The identification in those papers then comes from another dimension, usually time. However, this strategy requires making the assumption that $X_n P^{\sigma-1}_n$ does not vary over time when τ_n does. This is inconsistent with a theory where trade costs enter the price index. Also the time dimension of variance in tariffs might be problematic since Fitzgerald and Haller (2015) note that the changes in tariffs over time are small relative to the cross-sectional variations.

remain theory-consistent, one therefore needs to use at least two sets of exporters, based in countries that face different levels of tariffs applied by n. We do combine firm-level customs data for France and China ($i=[\text{FR},\text{CN}]$), where the value of export flows is available at the firm–HS6–destination level in each year. We measure bilateral tariffs t^p_{ni} using WITS at the HS6–destination country level in each year. Proxies for D_{ni} include distance, contiguity, colonial linkage and common language all obtained from the CEPII gravity database. Section 4.2 gives more detail about each of those data sources.

4.1.2. The fixed effects curse

The second challenge relates to the number of fixed effects to be estimated. In addition to the market dimension (A^p_n), we need a set of fixed effects at the firm level to capture marginal costs (αw_i) (and more generally all other unobservable firm-level determinants of export performance, such as quality of products exported, managerial capabilities…). Since there are tens of thousands of exporters in each origin country and several hundred thousand destination–product combinations, the Least Square Dummy Variable–brute force-approach is not feasible. There are two alternative implementable solutions that we consider. The first solution is to estimate Eq. (15) directly using the high-dimensional procedure that was developed by labor economists to deal with the very large number of fixed effects implied by employer–employee data. [12]

$$\ln x^p_{ni}(\alpha)=\text{FE}^\alpha_i+\text{FE}^p_n+(1-\sigma)\ln(1+t^p_{ni})+(1-\sigma)\delta\ln D_{ni}+\ln\epsilon^p_{ni}(\alpha). \tag{16}$$

We call this approach two-way fixed effects procedure, 2WFE, since we have two dimensions of unobserved heterogeneity to be controlled for (i.e. firm fixed effects FE^α_i and market fixed effects FE^p_n). The second solution is a ratio-type estimation inspired by Hallak (2006), Romalis (2007), Head et al. (2010), and Caliendo and Parro (2015) that removes observable and unobservable determinants for both firm-level and destination factors. This method uses four individual export flows to calculate ratios of ratios: an approach referred to as Tetrads from now on. Consider a given French firm j and a Chinese firm ℓ exporting to both n and a reference country k. The Independence of Irrelevant Alternatives (IIA) property of the CES demand system allows to manipulate Eq. (2) to write the following Tetrad:

$$\frac{x^p_n(\alpha_{j,\text{FR}})/x^p_k(\alpha_{j,\text{FR}})}{x^p_n(\alpha_{\ell,\text{CN}})/x^p_k(\alpha_{\ell,\text{CN}})}=\left(\frac{\tau^p_{n\text{FR}}/\tau^p_{k\text{FR}}}{\tau^p_{n\text{CN}}/\tau^p_{k\text{CN}}}\right)^{1-\sigma}\times\frac{\epsilon^p_n(\alpha_{j,\text{FR}})/\epsilon^p_k(\alpha_{j,\text{FR}})}{\epsilon^p_n(\alpha_{\ell,\text{CN}})/\epsilon^p_k(\alpha_{\ell,\text{CN}})}. \tag{17}$$

Denoting tetradic terms with a ~ symbol, one can re-write Eq. (17) as an estimable equation

$$\ln\tilde{x}^p_{\{j,n,k\}}=(1-\sigma)\ln\left(\widetilde{1+t^p_{\{n,k\}}}\right)+(1-\sigma)\delta\ln\tilde{D}_{\{n,k\}}+\ln\tilde{\epsilon}^p_{\{j,n,k\}}. \tag{18}$$

This approach involves a linear regression of log "tetraded" flows on log "tetraded" trade costs and does not require the estimation of any fixed effect. This method is therefore very simple computationally. It also lends itself easily to graphical analysis and will finally provide a natural test of non-constant aggregate elasticities, which we conduct in Section 5.3.

4.1.3. Firm-level zeroes (selection bias)

The third challenge is to account for the selection of firms into different export markets. Assuming that fixed export costs vary across markets and are paid using labor of the origin country, profits in this setup are given by $x^p_{ni}(\alpha)/\sigma-w_if^p_n$. From Eq. (15), we see that a firm with a low cost (αw_i) can afford having a low draw on $\epsilon^p_{ni}(\alpha)$ and still export profitably to n. The same logic applies for large (high A^p_n), and easy to reach (low τ^p_{ni}) markets. Concerning our variable of interest, higher tariff observations will be associated with firms having drawn higher $\epsilon^p_{ni}(\alpha)$, thus biasing downwards our estimate of the trade elasticity. The solution to this selection bias is not trivial in our case where a large set of fixed effects is included. Although we are unaware of a "perfect" estimator, we propose three alternative methods, that we confront to Monte Carlo evidence of a simulated version of the model.

First, one can focus the regressions on firms that have such a large productivity that their idiosyncratic destination shock is of second order. Inspired by Mulligan and Rubinstein (2008), Paravisini et al. (2015) and Fitzgerald and Haller (2015), we concentrate the analysis on large firms that serve almost all markets. This requires to decide on a variable likely to predict small levels of selection. Paravisini et al. (2015) use firm-level measures of total exports and credit, while Fitzgerald and Haller (2015) use a threshold of firm-level employment. We implement this approach with our data by restricting the sample to the largest exporter in each origin-product. Because this approach can accommodate our two-way fixed effects procedure (firm and destination) very easily, we call it 2WFE on top exporters. The second estimator relies upon the Tetrads method, with a similar strategy of restricting attention to large exporting firms that are the least likely to be affected by the selection bias. When taking ratios of ratios of individual trade flows, we focus on the top exporters of each country,[13] and look at their relative exports in different markets (compared to a reference country). We expect those two methods to give comparable results. The issue with both estimators is that they estimate the firm-level trade elasticity on a reduced sub-sample of the largest firms. Those might have different trade elasticities, for reasons outside of our model.[14] Our third estimator reinstates the full sample of exporters. Assuming a normally distributed $\ln\epsilon^p_{ni}(\alpha)$ in Eq. (15) yields a generalized structural Tobit, that we will refer to as EK-Tobit, since it was developed by Eaton and Kortum (2001). Crozet et al. (2012) apply EK-Tobit to the heterogeneous exporter model by using the theoretical equation for minimum sales, $x^{p,\mathtt{MIN}}_{ni}(\alpha)=\sigma w_if^p_n$, which therefore provides a natural estimate for the truncation point for each market. EK-Tobit is the best estimator for our theoretical framework, with an important caveat: We must reduce the number of included fixed effects because it seems computational unfeasible to estimate a generalized Tobit with the very large set of fixed effects our theory demands.[15]

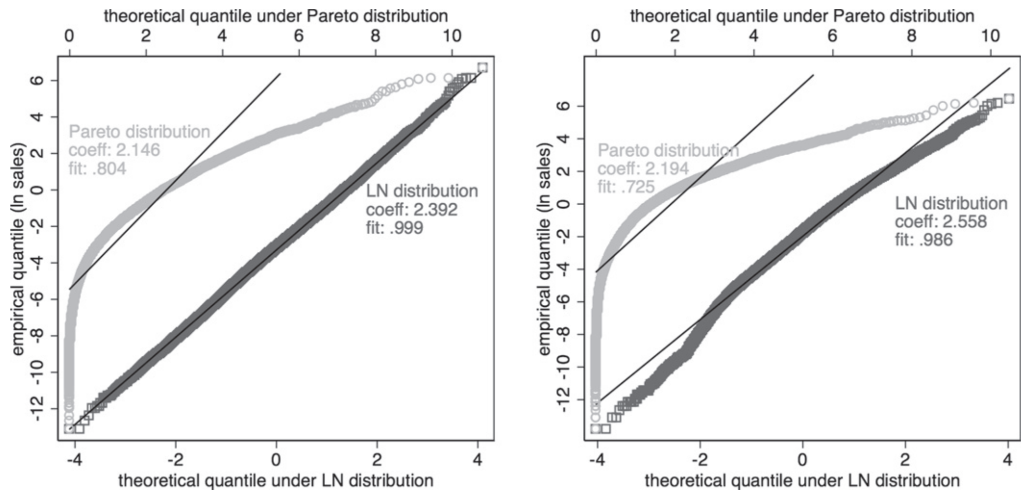
We therefore have three possible estimators, 2WFE on top exporters, Tetrads on top exporters, and EK-Tobit. We now proceed to test for the performance of our three imperfect estimators meant to correct for the selection bias using Monte Carlo simulations. The DGP uses Eq. (15) for the value $x_{ni}(\alpha)$ exported by 100,000 firms divided into two origin countries and selling in 80 (to roughly match the numbers we have in our sample). The true value of σ is set to 5. The fixed export costs f_n and the market size A_n are drawn from independent log-normal distributions calibrated to generate the same proportion

12 The fastest procedure to date available in Stata, `reghdfe`, has been developed by Sergio Correia building on Guimarães and Portugal (2010).

13 j and ℓ are chosen as the top exporters to k in value terms in Eq. (17).

14 Berman et al. (2012) show that several models featuring variable markups predict that large firms should face lower demand elasticities and therefore react less than small firms to a change in trade costs. Their finding that the response to exchange rate changes declines with productivity (confirmed by Chatterjee et al. (2013) for Brazilian exporters and Li et al. (2015) for Chinese exporters) suggests that the estimates in the present paper could be considered as a lower bound.

15 Although Greene (2004) shows that the Tobit model is much less subject to the incidental parameters problem than other non-linear models such as logit or probit, it is not possible to include the very large number of fixed effects in the EK-Tobit model due to computational burden. We also detect no sign of bias in our Monte Carlo simulations.



(a) French firms in Belgium (b) Chinese firms in Japan

Fig. 1. Distribution of firm-level sales. Note: In the QQ regressions reported, the dependent variable is the firm-level log of exports in 2000. The RHS is $\Phi^{-1}(\hat{F}_f)$ for log-normal and $\ln(1-\hat{F}_f)$ for Pareto, where $\hat{F}_f$ is the empirical CDF of log sales and Φ is the CDF of the standard normal. Under the usual CES (σ)/constant markup assumptions, coefficients have a direct interpretation in terms of structural parameters: $\frac{\sigma-1}{\theta}$ if productivity is Pareto with shape parameter θ, and $(\sigma-1)\nu$ for log-normal productivity with dispersion parameter ν. The fit measure is the R^2 of each regression.

of zero valued flows we find in the data (about 95%). For a given origin-destination, the firm-level sales distribution follows the distribution of $a = \alpha \times b(\alpha)$. Following results of firm-level sales distribution shown in Section 5, we assume a log-normal distribution for a. The key parameter of this distribution is its standard deviation, which is equal to the Quantile–Quantile(QQ) regression coefficient divided by $\sigma-1$ (see Head et al., 2014 and Section 5 for details). We set this parameter to be equal to the average of the two regression coefficients obtained for French and Chinese exporters in Fig. 1. An important aspect of the simulation lies in the choice of the relative importance of firm-level cost, α, and unobserved firm-destination shock $b_{ni}(\alpha)$ in the firms' sale Eq. (2). We calibrate the relative contribution of productivity and demand shocks, α and $b_{ni}(\alpha)$, fitting the correlation between the rank of a firm in total exports of the country and its rank in sales to each country n. Without the random demand term our model predicts a perfect correlation, while this rank correlation would approach zero if $b_{ni}(\alpha)$ is the only source of firm-level heterogeneity (see Appendix C for a theoretical discussion). Our data reveals that this correlation is around 66% in both the French and Chinese cases.

Table 1 summarizes our Monte Carlo results based on 1000 replications. Mean and standard deviation of the sampling distribution are reported for various statistics. When the exports are censored (setting unprofitable exports to 0), the correlation between τ_{ni} and the error term is about −14 %, sufficient to create a massive bias in the estimated trade elasticity which falls to about half its true value (rows 4 and 5). EK-Tobit with the appropriate set of fixed effects (row 6) recovers almost exactly the true coefficient. Perhaps more surprising, EK-Tobit without any fixed effect also is very close to the true trade elasticity. This is due to the fact that the simulation assumes no correlation between τ_{ni} and either A_n or α. Since EK-Tobit considers the full sample of potential flows, no selection bias can occur through that channel. However, there might be some correlation between τ_{ni} and A_n for instance in the true data, suggesting the need to introduce proxies of A_n in empirical implementations. Rows 8 and 9 report the results for the two other estimators. Both seems to be slightly biased, 8% for 2WFE on top exporters, 14% for Tetrads on top exporters, even though they do not differ significantly from the true value of σ.

Table 1
Monte Carlo results: firm-level elasticities wrt to a change in trade costs.

	Mean	s.d.
% of positive flows	0.050	0.008
Correlation between global and local rank	0.662	0.015
Correlation between $\ln\tau_{ni}$ and $\ln b_{ni}(\alpha)$	−0.138	0.045
σ 2WFE on full sample	5.000	0.004
σ 2WFE on censored sample	2.438	0.113
σ EK-Tobit	4.997	0.014
σ EK-Tobit (no FEs)	5.019	0.576
σ Tetrads	4.311	0.921
σ 2WFE on top exporter	4.603	1.068
# obs full sample (and EK-Tobit)	8,000,000	0
# obs censored sample	398,408.063	60,719.656
# obs Tetrads	326.036	48.222
# obs 2WFE on top exporter	133.779	14.097

Note: True σ is set to 5. There are 1000 replications, parameters on fixed costs of exports and size of the demand term have been calibrated so that the share of non-selected trade flows at the firm-destination level averages between 4 and 5%. For each elasticity, the first column reports the average value, while the second reports standard deviations of elasticities across the 1000 replications.

4.2. Data

We combine French and Chinese firm-level data sets from the corresponding customs administrations which report export value by firm at the HS6 level for all destinations in 2000. The firm-level customs data sets are matched with data on tariffs effectively applied to each exporting country (China and France) at the same level of product disaggregation for each destination. Focusing on 2000 allows us to exploit variation in tariffs applied to each exporter country (France/China) at the product level by the importer countries since it precedes the entry of China into WTO at the end of 2001.[16]

4.2.1. Trade

The French trade data comes from the French Customs, which provide annual export data at the product level for French firms.[17] The customs data are available at the 8-digit product level Combined Nomenclature (CN) and specify the country of destination of exports.

[16] We exploit the variation over time of trade and tariffs from 2000 to 2006 in a set of robustness checks reported in the working paper version.

[17] This database is quite exhaustive. Although reporting of firms by trade values below 250,000 FF – less than 39,000 euros –(within the EU) or 1000 euros (rest of the world) is not mandatory, there are in practice many observations below these thresholds.

The free on board (f.o.b.) value of exports is reported in euros and we converted those to US dollars using the real exchange rate from Penn World Tables for 2000. The Chinese transaction data comes from the Chinese Customs Trade Statistics (CCTS) database which is compiled by the General Administration of Customs of China. This database includes monthly firm-level exports at the 8-digit HS product-level (also reported f.o.b.) in US dollars. The data is collapsed to yearly frequency. The database also records the country of destination of exports. In both cases, export values are aggregated at the firm–product(HS6)–destination level in order to match with applied tariffs information that are available at the origin–product(HS6)–destination level.

4.2.2. Tariffs

Tariffs come from the WITS (World Bank) database.[18] We rely on the ad valorem rate effectively applied at the HS6 level by each importer country to France and China. In our cross-section analysis performed for the year 2000 before the entry of China into the World Trade Organization (WTO), we exploit different sources of variation within HS6 products across importing countries on the tariff applied to France and China. The first variation naturally comes from the European Union (EU) importing countries that apply zero tariffs to trade with EU partners (like France) and a common external tariff to extra-EU countries (like China). The second source of variation in the year 2000 is that several non-EU countries applied the Most Favored Nation tariff (MFN) to France, while the effective tariff applied to Chinese products was different (since China was not yet a member of WTO). In the Online Appendix, Fig. 1 illustrates those sources of variation in applied tariffs, graphing the average difference between the tariffs applied to France and China across industries by two major trading partners: Germany and Japan. While Germany naturally favors French exports across the board, Japan has the opposite policy, all industries featuring a preference towards Chinese exporters (except a few sectors where the difference is nil). Furthermore, those tariff differences show substantial variance across industries.

4.2.3. Gravity controls

In all estimations, we include additional trade barriers variables that determine bilateral trade costs, such as distance, common (official) language, colony and common border (contiguity). The data come from the CEPII distance database.[19] We use the population-weighted great circle distance between the set of largest cities in the two countries.

4.3. Empirical estimates of the firm-level trade elasticity

Table 2 reports the results for firm-level trade elasticity based on our three estimators applied to the French and Chinese firm-level exports in 2000. The first three columns use 2WFE on top exporters, the next 3 use Tetrads on top exporters, while the last three use EK-Tobit.

While 2WFE on top is straightforward to implement, one needs to define reference k countries for Tetrads (Eq. (18)). We choose those with two criteria in mind. First, these countries should be those that are the main trade partners of France and China in the year 2000, since we want to minimize the number of zero trade flows in the denominator of the Tetrad. The second criteria relies on the variation in the tariffs effectively applied by the importing country to France and China. Hence, among the main trade partners, we retain those countries for which the average difference between the effectively applied ad valorem tariffs to France and China is greater. These two criteria lead us to select the following set of 8 reference countries: Australia, Canada, Germany, Italy, Japan, New Zealand, Poland and the UK.

When implementing EK-Tobit, we need to fill in (with zero flows) the destinations that a firm found unprofitable to serve. The set of potential destinations for each product is given by all countries where at least one firm exported that good. When estimating Eq. (15) through EK-Tobit, we proxy for $\ln A_n^p$ with destination n fixed effects, and for firm-level determinants α with the count of markets served by the firm. An origin country dummy for Chinese exporters account for all differences across the two groups, such as wages, w_i.

For each of the three methods of estimation, the first column includes tariffs and the usual set of gravity variables (distance, contiguity, colonial link and common language). The second column adds a dummy variable set to one for active Regional Trade Agreements (RTA). The idea is to control for unobserved non-tariff barriers to trade that could potentially be correlated with ad valorem applied tariffs. This is a particularly demanding specification, since a lot of the variance in tariffs should come from the distinction between RTA members facing zero tariffs and non-member pairs facing positive ones. The third column of each method controls for all unobservable bilateral frictions. This is done by including destination–origin fixed effects for 2WFE and EK-Tobit (columns 3 and 9). Since Tetrads rely on a ratio of flows going to a destination compared to a reference country, all bilateral unobservables characteristics are taken into account by a destination–reference fixed effect (column 6).

The Tetrads and 2WFE methods on top exporters in columns (1) to (6) show quite similar patterns of results, as expected from the similarity in approach and Monte Carlo results. Distance has the usual negative coefficient, contiguity enters strongly positive, while colonial link and common language have a much more volatile and mostly insignificant effect at the firm level. RTAs enter with a very comparable and strong effect in both methods (approximately tripling trade flows), with the expected effect of reducing the impact of tariffs. Coefficients on tariffs relying on the EK-Tobit method (columns (7) to (9)) are slightly greater than in the previous methods. Overall the three methods point to similar coefficients with a reasonable value of the firm-level elasticity $(1-\sigma)$ averaging -4.4 across the 9 columns. This turns out to be a central value within the small set of papers estimating the response of firm-level flows to applied tariffs. Berthou and Fontagné (2016) obtain coefficients that would imply a preferred value of $\sigma = 3.5$ and report a larger response when restricting the sample to the largest exporters, as expected from our analysis of selection bias above. Fitzgerald and Haller (2015) report a very strong variance in firm-level response to tariffs, with implied value of σ strongly rising when restricting regressions to the largest firms. Their benchmark table imply σ ranging from 2.7 to around 25, the latter being relevant for the biggest firms in the most popular markets.

Our results are robust to several sensitivity tests that are presented and described in the Online Appendix. First, we re-estimate the firm-level trade elasticity on alternative levels of aggregation (firm–HS4(HS2)–destination and firm–destination) using the EK-Tobit method. Coefficients on tariffs are larger than in the baseline estimation and increase as the sample gets more aggregated.[20] Second, we have run our benchmark estimates for the years 2001 and 2006. Although estimates of the firm-level trade elasticity are slightly smaller in absolute value, they remain close to the results for the year

[18] Information on tariffs is available at http://wits.worldbank.org/wits/.

[19] This data set is available at http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

[20] This finding is consistent with a substantial effect of tariffs on the extensive margin of products, and is in line with the patterns of results in Fitzgerald and Haller (2015), when comparing the third columns of their Tables 8 and 12.

Table 2
Intensive margin elasticities in 2000: 3 methods.

Estimator:	2WFE on top			Tetrad on top			EK-Tobit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln (1 + Applied tariff)	-4.95^b	-3.14	-3.13	-4.75^a	-3.25^a	-2.57^a	-5.54^a	-5.54^a	-6.88^a
	(1.94)	(1.96)	(2.21)	(0.79)	(0.74)	(0.50)	(0.26)	(0.26)	(0.26)
ln Distance	-0.52^a	-0.19^a		-0.48^a	-0.16^a		-1.73^a	-1.66^a	
	(0.04)	(0.06)		(0.03)	(0.04)		(0.04)	(0.06)	
Common language	-0.36^a	-0.10		0.07	0.37^a		1.86^a	1.93^a	
	(0.11)	(0.12)		(0.09)	(0.08)		(0.12)	(0.12)	
Contiguity	0.99^a	0.89^a		0.57^a	0.52^a		1.43^a	1.41^a	
	(0.09)	(0.09)		(0.08)	(0.07)		(0.11)	(0.11)	
Colonial link	0.45	-0.14		0.31	-0.22		3.49^a	3.40^a	
	(0.53)	(0.54)		(0.29)	(0.29)		(0.21)	(0.21)	
RTA		1.13^a			1.07^a			0.25^b	
		(0.18)			(0.12)			(0.13)	
ln # of dest. by firm							1.69^a	1.69^a	
							(0.02)	(0.02)	
Chinese exporter							0.59^a	0.64^a	
							(0.06)	(0.05)	
Fixed effects:									
Firm	Yes	Yes	Yes						
HS6–destination	Yes	Yes	Yes						
Destination–origin			Yes			Yes			Yes
Destination							Yes	Yes	
Observations		14,124			37,706			49,067,666	
R^2	0.777	0.779	0.782	0.137	0.144	0.177	0.829/0.08		0.830/0.08

Notes: ^{a}, ^{b} and ^{c} denote statistical significance levels of 1, 5 and 10% respectively. Because Tetrads compare ratios of flows from the two origin countries going to destination n and to reference k, column (6) relies on destination × reference fixed effects to capture all destination–origin determinants. Standard errors are clustered by destination × reference country for columns (4) to (6) (Tetrads), and by HS6–origin–destination for columns (7) to (9). Those three last columns compute the R^2 as the squared correlation between the predicted and actual values of the dependent variable. The second R^2 does the same calculation on positive trade flows.

2000, and maintain similar patterns of statistical significance. Third, we present evidence on an alternative definition of largest firms as the set of exporters selling to the largest number of destinations. At the HS6 level of detail, a number of such origin–product combination have the top firm in terms of destinations sell to only a handful of destinations. We therefore restrict the attention to the set of those firms that export to more than 20 markets (the median destination count). The estimated elasticities of firm-level responses to tariffs are of comparable magnitude.

The working paper version of this paper concentrated on the Tetrads approach to estimation and provided many robustness checks of the firm-level trade elasticity. We briefly report a summary of those results here. There are essentially two sources of variance of tariffs in our setting: across products and across destinations. When focusing on the cross-destination dimension through the use of product fixed effects, the coefficients for the applied tariffs $(1-\sigma)$ range from -6 to -3.2. Restricting the sample to destination countries which apply non-MFN tariffs to France and China (Australia, Canada, Japan, New Zealand and Poland), yields results of similar magnitude, ranging from -5.47 to -3.24. We also consider two additional cross-sectional samples, one after China's entry into the WTO (2001), the other for the final year for which we have Chinese customs data (2006). Here again the results are qualitatively robust, although the coefficients on tariffs are lower as expected since the difference in the tariffs applied to France and China by destination countries is much reduced after 2001. With the caveat in mind that entry into the WTO combined with patchy data over time for many countries makes panel estimation quite difficult, we consider it over the 2000–2006 period. The coefficients are more volatile, but somehow close to the findings from the baseline cross-section estimations in 2000 (they range from -5.26 to -1.80). This set of robustness estimates combined with the ones in Table 2 points to a central value of the demand side parameter of our model located around 5. We use $\hat{\sigma} = 5$ in the coming section in order to calculate aggregate trade elasticities relevant for the same sample of French and Chinese exporters.

5. Bilateral aggregate trade elasticities

5.1. Theory with numbers: micro-based predictions of aggregate trade elasticities

In this section, we provide a theory-consistent methodology for predicting, from firm-level data, a set of three aggregate elasticities of trade with respect to trade costs (the reactions of total trade, number of exporters and average exports per firm to tariffs). Those *micro-based predictions* of aggregate elasticities are characterized by Eqs. (12), (13) and (14). In each of those, the distribution of rescaled labor requirement, the CDF $G(a)$ enters prominently. Specifying $G(a)$ is also necessary to invert the $\mathcal{H}(a^*)$ function, and reveal the bilateral cutoffs required to compute the bilateral trade elasticities. We consider two possibilities for $G(a)$: following the almost universally chosen Pareto approach, or going a natural alternative route motivated from data patterns, the log-normal distribution. The latter is a much better fit of the firm-level exports for the overall distribution (as Fig. 1 below shows), making it a credible and natural alternative to Pareto.[21]

Pareto-distributed rescaled productivity $\varphi \equiv 1/a$ translates into a power law CDF for a, with shape parameter θ and location parameter $\bar{a}$. A log-normal distribution of a retains the log-normality of productivity (with location parameter μ and dispersion parameter ν) but with a change in the log-mean parameter from μ to $-\mu$. Under those two distributional assumptions the CDFs for $a > 0$ are therefore given by

$$G^{\mathrm{P}}(a) = \max\left[\left(\frac{a}{\bar{a}}\right)^{\theta}, 1\right], \qquad \text{and} \qquad G^{\mathrm{LN}}(a) = \Phi\left(\frac{\ln a + \mu}{\nu}\right), \quad (19)$$

[21] It is worth noting that one can obtain a sales distribution with a similar shape as shown in Fig. 1 by combining Pareto-distributed productivity with log-normally-distributed demand shocks, as done in Eaton et al. (2011). This setup however maintains the property of a constant aggregate trade elasticity, which depends solely on the Pareto shape parameter, as in Chaney (2008).

where we use Φ to denote the CDF of the standard normal. Simple calculations using Eq. (19) in Eq. (8), and detailed in Appendix A, show that the resulting formulas for $\mathcal{H}$ are

$$\mathcal{H}^{\mathrm{P}}(a^*_{ni}) = \frac{\theta}{\theta - \sigma + 1}, \text{ and } \mathcal{H}^{\mathrm{LN}}(a^*_{ni}) = \frac{h\left[(\ln a^*_{ni} + \mu)/\nu\right]}{h\left[(\ln a^*_{ni} + \mu)/\nu + (\sigma - 1)\nu\right]}, \tag{20}$$

where $h(x) \equiv \phi(x)/\Phi(x)$, the ratio of the PDF to the CDF of the standard normal.

Calculating $G^{\mathrm{P}}(.)$, $G^{\mathrm{LN}}(.)$, $\mathcal{H}^{\mathrm{P}}(.)$ and $\mathcal{H}^{\mathrm{LN}}(.)$ requires knowledge of underlying key supply-side distribution parameters θ and ν.[22] For those, we rely on estimates from QQ regressions combined with our estimate of the demand side parameter ($\hat{\sigma} = 5$) obtained from the preceding section.

With CES demand and constant markups, Head et al. (2014) show that the distribution of sales in a given destination inherits the distribution of the firms' underlying performance variable (productivity). When the latter is distributed Pareto or log-normal, sales are also distributed Pareto and log-normal, the only substantial difference being a shift in the shape parameter of each of those distributions. While we do not observe productivity, we do observe the distribution of sales, and can use those to reveal the underlying structural parameters θ and ν. A very useful tool for that purpose is the QQ regression, where the empirical quantile (log sales) is regressed on the theoretical quantile under each alternative distributional assumption. Denoting with $\hat{F}_f$ the empirical CDF of log sales, and f now indexing firms in ascending order of individual sales, we have the theoretical quantiles

$$\mathbb{Q}^{\mathrm{P}}_f = q^{\mathrm{P}} - \frac{\sigma - 1}{\theta}\ln(1 - \hat{F}_f), \qquad \text{and} \qquad \mathbb{Q}^{\mathrm{LN}}_f = q^{\mathrm{LN}} + (\sigma - 1)\nu\Phi^{-1}(\hat{F}_f). \tag{21}$$

QQ regressions are thus linear in both Pareto and log-normal cases, and the slope reveals the shape parameter of the underlying distribution (the constant terms q^{P} and q^{LN} corresponding to location parameters). Fig. 1 reports those regressions for the two sets of exporters used in this paper. We focus on a major destination for each of those countries, Belgium and Japan respectively. The Pareto regression is represented in light gray and the log-normal one in dark gray. It is very clear that the log-normal QQ plot is much closer to the linear relationship that should obtain when assuming the correct distribution.

What are the implied values of structural distribution parameters? The log-normal case is simple, since it is a very good fit to the overall distribution, we simply have $\hat{\nu}_{\mathrm{FRA}} = 2.392/(\hat{\sigma} - 1) = 0.598$ and $\hat{\nu}_{\mathrm{CHN}} = 2.558/(\hat{\sigma} - 1) = 0.639$. The Pareto case is more tricky since the implied values of $\hat{\theta}$ for the overall estimation of the QQ regression (2.146 and 2.194) are incompatible with finite values of the price index. We therefore concentrate on the part of the distribution where the Pareto QQ relationship is approximately linear with a slope satisfying $\hat{\theta} > \hat{\sigma} - 1$, that is the extreme right tail (like most papers that estimate Pareto shape parameters on sales data). Concentrating on the top 1% of sales (in terms of value exported), we obtain QQ coefficients of 0.779 and 0.618 respectively, which yield $\hat{\theta}_{\mathrm{FRA}} = (1/0.779)(\hat{\sigma} - 1) = 5.134$ and $\hat{\theta}_{\mathrm{CHN}} = (1/0.618)(\hat{\sigma} - 1) = 6.472$.

Panel (a) of Fig. 2 depicts the theoretical relationship between the ratio of mean to minimum sales, $\mathcal{H}(a^*_{ni})$ in Eqs. (8) and (9), and the probability of serving the destination market, $G(a^*_{ni})$, spanning over values of the cutoff a^*_{ni}. Under Pareto heterogeneity, $\mathcal{H}$ is constant but this property of scale invariance is specific to the Pareto: $\mathcal{H}$ is increasing in G under log-normal. Panel (b) of Fig. 2 depicts the empirical counterpart of this relationship as observed for French and Chinese exporters in 2000 for all countries in the world. On the x-axis is the share of exporters serving each of those markets.[23] Immediately apparent is the non-constant nature of the mean-to-min ratio in the data, contradicting the Pareto prediction. This finding is very robust when considering alternatives to the minimum sales (which might be noisy because of statistical threshold effects) for the denominator of $\mathcal{H}$, that is different quantiles of the export distribution (results available upon request). In a further effort to minimize noise in the calculation of the mean-to-min ratio, the figures are calculated for each of the 99 HS2 product categories and averaged. In the rest of the section, we will stick to this approach for the calculation of elasticities, done at the HS2 level before being averaged, which also simplifies exposition (detailed sector-level results are provided in Section 6).

Fig. 3 turns to the predicted aggregate trade elasticities.[24] Functional forms Eqs. (19) and (20) combined with Eq. (11), are used to deliver bilateral aggregate trade elasticity of total flows, ε_{ni}, under the two alternative distributional assumptions:

$$\varepsilon^{\mathrm{P}}_{ni} = -\theta, \qquad \text{and} \qquad \varepsilon^{\mathrm{LN}}_{ni} = 1 - \sigma - \frac{1}{\nu} h\left(\frac{\ln a^*_{ni} + \mu}{\nu} + (\sigma - 1)\nu\right). \tag{22}$$

Parallel to Fig. 2, panel (a) of Fig. 3 shows the theoretical relationship between those elasticities and $G(a^*_{ni})$, while panel (b) plots the same elasticities evaluated for each individual destination country against the empirical counterpart of $G(a^*_{ni})$ (described in footnote 23). Again, the Pareto case has a constant prediction, while log-normal predicts a trade elasticity that is declining (in absolute value) with easiness of the market. Panel (b) confirms the large variance of trade elasticities according to the share of exporters that are active in each of the markets. It also shows that the response of aggregate flows to trade costs is reduced (in absolute value) when the market becomes easier. The intuition is that for very difficult markets, the individual reaction of incumbent firms is supplemented with entry of exporters selected among the most efficient firms. The latter effect becomes negligible for the easiest markets, yielding ε to approach the firm-level trade elasticity. This mechanism becomes very clear when looking at the patterns of the extensive margin and average export elasticities in Fig. 4.

The predicted elasticity on the extensive margin is also rising with market toughness as shown in panel (a) of Fig. 4. The inverse relationship is true for average exports per firm (panel b). When a market is very easy and most exporters make it there, the extensive margin elasticity goes to zero, and the response of average exports per firm goes to the value of the firm-level trade elasticity, $1 - \sigma$, as shown in Fig. 4 when the share of exporters increase. While this should intuitively be true in general, Pareto does not allow for this change in elasticities across markets, since the response of average exports per firm should be uniformly 0, while the total response is entirely due to the (constant) extensive margin elasticity.

In Table 3, we compute the mean and standard deviation of the bilateral trade elasticities calculated using the log-normal distribution, and presented in Figs. 3 and 4. The first column presents the statistics for the French exporters' sample, the second one is the Chinese

[22] We show in Appendix A that the values taken by $\bar{a}$ and μ do not affect calculations of the trade elasticity.

[23] On the x-axis we have the share of exporters serving each market n computed as the number of French and Chinese firms serving market n in 2000 conditional on exporting (divided by the total number of exporters in each origin). The exact empirical counterpart of $G(a^*_{ni})$ would require to divide the number of actual French and Chinese exporters to n by the set of *potential* exporters in France and China. We don't observe that last number. However, the theoretical and the empirical proportion of exporters differ only by a multiplicative constant, leaving the *shape* of the (logged) relationship unchanged.

[24] See Appendix A for details.

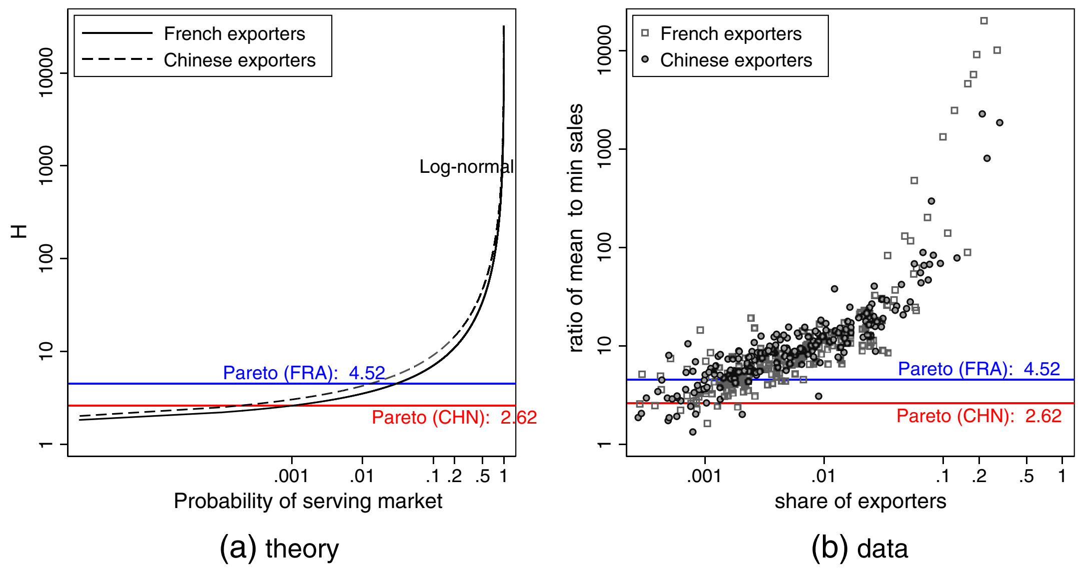


Fig. 2. Theoretical and empirical mean-to-min ratios.

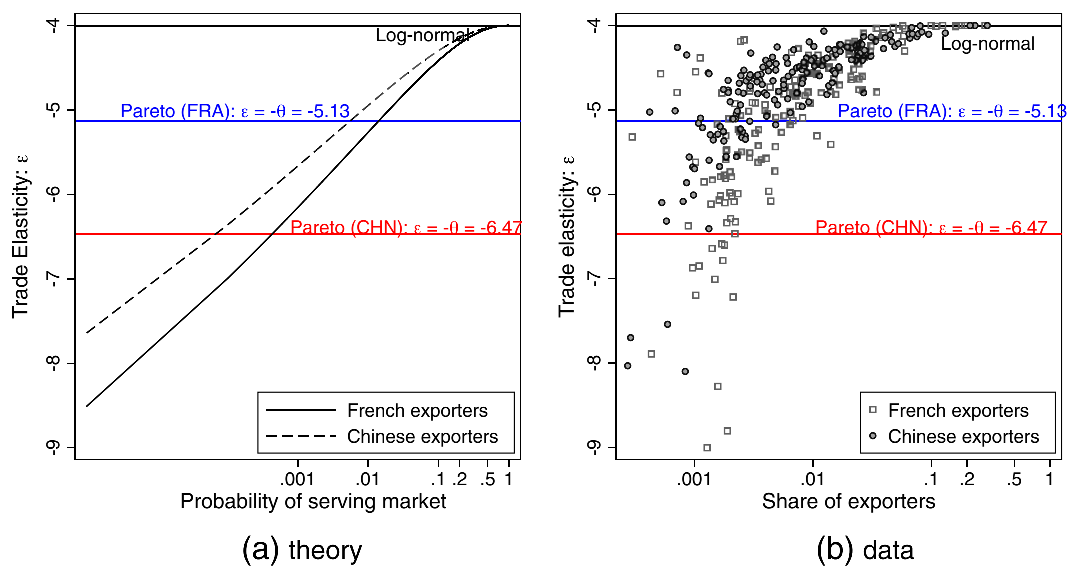


Fig. 3. Predicted bilateral aggregate trade elasticities: $\varepsilon_{n\mathrm{FR}}$ and $\varepsilon_{n\mathrm{CN}}$.

exporters' case, and the last column averages those. The mean elasticities obtained vary slightly between France and China, but the dominant feature is that the aggregate elasticity of total flows is in neither case confined to the extensive margin (i.e. number of exporters). In both cases, average exports per firm are predicted to react strongly to trade costs, a pattern we will confirm on actual data in the next subsection.

Fig. 5 groups our predictions of the bilateral aggregate elasticity of total flows (ε_{ni}) into ten bins of export shares for both France and China in a way similar to empirical evidence by Novy (2013) and Spearot (2013), who find that the aggregate trade cost elasticity decreases with bilateral trade intensity.[25] The qualitative pattern is very similar here, with the bilateral aggregate elasticity decreasing in absolute value with the share of exports going to a destination. One can use this variance in ε_{ni} to quantify the difference with respect to a constant response of exports to a trade liberalization episode. Taking China as an example, decreasing trade costs by 1% would raise flows by around 6.5% for countries like Somalia, Chad or Azerbaijan (first bin of Chinese exports) and slightly more than 4% for the USA and Japan (top bin). Since the estimate that would be obtained when imposing a unique elasticity would be close to the average elasticity (4.79), this would entail about 25% underestimate of the trade growth for initially low traders (1.7/6.5) and an overestimate of around 20% (0.8/4) for the top trade pairs (with the caveat that we have more variance in the predicted elasticities for low export bins than in the top ones).[26]

5.2. Comparison with macro-based estimates of aggregate trade elasticities

We now turn to empirical estimates of aggregate trade elasticities to be compared with our predictions. Those are obtained using

[25] Although Novy (2013) estimates variable distance elasticity, his Section 3.4 assumes a constant trade costs to distance parameter to focus on the equivalent of our ε_{ni}. Spearot (2013) estimates the aggregate trade elasticity directly through the impact of tariff changes on US imports between 1992 and 2004.

[26] We thank Steve Redding for suggesting this quantification.

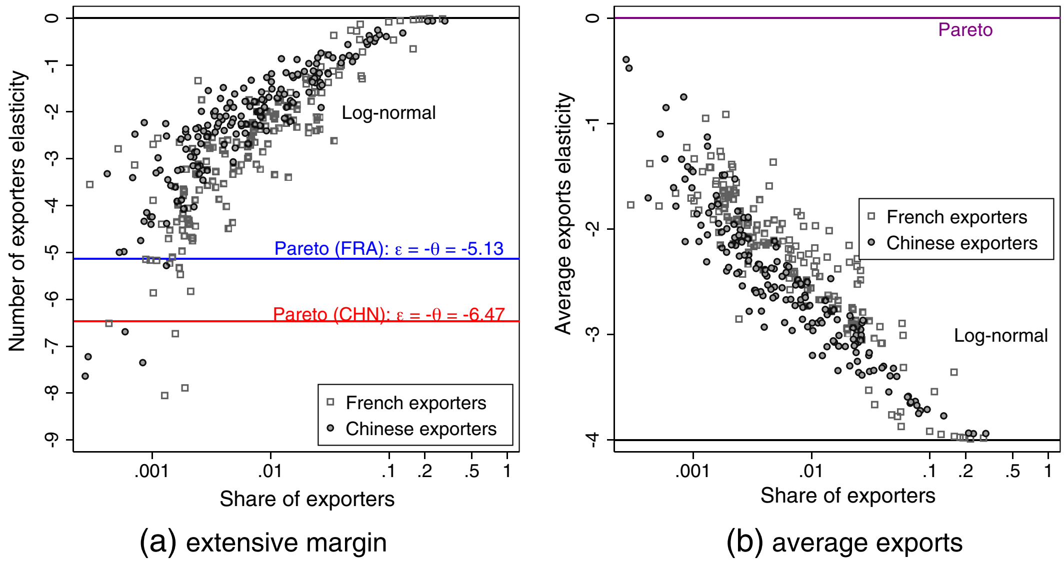


Fig. 4. Predicted bilateral aggregate elasticities: extensive and average exports.

aggregate versions of our estimating Tetrad equations presented above, which is very comparable to the method most often used in the literature: a gravity equation with country fixed effects and a set of bilateral trade costs covariates, on which a constant trade elasticity is assumed.[27] Column (1) of Table 4 uses the same specification as in column (4) of Table 2 but runs the regression on the Tetrad of aggregate rather than individual exports. Column (2) uses the same covariates but on the count of exporters, and column (3) completes the estimation by looking at the effects on average exports per firm.[28] An important finding is that the effect on average exports per firm is estimated at −2.77, and is significant at the 1% level, contrary to the Pareto prediction (in which no variable trade cost should enter the equation for average flows).[29] This finding is robust to controlling for RTA (columns 4–6) or destination–reference fixed effects (columns 7–9). It is interesting to note that our middle estimate of the trade elasticity on total flows is −4.79, reasonably close to the −5.03 found as the median estimate in a large set of regressions covered by Head and Mayer (2014).

Under Pareto, the aggregate elasticity should reflect fully the one on the number of exporters, and there should be no impact of tariffs on average exports per firm. This prediction of the Pareto distribution is therefore strongly contradicted by our results. As a first pass at assessing whether the data support the log-normal predictions, we compare the (unique) macro-based estimate of aggregate elasticities obtained in Table 4, to their corresponding micro-based predictions shown in Table 3. The sample mean obtained with the model's predictions are not out of the range of empirical estimates. Although this is not a definitive validation of the heterogenous firms model with log-normal distribution, our results clearly favor this distributional assumption over Pareto, and provide support for the empirical relevance of non-constant trade elasticities.

5.3. Direct evidence of non-constant bilateral aggregate trade elasticities

We can further use the Tetrad methodology in order to show *direct* empirical evidence of non-constant bilateral trade elasticities. Using aggregate bilateral flows from Eq. (3), and building Tetrads with a procedure identical to the one used at the firm level, we obtain the (FR,CN, n, k)—Tetrad of aggregate exports

$$\tilde{X}_{\{n,k\}} \equiv \frac{X_{n\mathrm{FR}}/X_{k\mathrm{FR}}}{X_{n\mathrm{CN}}/X_{k\mathrm{CN}}} = \left(\frac{\tau_{n\mathrm{FR}}/\tau_{k\mathrm{FR}}}{\tau_{n\mathrm{CN}}/\tau_{k\mathrm{CN}}}\right)^{1-\sigma} \times \frac{V_{n\mathrm{FR}}/V_{k\mathrm{FR}}}{V_{n\mathrm{CN}}/V_{k\mathrm{CN}}} \tag{23}$$

Taking logs, differentiating with respect to tariffs and using the expression for the cutoff Eq. (5), we obtain

$$d\ln\tilde{X}_{\{n,k\}} = (1-\sigma-\gamma_{n\mathrm{FR}}) \times d\ln\tau_{n\mathrm{FR}} - (1-\sigma-\gamma_{k\mathrm{FR}}) \times d\ln\tau_{k\mathrm{FR}} - (1-\sigma-\gamma_{n\mathrm{CN}}) \times d\ln\tau_{n\mathrm{CN}} + (1-\sigma-\gamma_{k\mathrm{CN}}) \times d\ln\tau_{k\mathrm{CN}}, \tag{24}$$

Table 3
Predicted bilateral aggregate trade elasticities (LN distribution).

	France	China	Average
Total flows	−5.14	−4.792	−4.966
	(1.069)	(.788)	(.742)
Number of exporters	−2.866	−2.274	−2.57
	(1.657)	(1.472)	(1.335)
Average flows	−2.274	−2.517	−2.396
	(.687)	(.731)	(.64)

Notes: This table presents the predicted elasticities (mean and standard deviation calculated across destinations) on total exports, the number of exporting firms, and average export flows. Required parameters are σ, the CES, and ν, the dispersion parameter of the log-normal distribution.

[27] Note that the gravity prediction on aggregate flows where origin, destination, and bilateral variables are multiplicatively separable and where there is a unique trade elasticity is only valid under Pareto. The heterogeneous elasticities generated by deviating from Pareto invalidate the usual gravity specification. Our intuition however is that the elasticity estimated using gravity/Tetrads should be a reasonable approximation of the *average* bilateral elasticities. In order to verify this intuition, we run Monte Carlo simulations of the model with log-normal heterogeneity and find that indeed the average of micro-based predictions of elasticities is very close to the unique macro-based estimate in a gravity/Tetrads equation on aggregate flows. Description of those simulations are in Appendix B.

[28] Note that the three dependent variables are computed for each HS6 product–destination, and therefore that the average exports per firm do not contain an extensive margin where number of products would vary across destinations.

[29] Fernandes et al. (2015) also show that the Melitz model combined with log-normal productivity can explain the reaction of average flows to distance. They refer to that response as the "intensive margin puzzle". We prefer to keep the terminology "intensive" for the firm-level response, and while we measure the trade elasticity directly through the impact of tariffs rather than distance, our results are totally in line with their main finding.

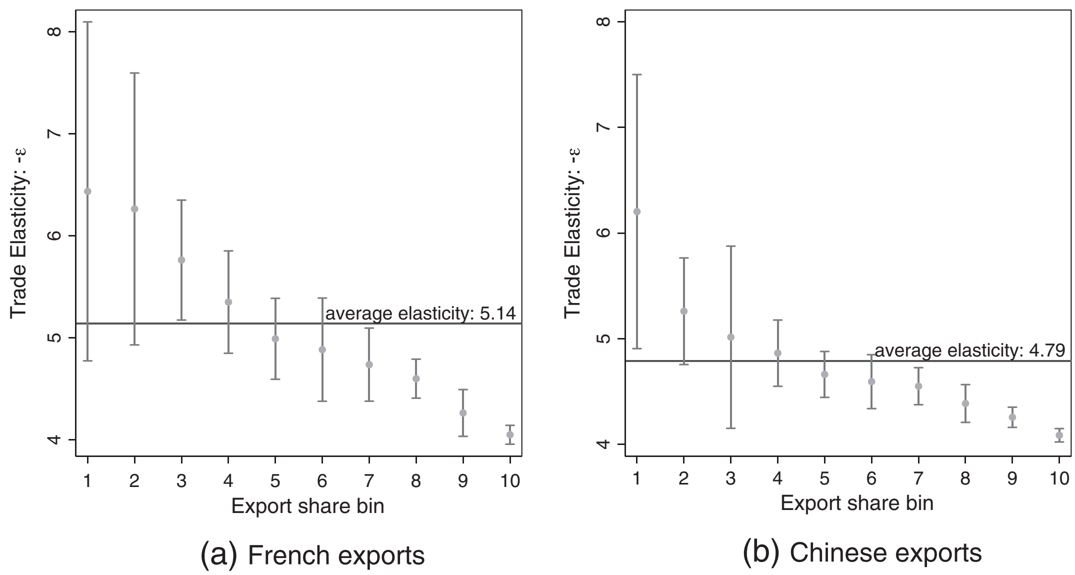


Fig. 5. Variance in bilateral aggregate trade elasticities: ε_{nFR} and ε_{nCN}. Note: The export share bins are constructed using the share in total export value to each destination country in 2000.

Table 4
Elasticities of total flows, count of exporters and average trade flows.

	Tot. (1)	# exp. (2)	Avg. (3)	Tot. (4)	# exp. (5)	Avg. (6)	Tot. (7)	# exp. (8)	Avg. (9)
ln (1 + Applied tariff)	-6.60^a	-3.83^a	-2.77^a	-4.79^a	-2.02^a	-2.77^a	-3.81^a	-2.45^a	-1.36^a
	(0.60)	(0.41)	(0.40)	(0.57)	(0.36)	(0.40)	(0.26)	(0.10)	(0.22)
ln Distance	-0.80^a	-0.54^a	-0.26^a	-0.43^a	-0.18^a	-0.26^a			
	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.03)			
Common language	0.05	0.10	−0.04	0.41^a	0.45^a	−0.04			
	(0.08)	(0.07)	(0.04)	(0.07)	(0.06)	(0.05)			
Contiguity	0.49^a	0.27^a	0.23^a	0.43^a	0.20^a	0.23^a			
	(0.09)	(0.06)	(0.05)	(0.08)	(0.05)	(0.05)			
Colonial link	0.89^a	0.78^a	0.12	0.28^c	0.16	0.12			
	(0.15)	(0.11)	(0.09)	(0.16)	(0.11)	(0.10)			
RTA				1.23^a	1.23^a	0.01			
				(0.12)	(0.07)	(0.08)			
R^2	0.336	0.543	0.081	0.346	0.581	0.081	0.382	0.635	0.109
rmse	2.56	1.09	2.13	2.54	1.04	2.13	2.47	0.98	2.10

Notes: 100,533 observations. ^{a}, ^{b} and ^{c} denote statistical significance levels of 1, 5 and 10% respectively. Standard errors are clustered by destination × reference country. The dependent variable is the tetradic term of the logarithm of total exports at the HS6–destination–origin country level in columns (1), (4) and (7); of the number of exporting firms by HS6–destination and origin country in columns (2), (5) and (8) and of the average exports per firm at the HS6–destination–origin country level in columns (3), (6) and (9). Applied tariff is the tetradic term of the logarithm of applied tariff plus one. Columns (7) to (9) present the estimations including destination–reference fixed effects that take into account all the unobservable bilateral frictions.

where γ_{ni} is the elasticity of the cost performance index to a rise in the easiness of the market, defined in Eq. (7). For general distributions of heterogeneity, this elasticity is not constant across country pairs as it depends on the bilateral-specific cutoff a^*_{ni}. Hence, our interpretation of Eq. (24) is that the contribution to the (tetraded) total exports of a change in bilateral tariffs is larger for country pairs that have a larger elasticity. Under Pareto, this elasticity is constant across country pairs, $\gamma^{\mathrm{P}}_{ni} = 1 - \sigma + \theta$. Combined with Eq. (24) this leads to

$$\tilde{\varepsilon}^{\mathrm{P}}_{\{n,k\}} = \frac{d\ln \tilde{X}_{\{n,k\}}}{d\ln \tilde{\tau}_{\{n,k\}}} = -\theta, \tag{25}$$

where $\tilde{\tau}_{\{n,k\}}$ is the vector of tetraded trade costs (each of the components of trade costs expressed as a ratio of ratios). This formula states that under Pareto, the elasticity of aggregate tetraded exports to tetraded tariffs is equal to the supply-side parameter θ. This transposes to the Tetrad environment the well-known result of Chaney (2008) on gravity. Under non-Pareto heterogeneity, the four elasticities in Eq. (24) will remain different, a prediction we can put to a test.

Results are shown in Table 5, where we pool observations for the years 2000 to 2006 (we also multiply $\ln\tau_{kFR}$ and $\ln\tau_{nCN}$ by −1, in order to have only negative figures–reflecting elasticities–in the Table, which eases interpretation). Columns (1), (2) and (3) are the equivalent of the first three columns of Table 4, with the trade costs Tetrads being split into its four components and the coefficients allowed to differ. The difference in coefficients on tariffs to destination country n is generally quite large, suggesting that the elasticity when considering France and China as an origin country differ, consistent with the non-Pareto version of heterogeneity.[30] Coefficients related to the reference importer k also differ substantially from each

[30] In terms of statistical significance, a Wald test reveals that the two coefficients on tariffs applied to France and China by destination n or by reference country k are indeed significantly different in columns (1), (2), (4) and (5) where the p-value of this test is zero. The p-value for the test on coefficients applied to France and China by destination n in column (3) is 0.79 and in column (6) is 0.02.

Table 5
Non-constant trade elasticity.

Dependent variable:	Tot. (1)	# exp. (2)	Avg. (3)	Tot. (4)	# exp. (5)	Avg. (6)
ln (1 + Applied tariff)$_{n,\text{FR}}$	−4.82[a]	−2.81[a]	−2.01[a]	−3.27[a]	−2.18[a]	−1.08[a]
	(0.23)	(0.17)	(0.15)	(0.20)	(0.14)	(0.13)
ln (1 + Applied tariff)$_{n,\text{CN}}$	−3.68[a]	−1.69[a]	−1.99[a]	−2.07[a]	−1.15[a]	−0.92[a]
	(0.21)	(0.16)	(0.13)	(0.18)	(0.12)	(0.12)
ln (1 + Applied tariff)$_{k,\text{FR}}$	−6.69[a]	−5.81[a]	−0.88[a]			
	(0.29)	(0.16)	(0.20)			
ln (1 + Applied tariff)$_{k,\text{CN}}$	−4.48[a]	−2.17[a]	−2.31[a]			
	(0.22)	(0.15)	(0.13)			
R^2	0.347	0.588	0.080	0.348	0.593	0.080
rmse	2.49	1.04	2.10	2.49	1.04	2.10

Notes: 1,119,993 observations in all columns. [a], [b] and [c] denote statistical significance levels of 1, 5 and 10% respectively. Estimations in columns (1) to (3) include a year fixed effect and the four components (n, FR; n, CN; k, FR; and k, CN) of each gravity control (distance, common language, contiguity and colony). Estimations in columns (4) to (6) include reference country–year fixed effects and the two components (n, FR; n, CN) of each gravity control. In all estimations, standard errors are clustered by destination × reference country and year level.

other, supporting further heterogeneity in the trade elasticities. A related approach is to confine identification on the destination country, neutralizing the change of reference country with a k fixed effect. Those results are shown in columns (4) to (6), where again most of the tariff elasticities differ across origin countries.[31]

6. Industry-level analysis

We now investigate the implications of cross-industry heterogeneity for our analysis. We proceed with the ISIC rev3 2-digit industry classification, which has an easy match with the HS6 product classification used in trade and tariff data. We first provide micro-based theoretical predictions of the bilateral aggregate trade elasticities at the sectoral level and compare them to their empirical gravity estimates. We then exploit cross-industry variations in firm-level elasticity and bilateral aggregate elasticity to show that both demand and supply determinants enter the aggregate elasticity — a finding which discriminates in favor of the log-normal distributional assumption.

6.1. Comparing predicted and estimated sectoral elasticities

We start by computing for each sector the bilateral aggregate trade elasticity predicted by the Melitz (2003) model under a log-normal distribution of productivity. To do so, we follow the approach of Sections 4 and 5 conducting the full analysis for each sector p. More precisely, for each sector, we estimate the firm-level trade elasticity σ^p according to a p-specific version of Eq. (18):

$$\ln \tilde{x}^p_{\{j,n,k\}} = (1-\sigma^p)\ln\left(\widetilde{1+t^p_{\{n,k\}}}\right) + (1-\sigma^p)\delta \ln \tilde{D}_{\{n,k\}} + \ln \tilde{\epsilon}^p_{\{j,n,k\}}. \quad (26)$$

We then estimate a sector-specific version of the QQ-regression Eq. (21), $\mathbb{Q}^{\text{LN},p}_f = q^{\text{LN},p} + (\widehat{\sigma^p}-1)\nu^p\Phi^{-1}(\hat{F}^p_f)$, in order to retrieve ν^p, the standard deviation of the underlying log-normal productivity distribution (QQ estimates at the sector level are available upon request). As explained in Section 5.1, those estimates $\widehat{\sigma^p}$ and $\widehat{\nu^p}$ are combined according to p-specific versions of Eqs. (12), (13) and (14) to predict three sector-specific aggregate elasticities of trade with respect to trade costs (the reactions of total trade, number of exporters and average exports per firm to tariffs).

Column (1) of Table 6 runs a Tetrad estimation of Eq. (26) over 2000–2006 for each 2-digit ISIC industry separately including destination–reference country (as in column 6 of our benchmark results table) and year fixed effects. Each cell reports the coefficient on the applied tariffs with its associated degree of statistical significance (1%, 5%, 10%), which corresponds to the estimated firm-level trade elasticity by industry ($1-\widehat{\sigma^p}$). Averaging over significantly negative coefficients, we obtain an elasticity of −5.39. The median response is −4.71.

Each figure in column (1) provides the firm-level trade elasticity ($1-\widehat{\sigma^p}$), needed for the calculation of p-specific bilateral aggregate trade elasticities. We summarize those elasticities by reporting cross-sector cross-destination means and standard deviations in Table 7. These statistics can then be compared to the empirical

Table 6
Firm-level and aggregate elasticities by industry.

	Firm-level	Aggregate		
Dependent variable:	Exports (1)	Tot. exports (2)	# exporters (3)	Avg. exports (4)
Agriculture	−3.89[a]	−2.71[a]	−1.63[a]	−1.08
Food	−3.53[a]	−4.54[a]	−1.23[a]	−3.31[a]
Textile	−4.18[a]	−1.95[a]	−1.63[a]	−.32
Wearing apparel	−4.03[a]	−4.21[a]	−.43	−3.79[a]
Leather	−2.84[a]	−6.62[a]	−.27	−6.34[a]
Wood	−7.82[b]	−11.04[a]	−2.26[a]	−8.78[a]
Paper	9.76[a]	5.43[c]	6.93[a]	−1.51
Publishing	−1.64	4.33[c]	6.81[a]	−2.48
Chemicals	−3.38[a]	.6	.08	.52
Rubber and plastic	−5[a]	−5.55[a]	−.93[b]	−4.61[a]
Non-metallic	4.56[a]	11.24[a]	5.61[a]	5.63[a]
Basic metals	−4.73[b]	−3.75[b]	−1.46[a]	−2.29
Metals	−2.39[b]	−1.41[b]	1.87[a]	−3.28[a]
Non-electrical machinery	−4.71[a]	−3.14[a]	−.11	−3.04[a]
Office machinery	−8.48	−20.75[a]	−2.43[c]	−18.33[a]
Electrical machinery	−2.5	.55	2.19[a]	−1.63[b]
Equip. radio, TV	−7.5[a]	−7.17[a]	−1.67[a]	−5.51[a]
Instruments	−4.76[a]	−.5	1.91[a]	−2.42[a]
Motor vehicles	−9.25[a]	−4.7[c]	−1.81[c]	−2.89
Other transport	−10.26[a]	−17.42[a]	−1.07[c]	−16.35[a]
Furniture	−3.11[a]	.31	1.29[a]	−.98[c]
Average	−5.39	−5.71	−.97	−4.74
Median	−4.71	−4.54	−1.23	−3.31
Pooled	−3.48[a]	−3.37[a]	−1.75[a]	−1.62[a]

Notes: [a], [b] and [c] denote statistical significance levels of 1, 5 and 10% respectively. Estimations run at the ISIC 2 digit level, for years 2000 to 2006. All estimations include destination–reference country and year fixed effects. Standard errors are clustered by product–reference country combinations. The lines "Average" and "Median" take the mean and median values of coefficients obtained for the 13 industries where both the firm-level elasticity and aggregate trade elasticity are estimated to be significantly negative at the 5% level. The line "Pooled" is reporting coefficients obtained in a regression pooling over all those industries. The cells report the coefficient on the tetraded applied tariffs for each industry.

[31] In the Online Appendix, we present results from the same estimations run for the two extreme years of our sample, 2000 and 2006, with significant evidence of non-constant elasticities in most cases.

Table 7
Predicted bilateral aggregate trade elasticities (LN distribution).

	France	China	Average
Total flows	−5.492	−5.513	−5.502
	(.369)	(.451)	(.308)
Number of exporters	−1.117	−.8	−.958
	(.788)	(.971)	(.762)
Average flows	−4.375	−4.713	−4.544
	(.608)	(.66)	(.564)

Notes: This table presents the predicted elasticities (mean and standard deviation over destinations and sectors) on total exports, the number of exporting firms, and average export flows. The CES for each sector, σ^p, is obtained from the first column of Table 6. Unreported sector-specific QQ-regressions yield ν^p, the standard deviation of log-normal productivity distribution.

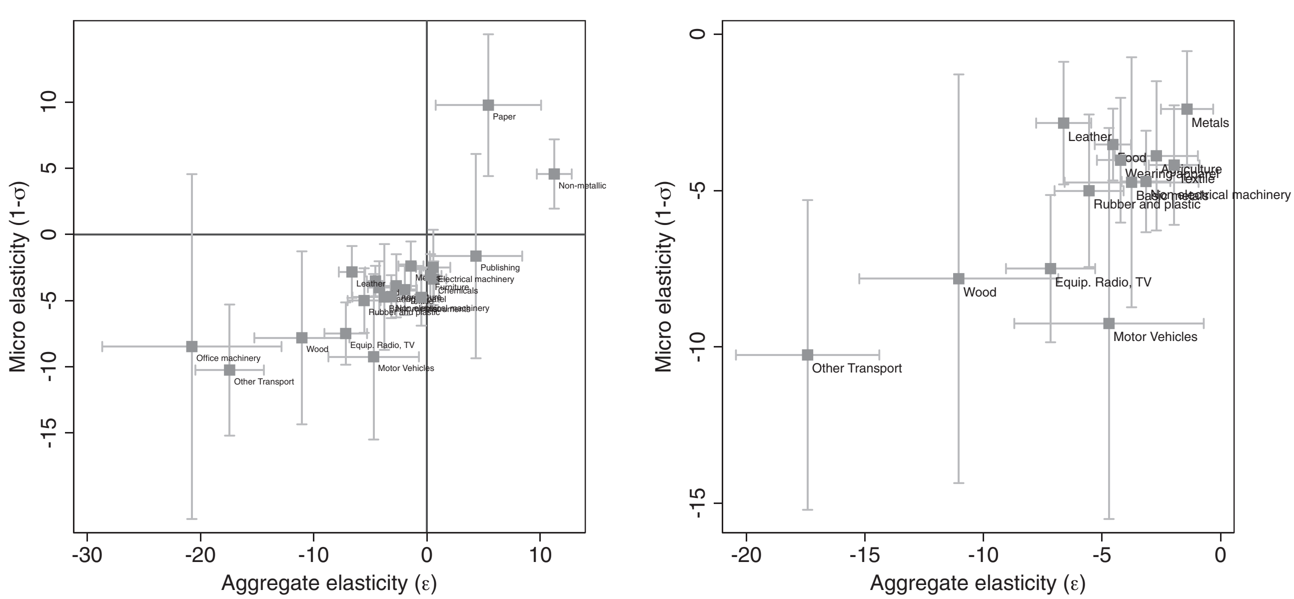
macro-based estimates of the same bilateral elasticities reported in columns (2), (3) and (4) of Table 6. Remarkably, for each of those three elasticities, the average micro-based theoretical prediction and the average macro-based estimate are quite close: −5.50 vs −6.04 for total export, −.96 vs −.97 for the number of exporter, −4.54 vs −4.74 for average exports. Another comparison can be made across sectors. For the 13 sectors with significantly negative joint estimates in Table 6, the pairwise correlations across sectors between predicted and estimated elasticities are large (respectively .77, .59, and .61). Overall, results from this sectoral analysis confirm the ones of Section 5.2: The micro-based theoretical predictions of the aggregate trade elasticities align reasonably well with their empirical macro-based estimates.

6.2. *Demand and supply side components of bilateral aggregate trade elasticity*

In this final section, we take advantage of our industry-level analysis to provide evidence that both demand and supply-side determinants enter the bilateral aggregate elasticity of total flows (ε_{ni}). Under the Pareto assumption, the firm-level elasticity and the aggregate elasticity have no reason to be correlated, since the firm-level elasticity is a measure of (inverse) product differentiation, while the aggregate one is capturing homogeneity in firms' productive efficiency (Chaney, 2008). Under alternative distributions like the log-normal, the aggregate elasticity includes both determinants and therefore should be correlated with the firm-level elasticity (see Eq. (1)). Fig. 6 reports graphical evidence that those correlations are large, both when considering all sectors (panel a), or when restricting attention to the 13 sectors with significantly negative joint estimates in Table 6 (panel b). This pattern exhibits overwhelming evidence in favor of the aggregate trade elasticity including demand side determinants.

7. Conclusion

We argue in this paper that knowledge of the firm-level response to trade costs is key for understanding the reaction of aggregate exports to trade costs. In other words, we need micro-level data to uncover the macro-level impact of trade costs, a central element in any trade policy evaluation. This need for micro-data is presumably true with the vast majority of possible distributional assumptions of firm-level productivity. There is one exception however where micro-data is not needed: The (unbounded) Pareto distribution. The literature has been concentrating on that exception for reasons of tractability that are perfectly legitimate. In particular, it maintains the simple log-linear gravity equation with a constant trade elasticity to be estimated with macro-data. However, the evidence presented in our paper points to systematic variation in bilateral aggregate trade elasticities that is both substantial and compatible with log-normal heterogeneity (which is also the assumption that best matches the micro-level distribution of export sales). We find in particular that the average values of bilateral aggregate trade elasticities predicted under a log-normal calibration are close to the empirical gravity estimates. By contrast, the Pareto-based calibration leads to predictions that seem invalidated by the data. Namely, the invariance of average exports per firm to ad-valorem tariff variations, the lack of correlation between firm-level and aggregate elasticities estimated industry by industry, and the constant aggregate trade elasticities. We are therefore tempted to call for a "micro-approach" to estimating those bilateral aggregate elasticities as opposed to the



(a) Full set of industries

(b) Industries with sig.negative coeffs.

Fig. 6. Country-pair and firm-level elasticities by industry. Note: The figure plots in each panel the firm-level elasticity against the aggregate elasticity estimated at the 2-digit industry level.

"macro-approach" that uses gravity specified so as to estimate a constant elasticity.

The micro- and macro-approaches differ substantially in several respects and both have positive and negative aspects. On the one hand, gravity is a more direct and parsimonious route for estimating aggregate elasticities: (i) parametric assumptions are reduced to a minimum while our micro-based procedure depends on the calibration of the productivity distribution; (ii) gravity is less demanding in terms of data and makes possible the use of easily accessible data sets of bilateral aggregate trade flows. On the down side, the log-linear specification of gravity is inconsistent with theory when dropping Pareto-heterogeneity. Also, gravity provides, for each origin country, only a cross-destination average of aggregate elasticities while the micro-based approach provides the full distribution of elasticities. Our Monte Carlo simulations show that gravity actually approximates this average of the true underlying bilateral aggregate elasticities quite decently. However, the gravity-predicted impact can be a bad approximation when trade liberalization occurs between countries that have either very low or very high levels of trade initially. When interested in policy experiments that involve this type of country pairs (distant and small countries or proximate and large ones for instance), one might strongly prefer the micro-approach.

Appendix A. Theoretical derivations of V, γ and $\mathcal{H}$ under Pareto and log-normal distributions

Our central Eq. (6) makes it clear that the heterogeneity of aggregate trade elasticity comes entirely from the term γ_{ni} that stems from endogenous selection of firms into export markets (see Eq. (7)). In turn, γ_{ni} depends on the cost-performance index V_{ni} as defined by

$$V_{ni} \equiv \int_0^{a^*_{ni}} a^{1-\sigma} g(a) da.$$

We therefore need to understand how these γ and V terms behave under alternative distributional assumptions on productivity, i.e. Pareto and log-normal. One important advantage of Pareto, pointed out by Redding (2011) is that if φ is distributed Pareto(θ) then φ^r is also distributed Pareto. The shape parameter becomes θ/r. This advantage is shared by the log-normal. If φ is log-$\mathcal{N}(\mu, \sigma^2)$ then φ^r is log-$\mathcal{N}(r\mu, r^2\sigma^2)$. This follows from a more general reproductive property reported by (Bury, 1999, p. 156) and recently used in Mrázová et al. (2016).

If productivity is Pareto, then the rescaled unit input requirement a has PDF $g(a) = \theta a^{\theta-1}/\bar{a}^\theta$, which translates into

$$V^{\mathrm{P}}_{ni} = \frac{\theta a^{*\theta-\sigma+1}_{ni}}{\bar{a}^\theta(\theta-\sigma+1)}. \tag{A.1}$$

The elasticity of V^{P}_{ni} with respect to a^* is

$$\gamma^{\mathrm{P}}_{ni} = \theta - \sigma + 1 > 0. \tag{A.2}$$

Hence, Pareto renders γ_{ni} constant, i.e. not depending on n nor i, and therefore transforms an expression generally yielding heterogeneous trade elasticities into a one-parameter elasticity $\frac{d\ln X_{ni}}{d\ln \tau_{ni}} = \theta$, that is related to the supply side of the economy only.

When productive efficiency is distributed log-normally,[32] things are very different. For φ ~log-$\mathcal{N}(\mu, \nu)$, the distribution of rescaled unit input requirements is a ~log-$\mathcal{N}(-\mu, \nu)$. (Jawitz, 2004, Table 1) expresses m_r, the absolute rth truncated moment in terms of the error function (erf). We convert his expression to be in terms of the more familiar, $\Phi()$, the CDF of the standard normal, using the relationship $\mathrm{erf}(x) = 2\Phi(x\sqrt{2}) - 1$. For $x \sim$ log-$\mathcal{N}(\mu, \nu)$ truncated between lower limit ℓ and upper limit u the Jawitz formula can be expressed as

$$m_r = \exp(r\mu + r^2\nu^2/2)\left[\Phi\left(\frac{\ln u - \mu - r\nu^2}{\nu}\right) - \Phi\left(\frac{\ln \ell - \mu - r\nu^2}{\nu}\right)\right].$$

We are considering the distribution of α which is the inverse of productivity so it has distribution log-$\mathcal{N}(-\mu, \nu)$ and it has a lower limit $\ell = 0$ and an upper limit $u = \alpha^*$. The limit of $\Phi(x)$ as $x \to -\infty$ is zero so the second term involving $\ln\ell$ disappears. Replacing μ with $-\mu$ and r with $1-\sigma$, we obtain:

$$V^{\mathrm{LN}}_{ni} = \exp\left[(\sigma-1)\mu + (\sigma-1)^2\nu^2/2\right]\Phi\left[(\ln a^*_{ni} + \mu)/\nu + (\sigma-1)\nu\right], \tag{A.3}$$

Differentiating $\ln V^{\mathrm{LN}}_{ni}$ with respect to $\ln a^*_{ni}$,

$$\gamma^{\mathrm{LN}}_{ni} = \frac{1}{\nu} h\left(\frac{\ln a^*_{ni} + \mu}{\nu} + (\sigma-1)\nu\right), \tag{A.4}$$

where $h(x) \equiv \phi(x)/\Phi(x)$, the ratio of the PDF to the CDF of the standard normal. Thus, γ_{ni} is no longer the constant which obtains for productivity distributed Pareto. Bilateral elasticities write as

$$\varepsilon^{\mathrm{P}}_{ni} = -\theta, \qquad \text{and} \qquad \varepsilon^{\mathrm{LN}}_{ni} = 1 - \sigma - \frac{1}{\nu} h\left(\frac{\ln a^*_{ni} + \mu}{\nu} + (\sigma-1)\nu\right). \tag{A.5}$$

The $\mathcal{H}$ function is a central element of our quantification exercise, as summarized by relationship Eq. (9), that reveals cutoffs and therefore aggregate bilateral elasticities. Comparing Eqs. (4), (7) and (8), we see that $\mathcal{H}$ and γ are closely related

$$\gamma_{ni} \times \mathcal{H}\left(a^*_{ni}\right) = a^*_{ni}\frac{g\left(a^*_{ni}\right)}{G\left(a^*_{ni}\right)}. \tag{A.6}$$

With Pareto, we make use of Eq. (A.2) to obtain

$$\mathcal{H}^{\mathrm{P}}\left(a^*_{ni}\right) = \frac{\theta}{\theta - \sigma + 1}\quad. \tag{A.7}$$

With a log-normal productivity, Eq. (A.4) leads to

$$\mathcal{H}^{\mathrm{LN}}\left(a^*_{ni}\right) = \frac{h\left[(\ln a^*_{ni} + \mu)/\nu\right]}{h\left[(\ln a^*_{ni} + \mu)/\nu + (\sigma-1)\nu\right]}. \tag{A.8}$$

Table A.1 summarizes all formulas for the variables used in this paper under both distributions.

An attractive feature of our quantification exercise relates to the small number of relevant parameters to be calibrated. Under Pareto, Eqs. (A.2) and (A.7) show that only the shape parameter θ matters. Similarly, under a log-normal, only the calibration of the second-moment of the distribution, ν, is necessary for inverting the $\mathcal{H}$ function to reveal the cutoff and for quantifying the aggregate elasticity: This last point stems from the fact that shifting the first moment, μ, affects Eqs. (A.4) and (A.8) in an identical way and so has no impact on the quantification.

[32] There are a number of useful properties of the normal distribution that we use here:

1. $\Phi(-x) = 1 - \Phi(x)$
2. $\phi(-x) = \phi(x)$
3. $\phi'(x) = -x\phi(x)$
4. $\Phi'(x) = \phi(x)$
5. $\partial(x\Phi(x) + \phi(x))/\partial x = \Phi(x) > 0$

Table A.1
Pareto vs log-normal: key variables.

Variable	Pareto	Log-normal
PDF: $g(a)$	$\frac{\theta a^{\theta-1}}{\bar{a}^{\theta}}$	$\phi\left(\frac{\ln a+\mu}{\nu}\right)/a\nu$
CDF: $G(a)$	$\frac{a^{\theta}}{\bar{a}^{\theta}}$	$\Phi\left(\frac{\ln a+\mu}{\nu}\right)$
$V_{ni}(a^*) \equiv \int_0^{a^*} a^{1-\sigma} g(a)da$	$\frac{\theta a_{ni}^{*\theta-\sigma+1}}{\bar{a}^{\theta}(\theta-\sigma+1)}$	$\exp\left[(\sigma-1)\mu+\frac{(\sigma-1)^2\nu^2}{2}\right]\Phi\left[\frac{(\ln a_{ni}^*+\mu)}{\nu}+(\sigma-1)\nu\right]$
$\gamma_{ni} \equiv \frac{d\ln V_{ni}}{d\ln a_{ni}^*}$	$\theta-\sigma+1$	$\frac{1}{\nu}h\left(\frac{\ln a_{ni}^*+\mu}{\nu}+(\sigma-1)\nu\right)$
$\frac{d\ln X_{ni}}{d\ln \tau_{ni}} = 1-\sigma-\gamma_{ni} = \epsilon_{ni}$	$-\theta$	$1-\sigma-\frac{1}{\nu}h\left(\frac{\ln a_{ni}^*+\mu}{\nu}+(\sigma-1)\nu\right)$
$\mathcal{H}(a^*) \equiv \frac{V(a^*)}{a^{*1-\sigma}G(\alpha^*)}$	$\frac{\theta}{\theta-\sigma+1}$	$\frac{h[(\ln a_{ni}^*+\mu)/\nu]}{h[(\ln a_{ni}^*+\mu)/\nu+(\sigma-1)\nu]}$

Note: The Pareto parameters of the unit input requirement distributions are θ and $\bar{\alpha}$. For the log-normal distribution, when $\varphi \sim$log-$\mathcal{N}(\mu,\nu)$, the distribution of rescaled unit input requirements is $a \sim$log-$\mathcal{N}(-\mu,\nu)$. We define $h(x) \equiv \phi(x)/\Phi(x)$, a non-increasing function.

Appendix B. Predicted/estimated elasticities: Monte Carlo evidence

In Section 5.2, we find that the macro-based estimate of the aggregate trade elasticity is quantitatively close to the average micro-based theoretical prediction when heterogeneity is calibrated as being log-normal. We interpret this finding as empirical support in favor of this distributional assumption. In this Appendix, we substantiate this last statement by embracing a more theoretical perspective. This is an important step in the argument because the theoretical relationship between the macro-based estimates and the micro-based predictions of the elasticities is unknown (except under Pareto where they are unambiguously equal). Hereafter, we provide simulation-based evidence that the similarity between micro-based predictions and macro-based estimates is not accidental, even under log-normal heterogeneity.

We proceed with Monte Carlo (MC) simulations of our generic trade model with heterogeneous firms. In the baseline simulations, we generate fake bilateral trade for 10 countries and 1 million active firms per country. Our data generating process uses the firms' sales in Eq. (2). Firm-level heterogeneity in terms of rescaled labor requirement, $a \equiv \alpha \times b(\alpha)$ is assumed to be Pareto or log-normally distributed with a set of parameters identical to the ones used in our empirical analysis (Section 5.1). We also retain $\sigma = 5$ as the parameter for the firm-level trade elasticity. Without loss of generality, in this partial equilibrium framework, we normalize the nominal wage, $w = 1$, and we draw A_n/f_{ni}, i.e. the ratio of destination n attractiveness over entry cost from a log-normal distribution. This distribution is calibrated such as to match an average bilateral share of exporting firms of 18%, an empirical moment that is also targeted by Head et al. (2014) and Melitz and Redding (2015) and that reflects the average fraction of US manufacturing firms that export (as reported in Bernard et al., 2007). Finally, the applied-tariffs $\tau_{ni} = 1 + t_{ni}$ are drawn from an uniform distribution over the range $[1, 2]$.

In each MC draw, we first generate a matrix of firm-level trade flows that are non-zero when sales exceed the bilateral entry cost, i.e. $x_{ni}(a) > \sigma w f_{ni}$. In a first stage we infer from this fake trade data set the micro-based predictions of the aggregate trade elasticities by applying the methodology of Section 5.1: We first retrieve mean-to-min ratios for all country-pairs and then predict the corresponding set of theoretical bilateral elasticities (Eqs. (10) and 11)). In a second stage, we turn to the macro-based estimates of the trade elasticity. To this purpose, we collapse firm-level trade flows at the country-pair level to construct a matrix of bilateral aggregate trade. We then run gravity regressions (both using country fixed effects and Tetrads) and retrieve the point estimate of applied tariffs. Hence, for each draw, we obtain one macro-based estimate of the trade elasticity that we compare to the average of the micro-based predictions. This procedure is replicated 1000 times.

The simulation results are displayed in Table A.2 for log-normal (col. 1–col. 4) and Pareto (col. 5–col. 8) and for different degrees of firm scarceness (from 1000 to 1 million firms per country). Each column reports averages and standard errors across replications.

Our baseline simulation under Pareto (col. 8) shows that the simulated economy with 1 million firms conforms to the theoretical prediction of a model with a continuum of firms. The micro-based predictions of the bilateral aggregate trade elasticity are relatively homogeneous across country pairs (the number in parenthesis reports the mean value of the standard deviation within each draw, and is quite small compared to the average elasticity) and their average (first row) is close to the macro-based estimates of the elasticities retrieved from Tetrad-like specification (second row) or standard gravity with fixed effects (third row). Finally, the elasticity of the average export (last row) is not significantly different from zero, as

Table A.2
Monte Carlo results: elasticities wrt to a change in trade costs.

Distribution:	Log-normal				Pareto			
# firms per country:	1 K	10 K	100 K	1 M	1 K	10 K	100 K	1 M
Total exports (micro-predictions)	−4.64	−4.51	−4.48	−4.47	−5.77	−5.50	−5.28	−5.20
	(1.08)	(0.66)	(0.54)	(0.53)	(0.95)	(0.77)	(0.33)	(0.17)
Total exports (macro/Tetrads)	−4.42	−4.56	−4.53	−4.53	−5.60	−5.33	−5.31	−5.23
	(0.49)	(0.26)	(0.09)	(0.03)	(0.95)	(0.59)	(0.42)	(0.27)
Total exports (macro/FE)	−4.40	−4.44	−4.43	−4.43	−5.51	−5.32	−5.27	−5.20
	(0.20)	(0.09)	(0.04)	(0.01)	(0.35)	(0.23)	(0.15)	(0.10)
Nb exporters (macro/FE)	−2.63	−2.72	−2.72	−2.71	−5.12	−5.09	−5.15	−5.14
	(0.07)	(0.04)	(0.01)	(0.00)	(0.20)	(0.08)	(0.04)	(0.01)
Avg. exports (macro/FE)	−1.77	−1.72	−1.71	−1.71	−0.43	−0.23	−0.13	−0.06
	(0.18)	(0.09)	(0.03)	(0.01)	(0.50)	(0.21)	(0.14)	(0.10)

Notes: 1000 replications for each cell, parameters on fixed costs of exports and size of the demand term have been calibrated so the share of exporters averages to 18% in all simulations. For each elasticity, the first line reports the average value. Standard deviations are in parentheses. For the “micro-predictions” elasticity, the number in parentheses is the average of standard deviations of the elasticity in each draw (quantifying the degree of heterogeneity in bilateral elasticities). For the aggregate elasticities, we report the standard deviation of elasticities across the 1000 replications.

expected from the theoretical prediction associated with Pareto heterogeneity and a continuum of firms. An important finding is that with small numbers of potential exporters, this elasticity gets larger (in absolute value), but remains a small share of the elasticity on total exports, and is never statistically different from 0, even with only 1000 firms (of which around 18% export). We conclude from this exercise that scarceness (sometimes referred to as granularity) does not seem to play a central role in our results (which is related to the findings of Fernandes et al., 2015).

From the baseline simulation under log-normal (column 4), we see that the macro-based estimate of the aggregate elasticity and the average of the micro-based predictions are quantitatively very close — i.e. equality cannot be rejected. This constitutes the main result of our Monte Carlo approach. It confirms that the similarity between micro-based predictions and the macro-based estimates in Section 5.2 can be safely interpreted as supportive of the log-normal distribution. Notice that the magnitude of the simulation results on the three bilateral trade elasticities (total exports, count of exporters and average exports per firm) is also close to what we obtain with the sample of French and Chinese firms. This is remarkable given that our Monte Carlo approach is minimal and shares only few features with the true data, i.e. the parameters of firm-level heterogeneity and the share of exporters.

Appendix C. Rank correlation in Section 4.1

In Section 4.1, we calibrate the relative contribution of α and $b(\alpha)$ using the correlation between the rank of a firm in total exports of the country and its rank in sales to country n. This Appendix aims at documenting in more details the predictions of our model in terms of those rank correlations. For a given destination market n, the *relative rank* in the unconstrained sales distribution of two firms α and α' from origin country i is an increasing step-function of the (log of the) ratio of their sales. Let us denote this log of sales ratio as $RR_{ni}(\alpha, \alpha') \equiv ln(x_{ni}(\alpha)/x_{ni}(\alpha'))$. Note that by construction, RR is equal to zero when firms have exactly the same rank. Eq. (2) yields

$$RR_{ni}(\alpha, \alpha') = (1-\sigma)(\ln\alpha - \ln\alpha') - (1-\sigma)(\ln b_{ni}(\alpha) - \ln b_{ni}(\alpha')). \quad (A.9)$$

The previous formula makes clear that, for a given pair of firms, the cross-market variations in $RR_{ni}(\alpha, \alpha')$ are driven by the second component only. In particular we see that the larger is the variance of $\ln b(\alpha)$ with respect to the variance of $\ln\alpha$, the more we should observe changes in the relative rank of α and α' across markets. When the variance of $\ln b(\alpha)$ goes to zero (infinite), the cross-market correlation in $RR_{ni}(\alpha, \alpha')$ approaches unity (zero).

The previous discussion relates to rank correlations in the unconstrained sales distribution. We must now take into account selection into export and this feature makes the analysis more complex. Moreover, obtaining closed-form results on rank statistics is challenging because of the non-differentiability of the step function. For these two reasons we now rely on Monte Carlo simulations to illustrate further how rank correlation is used for calibrating the relative contribution of α and $b(\alpha)$.

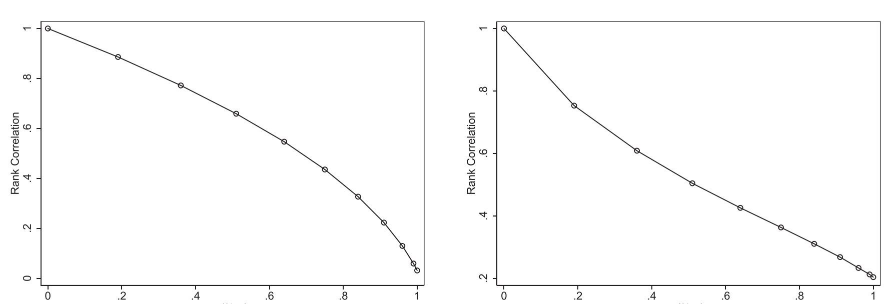
We simulate our model for different values of r, namely the ratio of the variance of $\ln b(\alpha)$ over the variance of $\ln\alpha$ (other parameters being unchanged). Fig. A.1 reports the correlation between the rank of a firm in total exports of the country and its rank in sales to each destination market n (vertical axis) for different values of $r/(1+r)$ (horizontal axis). The left panel displays rank statistics for the unconstrained sales distribution. The right panel considers the conditional sales distribution, i.e. when selection into export takes place. In that case all non-exporters have zero sales and so only exporting firms are ranked. In both cases, as expected, we see a negative relationship. Under selection the limited number of destinations prevents rank correlation to attain zero for infinite r. Crucial for the calibration is the fact that the relationship is monotonic. And we retain a value of r such as to match the empirical moment of 66% for the rank correlation obtained with conditional sales (right panel).

Appendix D. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jinteco.2017.05.001.

References

Anderson, J.E., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. Am. Econ. Rev. 93 (1), 170–192.
Arkolakis, C., 2010. Market penetration costs and the new consumers margin in international trade. J. Polit. Econ. 118 (6), 1151–1199.
Arkolakis, C., Costinot, A., Rodríguez-Clare, A., 2010. Gains from trade under monopolistic competition: a simple example with translog expenditure functions and Pareto distributions of firm-level productivity. Manuscript.
Arkolakis, C., Costinot, A., Rodriguez-Clare, A., 2012. New trade models, same old gains? Am. Econ. Rev. 102 (1), 94–130.
Atkeson, A., Burstein, A., 2008. Pricing-to-market, trade costs, and international relative prices. Am. Econ. Rev. 98 (5), 1998–2031.
Baier, S., Bergstrand, J.H., 2001. The growth of world trade: tariffs, transport costs, and income similarity. J. Int. Econ. 53 (1), 1–27.
Barba Navaretti, G., Bugamellli, M., Forlani, E., Ottaviano, G., 2015. It takes (more than) a moment: revisiting the link between firm productivity and aggregate exports. University of Milan Mimeo.
Berman, N., Martin, P., Mayer, T., 2012. How do different exporters react to exchange rate changes? Q. J. Econ. 127 (1), 437–492.
Bernard, A., Jensen, J., Redding, S., Schott, P., 2007. Firms in international trade. J. Econ. Perspect. 21 (3), 105–130.
Berthou, A., Fontagné, L., 2016. Variable trade costs, composition effects and the intensive margin of trade. World Econ. 39 (1), 54–71.
Broda, C., Weinstein, D., 2006. Globalization and the gains from variety. Q. J. Econ. 121 (2), 541–585.
Bury, K., 1999. Statistical Distributions in Engineering. Cambridge University Press.



(a) unconstrained sales (b) selection into exports

Fig. A.1. Rank correlations. Notes: The figure reports the correlation between the rank of a firm in total exports of the country and its rank in sales to each destination market n for 10 different values of $r/(1+r)$ where r is the ratio of the variance of $\ln b(\alpha)$ over the variance of $\ln\alpha$. Each cell is equal to the average rank correlation across 1000 Monte Carlo replications (standard deviations are small and unreported). In the left panel, the sales distribution is unconstrained. In the right panel, there is selection into export.

Caliendo, L., Parro, F., 2015. Estimates of the trade and welfare effects of NAFTA. Rev. Econ. Stud. 82 (1), 1–44.
Chaney, T., 2008. Distorted gravity: the intensive and extensive margins of international trade. Am. Econ. Rev. 98 (4), 1707–1721.
Chatterjee, A., Dix-Carneiro, R., Vichyanond, J., 2013. Multi-product firms and exchange rate fluctuations. Am. Econ. J. Econ. Pol. 5 (2), 77–110.
Costinot, A., Donaldson, D., Komunjer, I., 2012. What goods do countries trade? A quantitative exploration of Ricardo's ideas. Rev. Econ. Stud. 79 (2), 581–608.
Crozet, M., Head, K., Mayer, T., 2012. Quality sorting and trade: firm-level evidence for french wine. Rev. Econ. Stud. 79 (2), 609–644.
Eaton, J., Kortum, S., 2001. Trade in capital goods. Eur. Econ. Rev. 45 (7), 1195–1235.
Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70 (5), 1741–1779.
Eaton, J., Kortum, S., Kramarz, F., 2011. An anatomy of international trade: evidence from French firms. Econometrica 79 (5), 1453–1498.
Feenstra, R.C., 1994. New product varieties and the measurement of international prices. Am. Econ. Rev. 84, 157–177.
Feenstra, R.C., 2003. A homothetic utility function for monopolistic competition models. Econ. Lett. 78 (2), 79–86.
Feenstra, R.C., 2013. Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. UC Davis Mimeo.
Fernandes, A., Klenow, P., Meleshchuk, S., Pierola, M.D., Rodriguez-Clare, U.A., 2015. The intensive margin in trade: moving beyond Pareto. Mimeo.
Fitzgerald, D., Haller, S., 2015. Exporters and Shocks: Dissecting the International Elasticity Puzzle. Manuscript.
Freund, C., Pierola, M.D., 2015. Export superstars. Rev. Econ. Stat.
Gopinath, G., Neiman, B., 2014. Trade adjustment and productivity in large crises. Am. Econ. Rev. 104 (3), 793–831.
Greene, W., 2004. Fixed effects and bias due to the incidental parameters problem in the Tobit model. Econ. Rev. 23 (2), 125–147.
Guimarães, P., Portugal, P., 2010. A simple feasible procedure to fit models with high-dimensional fixed effects. Stata J. 10 (4), 628–649. (22).
Hallak, J.C., 2006. Product quality and the direction of trade. J. Int. Econ. 68 (1), 238–265.
Head, K., Mayer, T., 2014. Gravity equations: workhorse, toolkit, and cookbook. In: Helpman, E., Gopinath, G., Rogoff, K. (Eds.), Handbook of International Economics. vol. 4. Elsevier.,
Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. J. Int. Econ. 81 (1), 1–14.
Head, K., Mayer, T., Thoenig, M., 2014. Welfare and trade without Pareto. Am. Econ. Rev. 104 (5), 310–316.
Head, K., Ries, J., 2001. Increasing returns versus national product differentiation as an explanation for the pattern of US–Canada trade. Am. Econ. Rev. 91 (4), 858–876.
Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: trading partners and trading volumes. Q. J. Econ. 123 (2), 441–487.
Hummels, D., 1999. Towards a Geography of Trade Costs. Technical report 17, GTAP Working Paper,
Imbs, J., Méjean, I., 2015. Elasticity optimism. Am. Econ. J. Macroecon. 7 (3), 43–83.
Jawitz, J.W., 2004. Moments of truncated continuous univariate distributions. Adv. Water Resour. 27, 269–281.
Krugman, P., 1980. Scale economies, product differentiation, and the pattern of trade. Am. Econ. Rev. 70, 950–959.
Li, H., Ma, H., Xu, Y., 2015. How do exchange rate movements affect Chinese exports?—A firm-level investigation. J. Int. Econ. 97 (1), 148–161.
Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71, 1695–1725.
Melitz, M., Ottaviano, G., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75 (1), 295–316.
Melitz, M.J., Redding, S.J., 2015. New trade models, new welfare implications. Am. Econ. Rev.
Mrázová, M., Neary, J.P., Parenti, M., 2016. Sales and markup dispersion: theory and empirics.
Mulligan, C.B., Rubinstein, Y., 2008. Selection, investment, and women's relative wages over time. Q. J. Econ. 1061–1110.
Nigai, S., 2017. A tale of two tails: productivity distribution and the gains from trade. J. Int. Econ. 104, 44–62.
Novy, D., 2013. International trade without CES: estimating translog gravity. J. Int. Econ. 89 (2), 271–282.
Ossa, R., 2015. Why trade matters after all. J. Int. Econ.
Paravisini, D., Rappoport, V., Schnabl, P., Wolfenzon, D., 2015. Dissecting the effect of credit supply on trade: evidence from matched credit–export data. Rev. Econ. Stud. 82 (1), 333–359.
Redding, S.J., 2011. Theories of heterogeneous firms and trade. Annu. Rev. Econ. 3, 77–105.
Romalis, J., 2007. NAFTA's and CUSFTA's impact on international trade. Rev. Econ. Stat. 89 (3), 416–435.
Simonovska, I., Waugh, M.E., 2011. The Elasticity of Trade: Estimates and Evidence. Working Paper 16796, NBER.,
Spearot, A.C., 2013. Variable demand elasticities and tariff liberalization. J. Int. Econ. 89 (1), 26–41.
Yang, M.-J., 2014. Firm Heterogeneity and Trade in a Distorted Gravity World. University of Washington Mimeo.