Gains From Trade: Demand, Supply and Idiosyncratic Shocks

Ruben Dewitte, Bruno Merlevede and Glenn Rayp*

12th May 2022

Abstract

Firm-level sales are often used as a proxy for productivity to quantify welfare Gains from Trade (GFT) using firm-level data. This approach ignores the existence of transitory idiosyncratic shocks and heterogeneity other than productivity in firm-level sales. We demonstrate, theoretically and empirically, that a productivity measure proxied by firm-level sales conflates at least two heterogeneity sources: persistent productivity and transitory shocks to demand and supply. Conflating transitory shocks with productivity results in an over-dispersed distribution of productivity. Assigning this shock-inflated productivity to the modeled economy's supply-side results in overestimated GFT. We show how to obtain unbiased productivity estimates, aggregate trade elasticities, and GFT estimates by exploiting the revenue production function from a single source country.

Keywords: Productivity distribution, Trade Elasticity, Gravity, Gains From Trade

JEL Codes: L11, F11, F12

*Dewitte: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be, Merlevede: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, bruno.merlevede@ugent.be, Rayp: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, glenn.rayp@ugent.be. We thank Luca David Opromolla, Gonzague Vannoorenberghe, Gerdie Everaert, Marijn Verschelde, Angelos Theodorakopoulos and participants of the 2017 ETSG conference in Florence, the 2018 FEB research day in Ghent, the 2018 Compnet conference in Halle and the 2019 ETSG conference in Bern for their helpful comments and suggestions. Ruben Dewitte acknowledges financial support from Research Foundation Flanders (FWO) grant 12B8822N.

1 Introduction

Quantifying Gains From Trade (GFT) has gained importance in recent years. As globalization and trade agreements attract increasing scrutiny, it becomes all the more important to provide policymakers with reliable information on the effects of trade across countries, industries and firms. In this light, the development of trade models that allow for heterogeneity at the firm level brought trade theory much closer to businesses and policy-makers (Cernat, 2014). In such models, a correct measurement of the aggregate trade elasticity, i.e. the response of aggregate trade flows to a change in trade costs, is paramount to obtain a correct evaluation of the effects of trade liberalization (Chaney, 2008; Arkolakis et al., 2012; Bas et al., 2017).

To calculate aggregate trade elasticities, one requires an approximation of the complete firm-level productivity distribution (Melitz and Redding, 2015; Bas et al., 2017). As firm-level productivity is usually unobserved by the researcher, the trade literature tends to rely on firm-level sales a a proxy for productivity (Bas et al., 2017; Nigai, 2017; Bee and Schiavo, 2018; Sager and Timoshenko, 2019). In doing so, however, the literature fails to account for idiosyncratic shocks and heterogeneity other than productivity that might be captured by firm-level sales (di Giovanni et al., 2011; Amand and Pelgrin, 2016).[1]

This paper demonstrates that a productivity measure proxied by firm-level sales conflates at least two heterogeneity sources: persistent productivity and transitory shocks to demand and supply. The difference between both matters: whether a firm's competitive (dis)advantage is transitory (for instance, due to an unexpected cyber attack) or persistent (for instance, due to its geographical location and/or industrial affiliation) will affect its profit maximizing decisions. We investigate, theoretically and empirically, the bias in the aggregate trade elasticity and subsequent GFT calculations that arises from the presence of transitory idiosyncratic shocks in current measures of firm-level productivity. The premise is straightforward: *conflating transitory shocks to demand and supply with firm-level productivity results in an over-dispersed productivity distribution. This is because persistent productivity and transitory shocks are variance independent, meaning that the variance of their sum equals the sum of their variances. As a higher productivity dispersion implies higher trade elasticities (i.e., less elastic trade) and GFT (Chaney, 2008; Head et al., 2014), relying on shock-inflated productivity will result in overestimated trade elasticities and GFT.* We show how to obtain unbiased productivity estimates, aggregate trade elasticity, and GFT estimates by exploiting the revenue production function from a single source country.

Idiosyncratic shocks may bias aggregate trade elasticities and GFT through two channels: (i) mismeasurement of firm-level productivity and (ii) theoretical model misspecification. We provide a general framework to identify and evaluate the importance of both channels, demonstrating that only the mismeasurement of firm-level productivity matters. First, we rely on an open economy heterogeneous firms model (Melitz, 2003) augmented with idiosyncratic shocks to

[1] A similar argument relates to the deterministic bias in firm efficiency or productivity obtained from non-parametric data envelopment analysis or free disposable hull-estimations, see for instance (Van Biesebroeck, 2007; Sickles and Zelenyuk, 2019) and to the differentiation between persistent and transient inefficiency in stochastic frontier models (Tsionas and Kumbhakar, 2014).

demand and supply (Das et al., 2007; De Loecker, 2011; Kasahara and Lapham, 2013; Gandhi et al., 2020) to demonstrate how firm-level sales conflate heterogeneity in productivity with transitory shocks to demand and supply. This approach, therefore, overestimates the variance of firm-level productivity.[2] Second, we argue that transitory shocks to demand and supply result in biased aggregate trade statistics due to firm-level productivity mismeasurement, not due to model misspecification. Under standard assumptions regarding the distribution of transitory shocks, modeled aggregate trade statistics are equivalent to those obtained from prevalent heterogeneous firms models that do not feature these shocks (Melitz, 2003). The intuition for this equivalence result is as follows. Transitory shocks and the economic impact of transitory shocks are not influenced by exogenous developments such as a change in trade costs. Therefore, transitory shocks cancel out when evaluating the modeled impact of a change in trade costs. This impact will be equivalent to the outcome from a model that does not feature transitory shocks to demand and supply.

We then propose a theoretically underpinned identification strategy exploiting panel data on the production of a single source country to obtain unbiased productivity estimates and identify the aggregate trade elasticity. We combine a firm-level production function with a CES demand system into a revenue production function to identify the individual components of the aggregate trade elasticity, the demand-side elasticity of substitution between varieties and the supply-side distribution of productivity (Chaney, 2008; Bas et al., 2017), while controlling for transitory shocks to demand and supply. We rely on structural production function estimation techniques (Klette and Griliches, 1996; De Loecker, 2011; Ackerberg et al., 2015) to control for endogeneity concerns and to obtain consistent parameter estimates. In contrast, existing identification frameworks for the aggregate trade elasticity require cross-sectional firm-level export sales (Bas et al., 2017) or quantity (Sager and Timoshenko, 2020) data of at least two countries. Moreover, the reliance on cross-sectional data implies that these frameworks can not differentiate between persistent and transitory shocks.

We evaluate the influence of transitory shocks in supply and demand using French firm-level data over the years 1998–2006. We find that the variance of productivity increases by approximately 10% when productivity is conflated with idiosyncratic shocks. The impact of this residual is not homogeneously distributed, but is larger in the tails of the distribution. This results in an absolute aggregate trade elasticity estimate, when not controlling for transitory shocks, that is overestimated by about 10.9% in foreign markets where 25% of the domestic firms would be active, and increases as foreign markets become more difficult to reach. The trade elasticity is overestimated by about 12.1% in foreign markets where 10% of the domestic firms would be active. Bas et al. (2017) demonstrate that the majority of export markets have a probability of exporting smaller than 10%. GFT, then, calculated as a shift from autarky to variable trade costs of $\tau = 1.96$ in a stylized symmetric 2-country model, is overestimated with about 20.45% when transitory shocks are not controlled for. GFT from autarky to iceberg trade costs of $\tau = 2.2$, are overestimated by 36.24%. These large differences in GFT for different values

[2]We adhere to to the definition of productivity set out in the (Melitz, 2003)-model to denote productivity as the persistent unexplained variation in output which determines firm-level input choices. Transitory shocks to demand and supply, then, are defined as the transitory component of unexplained variation in output.

of the iceberg trade costs can be attributed to the distribution of idiosyncratic shocks. As transitory shocks mainly distort the tails of the firm-level productivity distribution, a larger bias in aggregate trade statistics will be observed when foreign markets are less accessible, i.e. when exporting cutoffs are located more in the tails of this distribution. Therefore, we emphasize the importance of evaluating the relative differences in GFT for this stylized model. Overall, we find conclusive evidence that controlling for transitory idiosyncratic shocks is economically relevant when calculating the impact of trade costs on trade flows and welfare.

The remainder of the paper is organized as follows. In section 2, we provide an overview of the related literature. We present our theoretical framework in section 3. This framework allows us to define the identification strategy in section 4 and apply this strategy to French firm-level data. Section 5 evaluates the impact of transitory shocks on aggregate trade elasticities and GFT. Section 7 concludes.

2 Literature review

As stated in the introduction, the aggregate trade elasticity is identified once its individual components, the demand-side elasticity of substitution between varieties and the supply-side productivity distribution, are determined (Bas et al., 2017). The identification of these components is burdened by the existence of transitory idiosyncratic shocks to demand and supply. Below, we discuss the prevalent approach of (i) recovering the supply-side productivity distribution parameters from the sales distribution and (ii) identifying the elasticity of substitution from firm-level gravity, in light of the existence of transitory shocks.

On the supply side, identifying the productivity distribution parameters is difficult as firm-level productivity is unobservable to the researcher. The trade literature, therefore, resorts to *sales as a proxy for productivity* (see, for instance Head et al. (2014); Nigai (2017); Bas et al. (2017)). Under the assumptions of the dominant heterogeneous firms model (Melitz, 2003) with productivity following a distribution that is closed under power-law transformations,[3] it can be shown there is an approximate one-to-one mapping between sales (x) and productivity (ω): $x \dot{\sim} e^{(\sigma-1)\omega}$, up to the elasticity of substitution (σ).[4] Acknowledging the existence of transitory firm-level shocks and measurement error, however, productivity can only be identified from firm-level sales up to an independent firm-level stochastic component (ε^T): $x \dot{\sim} e^{(\sigma-1)\omega+\varepsilon^T}$ which increases the overall variance of productivity.[5] This component is not expected to be negligible. Kasahara and Lapham (2013) show that the residual (e^{ε^T}) accounts for approximately 10% of the variation in firm-level sales vis-à-vis productivity. It is yet unclear how the presence of such residual impacts aggregate statistics in dominant heterogeneous firms models.

Moreover, possible additional sources of firm-level heterogeneity can originate from the definition of sales. Sales has been interpreted to signify total sales (Axtell, 2001), exporting sales (di Gio-

[3]Most common distributions used in the economic literature are closed under power-law transformations (see, for instance, Mrázová et al. (2021) and Dewitte et al. (forthcoming).

[4]This almost one-to-one mapping with the distribution of productivity also appears for prices, profits, output and employment (Melitz and Redding, 2014, p. 12).

[5]For independent random variables X and Y, the variance of their sum or difference equals the sum of their variance.

vanni and Levchenko, 2013; Head et al., 2014), or domestic sales (Nigai, 2017). As soon as the country of study is an open economy with firm-level heterogeneity in the exporting/destination-decision, the one-to-one mapping between firm-level total sales and productivity disappears (di Giovanni et al., 2011; Amand and Pelgrin, 2016). Exporting sales only captures part of the firm population, with the size of that part depending on, among others, the size of the exporting destination and trade costs. The estimation of distribution parameters needs to be adapted accordingly, for instance by relying on truncated distributions (Sager and Timoshenko, 2019), and increases the possibility of finite sample biases. As for domestic sales, the productivity distribution can only be identified conditional on the correct identification of the elasticity of substitution (σ).

This second component of the aggregate trade elasticity, the elasticity of substitution, can be determined from the *firm-level gravity* equation. Specifically, it can be identified as the response of firm-level sales to cross-sectional and/or time variation in tariffs (Bas et al., 2017).[6] To do so, however, the researcher needs to capture the multilateral resistance terms accordingly and deal with other known issues as selection bias (the existence of firm-level zeros in trade data), heteroskedasticity and the difficulty in approximating trade costs (Yotov et al., 2016).[7] To date, we have no knowledge of gravity estimates that exploit the panel dimension of single-origin firm-level trade data while perfectly controlling for the multilateral resistance terms and/or stochastic productivity. As (Bas et al., 2017, footnote 11 on p.5) note, doing so with current techniques requires assumptions that are inconsistent with the underlying static trade theory. Bas et al. (2017), therefore, propose to rely on multiple-origin (minimum two) firm-level trade data and exploit solely the cross-sectional variation of the firm-level gravity equation to obtain theory-consistent elasticity of substitution estimates.

The proposed identification strategy in this paper solves the above described difficulties of prevalent identification strategies. Estimating a revenue production function allows us to exploit the panel dimension of firm-level production data from a single source country to identify both the productivity distribution and elasticity of substitution while controlling for transitory shocks to demand and supply.

This paper is also related to the literature that identifies aggregate trade elasticities from aggregate rather than firm-level data. Under the assumption of Pareto-distributed productivity, the two components of the aggregate trade elasticity collapse to a constant trade elasticity that can be identified from industry-level structural gravity equations.[8] Despite the popularity of this assumption, recent evidence exposes the superior performance of alternative distributional forms to capture firm-level heterogeneity (Head et al., 2014; Melitz and Redding, 2015; Nigai, 2017; Bee and Schiavo, 2018; Sager and Timoshenko, 2019). Adão et al. (2020), then, provide

[6] Note that one can also rely on the firm-level export demand equation to identify the elasticity of substitution from the variation in firm-level output due to variation in firm-level prices, if prices are correctly instrumented (see, for instance Fontagné et al. (2018); Fitzgerald and Haller (2018); Piveteau and Smagghue (2019)).

[7] Berthou and Fontagné (2016); Bas et al. (2017); Fitzgerald and Haller (2018) note that time variation in tariffs is small relative to the cross-sectional variation.

[8] See Arkolakis et al. (2012) for an exposition on the identification of aggregate trade elasticities from aggregate trade data under Pareto-distributed productivity and for a non-exhaustive list of works relying on this distributional assumption.

a general identification strategy for the aggregate trade elasticity using aggregate trade data, assuming that both the firm-level productivity distribution and elasticity of substitution are equal across countries. Additionally, the paper relates to the literature that studies the importance of the distributional assumption on trade forecasts (see, for instance di Giovanni et al. (2011); Head et al. (2014); Bas et al. (2017); Nigai (2017); Fernandes et al. (2018); Bee and Schiavo (2018); Sager and Timoshenko (2019)). However, distinct from this work, we consider the distribution of estimated productivity rather than using sales as a proxy for productivity. Egger et al. (2020) also relies on estimated productivity to recover firm-level heterogeneity, but do not focus on the importance of transitory shocks to supply and demand. Complimentary to our work, Sager and Timoshenko (2020) focus on the importance of export-market specific demand uncertainty when firms make an export decision and the subsequent impact on trade and welfare. The empirical analysis of Sager and Timoshenko (2020) relies on cross-sectional firm-level export quantity data and the gravity framework. The reliance on cross-sectional data implies that their framework can not differentiate between persistent and transitory shocks, the reliance on export rather than domestic quantity data implies that firm-size distribution estimates are systematically impacted by international trade (di Giovanni et al., 2011), and the reliance on quantity data implies Sager and Timoshenko (2020) can not account for transitory supply shocks. In contrast, the approach proposed in this paper allows disentangling transitory from persistent firm-level shocks to supply and demand and identifies the individual components of the aggregate trade elasticity (the elasticity of substitution parameter and firm-level productivity), solely requiring panel data on domestic production from a single source country. Our paper also relates to the granularity literature (Gabaix, 2011; di Giovanni and Levchenko, 2012; Eaton et al., 2012; Carvalho and Grassi, 2019), which discusses the transmission of firm-level shocks to the aggregate level. Our analysis demonstrates that independent transitory shocks have no aggregate implications in prevalent static heterogeneous firms models.

3 Heterogeneous firms model with idiosyncratic shocks

We specify a static open economy model with transitory idiosyncratic shocks to demand and supply. The core elements are derived from the (Melitz, 2003)-model augmented with transitory demand and supply shocks. The specification of these shocks follows the structural revenue production function estimation literature (see for instance Das et al. (2007); De Loecker (2011); Kasahara and Lapham (2013); Gandhi et al. (2020)). This will provide us with a general framework to evaluate the empirical methods that deduce firm-level productivity from sales data. We demonstrate that firm-level sales conflates heterogeneity in productivity with transitory shocks to demand and supply and, therefore, overestimates the variance of firm-level productivity. However, we reveal that if unbiased productivity estimates are available, modeled aggregate trade statistics are equivalent to those obtained in prevalent static heterogeneous firms models that do not feature supply and demand shocks (Melitz, 2003). We refer the reader to Online Appendix B for a detailed elaboration of the model.

Demand The preferences of a representative consumer in country $j \in J$ are defined over a continuum of horizontally differentiated varieties originating from country $i \in I$ ($\varpi \in \Omega^i$) and are assumed to take the Constant Elasticity of Substitution (CES) utility (U) form,

$$U^j = \left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu(\varpi)} y^{ij} (\varpi)^{\frac{\sigma-1}{\sigma}} d\varpi \right)^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

with σ the elasticity of substitution between varieties and $y^{ij}(\cdot)$ the quantity of a variety shipped from i that arrives in j. Utility maximization defines the optimal consumption and expenditure decisions over the individual varieties

$$\frac{y^{ij}(\varpi)}{Y^j} = \left[\frac{p^{ij}(\varpi)}{P^j} \right]^{-\sigma} e^{\nu(\varpi)}, \tag{2}$$

up to a variety-specific demand shock $e^{\nu(\varpi)}$ which is independent and identically distributed (i.i.d.) across varieties (see subsection 4.1 for a discussion of this assumption). The set of varieties consumed is considered as an aggregate good $Y^j \equiv U^j$ (Melitz, 2003) and P^j is the CES aggregate price index.

Supply There is a continuum of businesses, or firms, ($b \in B$) which choose to supply a distinct horizontally-differentiated variety. They are heterogeneous in terms of their productivity $\omega_b \in [0, \infty]$ drawn from the unconditional Cumulative Distribution Function (CDF) $G(\omega_b)$ after paying a fixed cost fe^i to enter the market.[9] The firm's productivity is assumed constant, such that $\omega_b \in \mathcal{I}_b$.[10] The information set $\mathcal{I}_b$ is a set of random variables that contains all the information a firm has available the moment it decides on the production level.

Production relies on a composite factor of production $\boldsymbol{A}_b^{ij}(\boldsymbol{\beta})$ (Melitz and Redding, 2014) subject to shocks to the production function e^{ϵ_b} which are i.i.d. across firms (see Subsection 4.1 for a discussion of this assumption):[11]

$$y_b^{ij} = q_b^{ij} e^{\epsilon_b} = \boldsymbol{A}_b^{ij}(\boldsymbol{\beta}) e^{\omega_b + \epsilon_b} \tag{3}$$

Supply of the production factor to the individual firm is perfectly elastic, so that firms are effectively price (W^i) takers on the input market.

Firms from country i have to pay a fixed cost f^{ij} to produce goods destined for country j

[9]We follow Asker et al. (2017) in differentiating all fixed costs from factors of production. "In their financial statements, firms report overhead costs as Selling, General and Administrative Expenses (SG&A). These expenses are not directly related to production, and include sales, advertising, marketing, executive compensation, ... and can in part be interpreted as expenses on intangible capital." (Asker et al., 2017, p. 4). We assume all fixed cost expenses are equally distributed within the source market.

[10]See Section 4 and Online Appendix B for an extension to a dynamic productivity specification.

[11]This composite factor can, for instance, be a Constant Returns to Scale Cobb-Douglas function of Z fixed (F) and V variable (L) factors of production respectively: $\boldsymbol{A}_b^{ij}(\boldsymbol{\beta}) = \prod_{z=1}^{Z} \prod_{v=1}^{V} (F_{bz}^{ij})^{\beta_z^i} (L_{bv}^{ij})^{\beta_v^i}$, where variable production factors can be adjusted after the realization of the information set $\mathcal{I}_b$ while fixed production factors can not be adjusted after the realization of the information set. See Online Appendix B for a model workout with such distinction between production factors.

denominated in final goods, and variable iceberg trade costs, $\tau^{ij} > 1$, denominated in units of labour of the origin country.

Profit maximization, then, results in an optimum quantity:

$$q_b^{ij} = \left(\frac{\sigma - 1}{\sigma} \frac{e^{\omega_b}}{\tau^{ij} W^i} \right)^{\sigma} Y^j \left(P^j\right)^{\sigma} \mathbb{E}_{\varepsilon^T} \left[e^{\varepsilon_b^T} \right]^{\sigma}. \tag{4}$$

where ε_b^T gathers the transitory demand and supply shocks $\left(\varepsilon_b^T = e^{\frac{\nu_b}{\sigma} + \frac{\sigma-1}{\sigma}\epsilon_b} \right)$. For each firm b, the timing assumptions of the model can thus be summarized as follows:

1. Observe the vector of state variables $\mathcal{I}_b$, with $\omega_b \in \mathcal{I}_b$;
2. Start producing the optimal quantity q_b;
3. Observe deviations from expectations regarding supply (ϵ_b) and realize final output y_b;
4. Observe deviations from expectations regarding demand (ν_b) and sell at a market-clearing price determined by the demand function.

The operational revenue for firms from country i selling in destination j at time t can be obtained as the product of output y_b with the market-clearing price p_b given by eq. 2:

$$\begin{aligned} x_b^{ij} = p_b^{ij} y_b^{ij} &= \left(y_b^{ij} \right)^{\frac{\sigma-1}{\sigma}} \left(Y^i\right)^{\frac{1}{\sigma}} e^{\frac{\nu_b}{\sigma}} P^j \\ &= \left(\frac{\sigma}{\sigma - 1} \tau^{ij} W^i \right)^{1-\sigma} Y^j \left(P^j\right)^{\sigma} \mathbb{E}_{\varepsilon^T} \left[e^{\varepsilon_b^T} \right]^{\sigma-1} e^{(\sigma-1)\omega_b + \varepsilon_b^T}. \end{aligned} \tag{5}$$

Equation 5 shows that firm-level variation in sales originates from (at least) two different sources: productivity ω_b, and transitory shocks ε_b^T, such that $x_b^{ij} \dot{\sim} e^{(\sigma-1)\omega_b + \epsilon_b^T}$. This stands in contrast with the prevalent method of using sales as a proxy for productivity which, as explained in the literature review above, attributes variation in sales solely to variation in productivity up to the elasticity of substitution: $x_b^{ij} \dot{\sim} e^{(\sigma-1)\omega_b}$. The i.i.d. nature of the idiosyncratic shocks implies that the variance of productivity measures which conflate productivity with these shocks will always be higher than the variance of productivity in itself: $Var\left(e^{(\sigma-1)\omega_b + \epsilon_b^T}\right) > Var\left(e^{(\sigma-1)\omega_b}\right)$.

Operating decisions The productivity cutoffs for serving each market are determined by two equations. First, a firm decides whether to exit or enter/stay in a market based on its ability to generate a positive profit π^{ij}, resulting in a zero-profit condition:

$$0 = \mathbb{E}_{\varepsilon^T} \left[\pi^{ij} \left((\sigma - 1)\omega^{ij*} + \varepsilon_b^T \right) \right]. \tag{6}$$

Second, a subset of active firms make positive profits net of the sunk entry cost. Free entry implies that in equilibrium, this expected measure of ex-ante profits (inclusive of the entry cost) must be equal to zero

$$fe^i = \left[1 - G(\omega^{ii*})\right] \mathbb{E}_{\omega,\varepsilon^T} \left[\sum_{j=1}^{J} \pi^{ij} \left((\sigma-1)\omega_b + \varepsilon_b^T \right) \Big| \omega_b > \omega^{ii*} \right]. \tag{7}$$

Aggregation With the productivity cutoffs determined, we can sum equation 5 across all active firms trading between i and j (M^{ij}) to obtain an expression for aggregate trade between country i and j as the product of the number of firms and average sales:

$$\begin{aligned} x^{ij} &= \frac{M^{ij}}{1 - G(\omega^{ij*})} \int_{\omega^{ij*}}^{\infty} \int_{-\infty}^{\infty} x_b^{ij} dG(\omega_b) dG(\varepsilon_b^T) \\ &= \frac{M^{ij}}{1 - G(\omega^{ij*})} \left(\frac{\sigma}{\sigma-1} \tau^{ij} W^i \right)^{1-\sigma} Y^j \left(P^j\right)^{\sigma} \mathbb{E}_{\varepsilon^T} \left[e^{\varepsilon_b^T} \right]^{\sigma} \int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b). \end{aligned} \tag{8}$$

From this aggregate revenue expression, we observe that transitory shock-induced idiosyncratic heterogeneity aggregates up to a constant $\mathbb{E}_{\varepsilon^T} \left[e^{\varepsilon_b^T} \right]^{\sigma}$.

The partial sensitivity of aggregate trade to changes in variable trade costs, the aggregate trade elasticity, can then be defined as (Chaney, 2008; Arkolakis et al., 2012; Melitz and Redding, 2014; Bas et al., 2017):

$$\begin{aligned} \gamma^{ij} \equiv \frac{\partial ln X^{ij}}{\partial ln \tau^{ij}} \bigg|_{\omega^{ii*}} &= 1 - \sigma - \frac{e^{\sigma \omega^{ij*}} g(\omega^{ij*})}{\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)} \\ &= \underbrace{1-\sigma}_{\text{intensive margin}} - \underbrace{\frac{e^{(\sigma-1)\omega^{ij*}} (1 - G(\omega^{ij*}))}{\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)}}_{\text{min-to-mean ratio}} \times \underbrace{\frac{d \ln M^{ij}}{d ln \tau^{ij}}}_{\text{extensive margin}}. \end{aligned} \tag{9}$$

This elasticity is decomposed into the sum of the intensive margin to trade (the variation in bilateral trade flows due to changes in average exporter size) and the (weighted) extensive margin $\frac{dlnM^{ij}}{dln\tau^{ij}} = \frac{e^{\omega^{ij*}} g(\omega^{ij*})}{1-G(\omega^{ij*})}$. The extensive margin measures the variation in bilateral trade flows due to changes in the number of exporters. This extensive margin is weighted by the min-to-mean ratio. Intuitively, the weight of the extensive margin will be decreasing when the market gets easier. Easier markets have a larger presence of weaker firms, which reduces the min-to-mean ratio. As a result, the marginal entrant's influence on aggregate exports decreases. In the limit, the weight of the extensive margin becomes negligible and the aggregate elasticity is completely determined by the intensive margin (Bas et al., 2017).

It can be observed that, if the productivity measure ω_b is unbiased, the aggregate trade elasticity is independent of transitory shocks to demand and supply. Whereas the firms' expectations of transitory shocks to demand and supply affect aggregate trade costs levels, the independent nature of these shocks renders these expectations invariable to a change in variable trade costs. As such, transitory shocks do not affect the changes in trade flows as the result of a change in variable trade costs.

Similarly, the changes in welfare (measured in terms of real wages) from a change in variable trade costs ($\boldsymbol{\tau} \to \boldsymbol{\tau}'$) can, upon choosing the composite wage as the numeraire $W^i = 1$, be written as a ratio of the aggregate price indices. Whereas the firms' expectations of transitory shocks to demand and supply affect aggregate price levels, the independent nature of these shocks renders these expectations invariable to a change in variable trade costs. As such, the transitory shocks do not affect the changes in the price indices as the result of a change in variable trade costs. Transitory shocks have no influence on the aggregate gains from trade:

$$\frac{(\mathbb{W}^i)'}{\mathbb{W}^i} = \frac{P^i}{(P^i)'}. \tag{10}$$

Overall, the intuition for the role of shocks in the model is as follows. As shocks to supply and demand are assumed i.i.d., a productivity measure that conflates productivity with transitory shocks to demand and supply will always have a higher variance than productivity in itself. However, the expectation of these shocks reduces to a constant due to their i.i.d. nature. As a result, all aggregate variables in the model are determined up to a constant. A comparison of the modeled impact of exogenous developments (that do not change the nature of these shocks) cancel out and will, conditional on an unbiased measurement of productivity, be equal to the impact from a model without transitory shocks. This reasoning mimics the logic related to the need to identify the productivity distribution up to a constant only (Bee and Schiavo, 2018).

4 Identification and Estimation

From the literature review in section 2, it was apparent that the current identification of aggregate trade elasticities based on firm-level cross-sectional trade data is hampered by transitory shocks to demand and supply. Moreover, we have no knowledge of firm-level gravity estimation techniques that can rely on single-origin data to identify the aggregate trade elasticity. Therefore, we extend our theoretical setting specified above to exploit the variation of a firm's productivity over time using panel data from a single source country. We rely on the productivity estimation literature to propose a theoretically underpinned alternative to the prevalent gravity/sales as proxy for productivity identification scheme. Following Klette and Griliches (1996), we combine the production function with the CES demand system to simultaneously identify the two components of the aggregate trade elasticity, productivity and the elasticity of substitution, while controlling for transitory shocks to demand and supply.

4.1 Theoretical model extension to uncertainty in future productivity

To allow for the stylized theoretical model specified above to be identifiable in panel data from a single source country, we extend the theoretical model with stochastically evolving productivity over time (Hopenhayn, 1992).[12] We assume that productivity follows a Markov process independent across firms with conditional distribution $G(\omega_{bt+1}|\mathcal{I}_{bt})$ such that:[13]

$$e^{\omega_{bt+1}} = \mathbb{E}_{\omega}\left[e^{\omega_{bt+1}}|\mathcal{I}_{bt}\right] e^{\eta_{bt+1}}. \tag{11}$$

Productivity at time $t+1$ is specified as a function of the information set of the firm at time t, $\mathcal{I}_{bt}$ with $\omega_{bt} \in \mathcal{I}_{bt}$, and a productivity shock η_{bt+1}. See Online Appendix B for a full elaboration of the dynamic model setup.

Time-varying domestic sales, then, can be specified as the log-linearized combination of the time-varying domestic production function (eq. 3), augmented with dynamic productivity, and the time-varying domestic CES demand system (eq. 2) (Klette and Griliches, 1996; De Loecker, 2011):

$$\begin{aligned} ln\left(\frac{x_{bt}^{ii}}{P_t^i}\right) = ln\left(\frac{p_{bt}^{ii}y_{bt}^{ii}}{P_t^i}\right) &= ln\left(\left(y_{bt}^{ii}\right)^{\frac{\sigma-1}{\sigma}}\left(Y_t^i\right)^{\frac{1}{\sigma}} e^{\frac{\nu_{bt}}{\sigma}}\right) \\ &= \frac{\sigma-1}{\sigma}\boldsymbol{A}_{bt}^{ii}(\boldsymbol{\beta}) + \frac{1}{\sigma}lnY_t^i + \frac{\sigma-1}{\sigma}\omega_{bt} + \varepsilon_{bt}^T, \end{aligned} \tag{12}$$

where productivity ω_{bt} and the transitory shocks ε_{bt}^T are unobserved. A domestic revenue function specification ($i = j$) avoids conflating productivity with heterogeneity originating from export output (Amand and Pelgrin, 2016; Nigai, 2017). Notice the difference between this 'production function specification' of the revenue equation and the 'gravity specification' of the revenue equation in equation 5. The gravity specification relies on the profit-maximizing optimal input mix to rewrite firm-level revenue as a function of aggregate variables and two sources of firm-level heterogeneity: productivity ω_{bt}, and transitory shocks ε_{bt}^T. This production function specification, on the other hand, relies on actual firm-level heterogeneous input use.

This specification reiterates the idea that relying on firm-level sales as a proxy for productivity results in ignoring a component of stochastic variation present in firm-level sales that is economically relevant. Idiosyncratic shocks $\left(\varepsilon_{bt}^T = \frac{\nu_{bt}}{\sigma} + \frac{\sigma-1}{\sigma}\epsilon_{bt}\right)$ combine supply- and demand-side deviations that are not expected to be negligible. Kasahara and Lapham (2013) show that this residual component accounts for approximately 10% of the variation in firm-level sales vis-à-vis productivity $e^{\omega_{bt}}$.

The supply-side specification of these deviations, ϵ_{bt}, represents common practice in the productivity estimation literature to allow for deviations from productivity. Whereas persistent productivity "$[\omega_{bt}]$ might represent variables such as the managerial ability of a firm, expected

[12] The theoretical framework can be interpreted as a dynamic setting where $G(\omega_{bt+1}|\mathcal{I}_t)$ is such that $\omega_{bt+1} = \omega_{bt} = \omega_b$ (Melitz, 2003), while still allowing for uncertainty in realized supply and demand. For a discussion on the implications of productivity dynamics on the economy, see Impullitti et al. (2013); Alessandria and Choi (2014); Ruhl and Willis (2017). Limiting the dynamics in our main model specification allows for clear analytical expressions for the equilibrium variables. Moreover, it allows us to demonstrate the influence of transitory shocks on the trade elasticity and GFT compared to the predominant Melitz (2003)-model (see for instance Head et al. (2014); Melitz and Redding (2015); Nigai (2017); Bee and Schiavo (2018)) in a straightforward manner.

[13] $\mathbb{E}_x[\ldots] = \int \ldots f(x)dx$.

down-time due to machine breakdown, expected defect rates in a manufacturing process, soil quality, or the expected rainfall at a particular farm's location", transitory supply side shocks "[ϵ_{bt}] might represent deviations from expected breakdown, defect, or rainfall amounts in a given year" (Ackerberg et al., 2015, p.2414).

Our demand-side specification of these deviations, ν_{bt}, follows De Loecker (2011) in assuming them to be firm-specific, transitory and unobserved residual demand shocks. Notice that the demand-side of our model can account for serial correlation in these shocks. First, the model controls for aggregate demand shocks through the aggregate demand shifter Y_t^i. Second, it can allow for firm-specific demand-side deviations to be correlated over time $\tilde{\nu}_{bt} \in \mathcal{I}_{bt}$ next to the transitory demand shocks ν_{bt}. Whereas persistent demand-side deviations $\tilde{\nu}_{bt}$ might represent variables such as the marketing abilities of a firm, expected drop in demand due to external conditions such as roadworks, expected demand departures from the aggregate demand trends, ... transitory demand shocks ν_{bt} might represent deviations from expected external conditions, unexpected departures from the aggregate demand trends, ... If the dynamics of these correlated demand shocks are similar to those of productivity (see De Loecker (2011); Gandhi et al. (2020)), the variable capturing persistent heterogeneity in equation 5 would become $\tilde{\omega}_b = \omega_{bt} + \frac{\tilde{\nu}_{bt}}{(\sigma-1)\sigma}$. This variable is referred to as business conditions (Bloom, 2009) or profitability conditions (Sager and Timoshenko, 2020) rather than productivity ω_{bt}. As the distinction between persistent heterogeneity and transitory shocks to supply and demand remains, allowing for firm-level serial correlation in demand affects the interpretation of the persistent heterogeneity component but does not affect the main results of the model specification.[14,15]

4.2 Identification

The data on which we rely (see subsection 4.2) only contains information on total input use, not market-specific input use. To resolve this issue, we rely on Rivers (2010) who shows that when the elasticities of demand are the same across markets, firms choose to allocate output such that the prices received by the firm across markets are equal. Since the prices are equal, this implies that the fraction of quantities across markets, which is not observed in the data, is equal to the fraction of revenues across markets, which is observed in the data: $\theta_{bt}^{ij} = \frac{x_{bt}^{ij}}{\sum_{j=1}^{J} x_{bt}^{ij}} = \frac{q_{bt}^{ij}}{\sum_{j=1}^{J} q_{bt}^{ij}}$. This observation allows us to rewrite domestic revenue as the product of the fraction of domestic revenue in overall revenue and overall revenue, which is determined by *observed* total input use:

[14]If we would have access to firm-level quantity rather than sales data, our methodology could account for serially correlated demand shocks of which the dynamics differ from the productivity dynamics. In the absence of such data, however, allowing for such correlation directly affects the estimation approach specified below by introducing an additional serially correlated unobserved state variable in the model, and this affects both the invertibility conditions and the ability to identify the parameters (De Loecker, 2011).

[15]The assumption on identically distributed deviations, on the other hand, is more difficult to be relaxed. It is a restrictive necessity for our theoretically underpinned identification strategy provided the data available (see Section 3) and to obtain equivalent aggregate trade statistics between a static version of this model and the prevalent heterogeneous firms model (see Section 5). Demand shocks could, for instance, be specified as partly consisting of serially correlated market-specific shocks (Sager and Timoshenko, 2020) without altering our main theoretical conclusions. The proposed estimation procedure, however, would require information on firm-level market-specific factor input use in that case. Notice that our specification already controls for market-specific aggregate demand shocks through the aggregate demand shifter Y_t^i.

$$
\begin{aligned}
ln\left(\frac{x_{bt}^{ii}}{P_t^i}\right) &= \frac{\sigma-1}{\sigma} ln\theta_{bt}^{ii} + ln\frac{x_{bt}^{i}}{P_t^i} \\
&= \frac{\sigma-1}{\sigma} ln\theta_{bt}^{ii} + \frac{\sigma-1}{\sigma}\boldsymbol{A}_{bt}^{i}(\boldsymbol{\beta}) + \frac{1}{\sigma} lnY_t^i + \frac{\sigma-1}{\sigma}\omega_{bt} + \varepsilon_{bt}^{T}.
\end{aligned} \tag{13}
$$

Equation 13 is our main estimating equation. The parameter identification for this equation is not straightforward, as productivity ω_{bt} is unobserved. A Nonlinear Least Squares (NLLS) estimation, for instance, will deliver biased coefficients as factor components of composite production factor $\boldsymbol{A}_{bt}^{i}(\boldsymbol{\beta})$ and the demand shifter Y_t^i are correlated with current productivity: $E\left[Y_t^i\left(\frac{\sigma-1}{\sigma}\omega_{bt} + \varepsilon_{bt}^{T}\right)\right] \neq 0$. Therefore, we rely on a structural productivity estimation technique like De Loecker (2011).[16] This technique uses the Ackerberg et al. (2015) proxy-variable approach (ACF) to separate the residual transitory component from our main estimation equation (eq. 13) in a first stage. In line with the specified theoretical model (which does not feature intermediate inputs), Ackerberg et al. (2015) typically rely on intermediate inputs as a proxy variable and use a value-added specification of the production function.[17] In a second stage, the estimation procedure relies on the Markov assumption for productivity (see eq. 11) to avert endogeneity problems and obtain consistent parameter and productivity estimates. We recover an estimate of the elasticity of substitution σ from the variation in the demand shifter Y_t^i over time (Klette and Griliches, 1996; De Loecker, 2011; Halpern et al., 2015). The production function parameters $\boldsymbol{\beta}$ can be recovered from this elasticity substitution estimate and the estimates of the production function parameters up to the elasticity of substitution $\frac{\sigma-1}{\sigma}\boldsymbol{\beta}$, which are obtained exploiting the variation in the firm-level production factors over time. A consistent identification of all revenue production function parameters allows us to identify firm-level productivity ω_{bt} separately from transitory shocks to supply and demand ϵ_{bt}^{T}.

4.3 Data

For our empirical analysis, we rely on a large panel of French firms extracted from the Amadeus database by Bureau Van Dijk Electronic Publishing. The Amadeus database contains financial information (balance sheet and profit and loss account) as well as information on firms' location, activity, ownership, etc. We construct a sample covering the period 1998-2006 using multiple issues of the database (October releases from 1998 till 2015)[18]. Despite its wide geographical, time and sectoral range, the sample of firms can vary considerably. The providers of the database rely on national data sources, which are subject to change. In addition, for firms that do not provide information for more than three consecutive years, all (historical) information is removed. The estimation of firm productivity and the its subsequent analysis are therefore based on an extended version of Amadeus (now Orbis), as described in Merlevede et al. (2015) (see also (Kalemli-Ozcan et al., 2015) for a discussion on the construction of nationally representative

[16]See Online Appendix C for an elaborate description of the estimation strategy.

[17]As the differences in overall demand across foreign markets is captured by the θ-term in eq. 13, this proxy-variable approach is not affected by these demand differentials.

[18]A single issue is only a snapshot of the ownership information and firms that exit are dropped from the next issue released. A single issue further only contains 10 years of financial data at maximum. In order to get a full overview of activity, location, ownership and financials through time, multiple issues are required.

firm-level data based on the Orbis database). Compiling annual versions of Amadeus, the extended database attenuates variability in the sample composition.

We restrict the dataset to manufacturing firms (NACE 2 class 10–33) that report positive total and domestic operating sales, tangible fixed assets, number of employees, costs of employees, material inputs and value added.[19] All monetary variables are deflated using the appropriate NACE 2-digit deflator from the EU-KLEMS database. Real output are sales deflated with producer price indices. Capital are tangible fixed assets deflated by the average of the deflators for five NACE 2-digit industries according to Javorcik (2004). Log aggregate sales, defined as the market share weighted sum of log deflated sales acts as the aggregate demand shifter in our main estimation equation (Klette and Griliches, 1996; Rivers, 2010; De Loecker, 2011).[20] Real material inputs are obtained by deflating material inputs with an intermediate input deflator as a weighted average of output deflators where the country-industry-time specific weights are based on intermediate input uses retrieved from input-output tables. Value added is then obtained as the difference between real output and real material inputs. Labor is simply the number of employees.

Our focus on France is motivated by the presence of information on total firm-level exports in the Amadeus database for French firms. This allows us to use domestic firm-level sales, calculated as the difference between total firm-level sales and total firm-level export sales, avoiding conflating firm-level productivity with heterogeneity in export output (see also Section 2). Furthermore, the trade literature has mainly focused on France concerning research on the characterization of the productivity distribution (see, for instance, di Giovanni and Levchenko (2013); Head et al. (2014); Nigai (2017); Bee and Schiavo (2018)). Our final database contains 379,765 observations from 86,959 unique French firms over the years 1998–2006. Summary statistics in Appendix Table 1 reveal that our database covers a wide range of the firm universe in France.

4.4 Estimation results

We apply both the Nonlinear Least Squares (NLLS) and the structural productivity estimation procedure (ACF) as described in Section 4.1 for a value-added Cobb-Douglas production function to our complete French firm-level dataset. The resulting parameter estimates are displayed in Table 1. The capital and labor elasticities ($\boldsymbol{\beta}$) are slightly overestimated by the NLLS procedure compared to the ACF estimation procedure, which controls for endogeneity of productivity. Both estimators report increasing returns to scale (RTS), which is common after removing the contribution of the elasticity of substitution to the overall production function estimates ($\frac{\sigma-1}{\sigma}\boldsymbol{\beta}$) (De Loecker, 2011; Halpern et al., 2015). The ACF elasticity of substitution estimate in this paper takes a value of 4.59, which is in line with previously reported estimates obtained from

[19] We clean the data both on levels and on growth rates to prevent effects of extreme outliers and extreme noise on the analysis. Specifically, we limit the sample to observations with a labor use larger than 1 and limit deflated turnover, deflated materials and deflated capital to values larger than 1,000 euro. Further, we removed the yearly lowest and highest percentile of the included variables (domestic sales, capital, labour, and materials) and dropped observations with yearly growth rates of included variables higher than 100 in absolute values.

[20] The firms market share's are assumed to be equivalent to the weights used to construct the aggregate price deflator (see (Klette and Griliches, 1996)). The validity of aggregate sales as an identifying variable for the demand elasticity follows from the optimal consumption specification (see equation 2).

variation in demand shifters on firm-level revenue (Rivers, 2010; De Loecker, 2011; Kasahara and Lapham, 2013) as well as from variation in trade costs on firm-level exports (Bas et al., 2017).

Table 1: Production function estimation results

	$\frac{\sigma-1}{\sigma}\boldsymbol{\beta}$			$\boldsymbol{\beta}$			
	Capital	**Labor**	**RTS**	**Capital**	**Labor**	**RTS**	**Elasticity of Substitution**
NLLS	0.145	0.876	1.021	0.177	1.072	1.249	5.482
	(0.002)	(0.007)	(0.008)	(0.002)	(0.007)	(0.008)	(0.157)
ACF	0.130	0.818	0.948	0.166	1.045	1.212	4.590
	(0.002)	(0.007)	(0.007)	(0.002)	(0.007)	(0.007)	(0.110)

Notes: Standard errors displayed between brackets are obtained from wild bootstrap clustered at the firm level with 99 repliations. Estimates obtained from French firm-level database over the years 1998–2006 with 379,765 observations from 86,959 firms.

With consistent estimates of the elasticity of substitution, productivity, and transitory shocks at hand, we turn our attention to the distribution of productivity and the influence of shocks on this distribution. Next to estimated *productivity* $\left(e^{\hat{\omega}_{bt}}\right)$, which is free from transitory shocks, we construct three additional measures of productivity based on equation 5. These additional measures will allow for a straightforward comparison with measures of productivity currently used in the literature.

We augment productivity with transitory shocks to obtain *shock-included productivity* $\left(e^{\hat{\omega}_{bt}+\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$. Additionally, we consider domestic *sales as a proxy* for productivity $\left(\left(x_{bt}^{ii}\right)^{\frac{1}{\hat{\sigma}-1}}\right)$ in line with the trade literature (see Section 2). This measure conflates productivity with transitory shocks to demand and supply and, thanks to our identification procedure, can now be compared to *shock-excluded sales as a proxy* for productivity $\left(\left(x_{bt}^{ii}\right)^{\frac{1}{\hat{\sigma}-1}} e^{-\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$.[21]

We report the variance and ratios of the 75-25, 90-10, 95-5, and 99-1 quantiles of the respective heterogeneity variables in Table 2 (See Online Appendix Figure 1 for the accompanying Kernel densities). The reported variance in this Table increases by ±7% (for sales as a proxy) or ±53% (for estimated productivity) when productivity is conflated with shocks to demand and supply. This confirms the numbers reported by Kasahara and Lapham (2013) on the importance of idiosyncratic shocks when measuring heterogeneity. Moreover, the overdispersion induced by these residuals grows towards the tails of the distribution. There are larger differences in the tail ratio (the 99-1 quantile) than the 75-25 ratio.

Overall, our theoretically underpinned identification strategy provides sensible parameter estimates and the estimation results confirm the first part of our premise: conflating transitory shocks

[21]Notice that shock-excluded sales can differ from productivity. This difference can be attributed to the differing underlying assumptions. Whereas firm-level sales relies on profit maximization to assume away firm-level heterogeneous input expenditures, the revenue production function explicitly controls for these expenditures (see also the discussion below equation 12). Therefore, if, for instance, firm-level distortions are present (Hsieh and Klenow, 2009), the observed input expenditures will deviate from those expected under profit maximization without taking distortions into account. Additionally, sources of heterogeneity can originate from a wrong production function specification or firm-level heterogeneity in the demand function (for instance markups) that affects the firm-level sales specification differently than the firm-level revenue production specification. While this poses an interesting puzzle, pursuing the solution of this puzzle falls outside the scope of this paper.

Table 2: Variance and quantile ratios for sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-augmented productivity, and productivity in the year 2006.

Variable	Variance	75/25	90/10	95/5	99/1
Sales as proxy $\left((x_{bt}^{ii})^{\frac{1}{\hat{\sigma}-1}}\right)$	0.224	1.852	2.949	3.848	5.905
Shock-excluded Sales as proxy $\left((x_{bt}^{ii})^{\frac{1}{\hat{\sigma}-1}}\, e^{-\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$	0.209	1.766	2.813	3.633	5.499
Shock-included Productivity $\left(e^{\hat{\omega}_{bt}+\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$	0.101	1.495	2.093	2.566	4.158
Productivity $\left(e^{\hat{\omega}_{bt}}\right)$	0.066	1.415	1.865	2.248	3.538

Notes: Values obtained from sample of 34,339 French firms in 2006.

with productivity results in an overdispersed distribution of measured productivity.

5 Aggregate implications

Having established that idiosyncratic transitory shocks exist, are sizable and result in overdispersed productivity, we want to evaluate its influence on aggregate trade statistics such as the trade elasticity and GFT. To this end, we quantify the bias in the aggregate trade elasticity and subsequent GFT that results from the in the previous section established bias in firm-level productivity measurement. As established in the theoretical section (Section 3), any observed bias can solely be ascribed to firm-level productivity mismeasurement, and not to model misspecification.

5.1 Aggregate trade elasticity

Quantifying the aggregate trade elasticity (eq. 9) requires a value for the elasticity of substitution and a fitted parametric distribution function. We assume the previously obtained value for the elasticity of substitution of 4.59 and fit a Lognormal distribution to four specifications of firm-level productivity: sales as a proxy for productivity $\left((x_{bt}^{ii})^{\frac{1}{\hat{\sigma}-1}}\right)$, transitory shock-excluded sales as a proxy for productivity $\left((x_{bt}^{ii})^{\frac{1}{\hat{\sigma}-1}}\, e^{-\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$, transitory shock-included productivity $\left(e^{\hat{\omega}_{bt}+\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$ and estimated productivity $\left(e^{\hat{\omega}_{bt}}\right)$. We plot the resulting trade elasticities γ^{ij} in function of the probabilities for a firm from country i to be active in market j, the probability of serving an export market $1 - G(\omega^{ij*})$, in Figure 1. Two conclusions can be drawn from this Figure.

First, for a given ease of market access, productivity variables augmented with transitory shocks (that is, sales as a proxy and shock-included productivity) result in less elastic trade than their shock-free counterparts. This phenomenon can be ascribed to the larger variance of productivity variables when augmented with these shocks (see Table 2). In more heterogeneous economies, highly productive firms represent a larger fraction of firms. Therefore, in more heterogeneous economies, aggregate exports are less sensitive to changes in transportation costs as the extensive margin reaction (the number of firms exiting and entering) is smaller relative to the intensive margin reaction (the change in average exports of incumbents) when variable costs fluctuate (Chaney, 2008). Neglecting the existence of transitory shocks, therefore, results in overestimated trade elasticities (i.e., less elastic trade).

Moreover, the bias in calculating trade elasticities increases as the ease of market access decreases. At a probability of 0.25 for a domestic firm to serve a foreign market, for instance, the distribution of shock-included productivity results in a trade elasticity about 10.9% (-5.49/-6.13) larger than when calculated based on the productivity distribution. This difference increases to an approximate 12.1% difference at a probability of 0.1 for a domestic firm to serve a foreign market, and continues to increase as the export market becomes more difficult. Bas et al. (2017) demonstrate that the majority of export markets have a probability of exporting smaller than 0.1.[22] Such results are in line with results reported in Table 2: heterogeneity differences between variables augmented with and purged from shocks are larger in the tail of the distribution. Trade elasticities obtained from shock-free productivity measures therefore diverge, as market access becomes increasingly restrictive, from the intensive margin of trade $(1 - \hat{\sigma})$ increasingly faster than their shock-augmented counterparts.

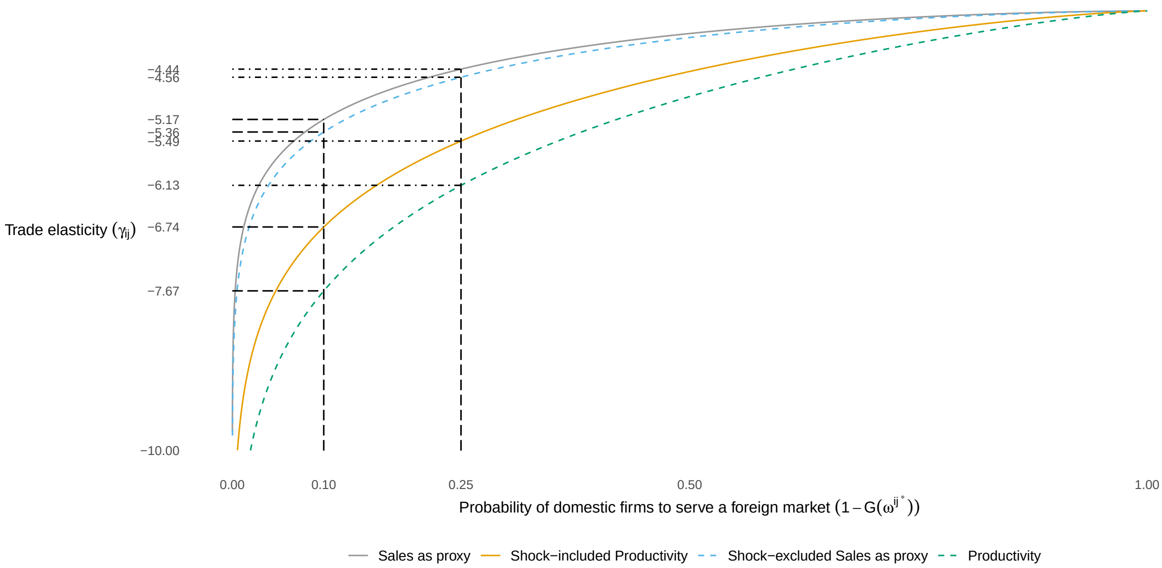

Figure 1: Trade elasticities for sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-included productivity and productivity in the year 2006.

Note: Trade elasticities calculated assuming a fitted Lognormal distribution function and an elasticity of substitution of 4.59.

5.2 Gains From Trade

We extend the results found for the (partial) trade elasticity in the previous subsection to the full trade elasticity and the accompanying welfare implications in this subsection. For simplicity, we rely on a stylized two-country symmetric heterogeneous firms model. This allows us to perform a GFT exercise in line with the current literature investigating the importance of distributional

[22] If we would have access to firm-level market-specific export data, our methodology would straightforwardly allow to determine the export productivity cut-offs as the export-market specific minimal productivity level. We are grateful to an anonymous referee for pointing this out.

assumptions on GFT (Head et al., 2014; Melitz and Redding, 2015; Bee and Schiavo, 2018) and investigate the influence of our alternative estimation procedure on welfare predictions.

The parameterization of our model is standard (Head et al., 2014; Melitz and Redding, 2015; Bee and Schiavo, 2018). We work with two symmetric countries i and j and choose labor in one country as the numeraire, so that $W^i = W^j = 1$. We calibrate the variable trade cost to match the average fraction of exports in firm sales in French manufacturing of 0.08, which amounts to $\tau^{ij} = 1.96$.[23] For this value of the variable trade cost, we calibrate fixed entry costs and fixed exporting costs relative to fixed domestic costs ($f^{ii} = 1$) to match the export participation rate (0.43 in our dataset) and the exogenously determined entry rate of 0.5 (Head et al., 2014; Melitz and Redding, 2015). The value of the elasticity of substitution remains at its estimated value, $\hat{\sigma} = 4.59$, and we continue to parameterize all heterogeneity variables assuming a Lognormal distribution for productivity.

We calculate the percentage changes in welfare from a reduction in variable trade costs relative to autarky ($\tau^{ij} = 10$). The resulting percentage GFT are displayed in Figure 2. We can immediately observe that the ranking in terms of heterogeneity and in terms of trade elasticities is preserved for GFT. Conflating productivity with firm-level noise results in overestimated GFT. GFT evaluated at the calibrated variable trade costs ($\tau^{ij}_{cal} = 1.96$) obtained using estimated productivity amounts to 0.93%, about 0.15 percentage points (or 20.45%) lower than the GFT predicted from shock-augmented productivity.[24] Moreover, these welfare gains from increased market access are realized faster for shock-included productivity. GFT estimates are overestimated by 36.24% for a reduction in variable trade costs to the calibrated value of 2.2 (see also Online Appendix Figure 2). This is in line with results reported for the partial trade elasticity, which approaches its intensive margin faster for transitory shock-included productivity than pure estimated productivity as market access improves. A similar picture emerges for sales as a proxy for productivity.

6 Robustness

It is possible that the size of the identified transitory shocks to demand and supply are affected by our modeling choices. Specifically, firm-level noise can be affected by a misspecified production function or data cleaning.[25] Additionally, it is possible that the heterogeneity in productivity is insufficiently captured assuming a Lognormal distribution. We attenuate these concerns with the following three robustness tests.

Firstly, we evaluate the importance of transitory shocks assuming a Translog production function. Results, available in Online Appendix Figures 3, 5, and 8 as well as Tables 2 and 4 reveal

[23] In a two-country symmetric heterogeneous firms model with $\tau^{ii} = 1$, we have that $\frac{x_b^{ij}}{x_b^{ii}+x_b^{ij}} = \frac{\tau^{1-\sigma}}{1+\tau^{1-\sigma}}$ (Melitz and Redding, 2015).

[24] A comparison in percentage rather than absolute differences is preferred due to the stylized model this calibration exercise relies on. Absolute differences are likely more sensitive to model specification and parametrization. See Costinot and Rodríguez-Clare (2014) for a discussion on the sensitivity of GFT on model specifications.

[25] Additionally, firm-level noise can also increase if our demand function is misspecified. As this assumption of a CES demand function is shared with current practices of measuring aggregate trade elasticities (see Section 2), we do not attempt to control for a possible misspecification of demand.

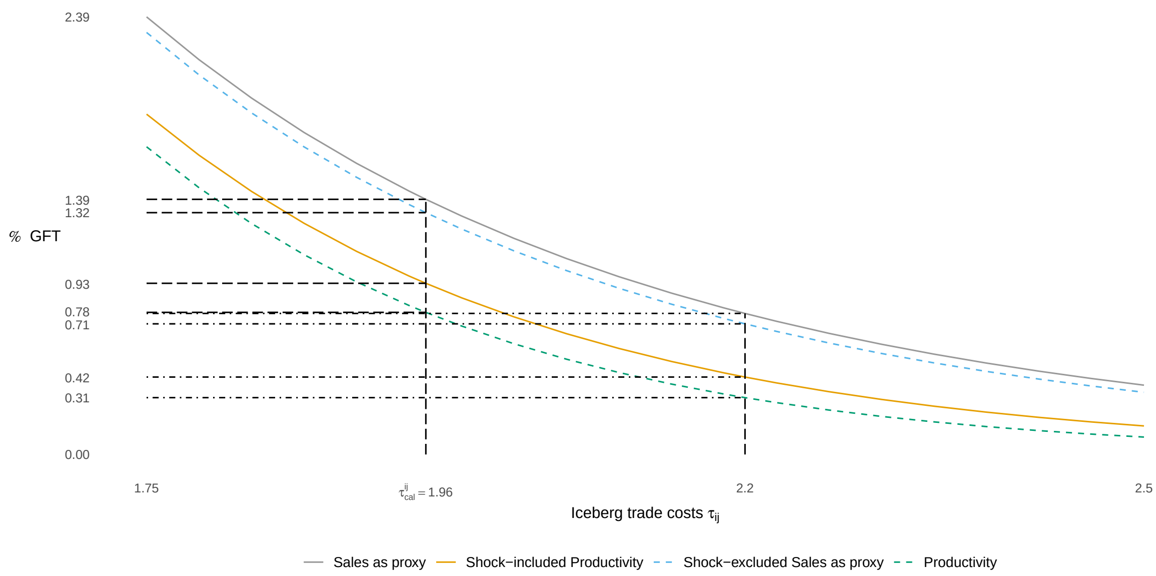

Figure 2: Percentage welfare gains om a reduction in variable trade costs relative to Autarky ($\tau = 10 \rightarrow \tau'$)

that assuming a Translog production function rather than a Cobb-Douglas specification does not alter the main results. Secondly, we re-evaluate our analysis with an uncleaned data sample, to ensure the results are not influenced by data cleaning. The results are displayed in Online Appendix Figures 4, 6 and 9 as well as Tables 3 and 5. The analysis on the uncleaned data sample results in a relatively high estimated elasticity of substitution, probably due to the presence of relatively large firms in the uncleaned data sample that might exert market power, but does not alter our conclusions regarding the importance of idiosyncratic shocks. Lastly, whereas the Lognormal distribution is a commonly used alternative for the Pareto distribution and has been argued to provide a relatively good fit to the data (Head et al., 2014), it is possible the heterogeneity in productivity is insufficiently captured assuming a Lognormal distribution. We therefore also calculate aggregate trade elasticities and GFT assuming a Weibull and Gamma distribution (see Online Appendix figures Figure 7 and 10).[26] Again, our main results stand: not controlling for transitory idiosyncratic shocks results in an overestimation of aggregate trade elasticities and GFT.

7 Conclusion

This paper identifies and evaluates transitory shocks to demand and supply as a source of bias in the measurement of firm-level productivity and the aggregate trade elasticities and Gains From Trade (GFT) that are derived from it. If productivity is conflated with transitory idiosyncratic shocks, we obtain an overdispersed distribution of measured productivity, and overestimated trade elasticities and GFT. In light of these shocks, prevalent methods to identify aggregate trade elasticities result in biased measurements. We propose a theoretically underpinned alternative

[26] A log-likelihood ratio test favors the Lognormal distribution over these alternative distributional forms for all definitions of productivity.

that estimates both components of the aggregate trade elasticity (elasticity of substitution and productivity distribution parameters) from a revenue production function. An empirical application to French firm-level data proves our identified source of mismeasurement to be economically relevant.

Our work highlights the possibilities and advantages of relying on the revenue production function to identify aggregate trade elasticities and resulting Gains From Trade. Estimating a revenue production function allows us to exploit the panel dimension of firm-level data from a single country to identify both the productivity distribution and elasticity of substitution while controlling for transitory shocks to demand and supply. Moreover, data to estimate such a production function is easily accessible for multiple countries through, for instance, the Orbis database.

The theoretical elaborations in this paper demonstrate that, under relatively light assumptions on the distribution of transitory shocks, the existence of firm-level shocks is not problematic for prevalent static firm-level heterogeneous models that do not feature these shocks. But, when quantifying these models, one does need to ensure transitory shocks to demand and supply are controlled for in order to obtain unbiased model inputs.

The results in this paper also point to future elaborations on the influence of firm-level shocks on aggregate trade statics. Whereas we demonstrate the role of *transitory* shocks in a static framework to allow comparison with current literature, an extension to a dynamic framework would allow to investigate the influence of *persistent* shocks in productivity on aggregate trade outcomes. Further research could also focus on the specification of transitory shocks. One could, for instance, extend the current definition of the shocks to allow for product-market specific deviations. Bas et al. (2017), for instance, specify their demand-side deviation as a firm-market specific cost to reach a market originating from, among others, differences in internal knowledge on how to reach consumers in that market. Sager and Timoshenko (2019, 2020) argue that the demand-side deviation represents, observed or unobserved, variety-specific demand that firms need to learn over time through market participation. The current specification captures the firm-level average of such market-specific shocks. Lastly, data limitations prohibit us to differentiate between supply and demand shocks and does not allow for firm-level markups. All these possibilities set out interesting research paths for expanding the current methodology to continuously improve our measurements and understanding of firm-level trade elasticities using firm-level data.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. *Econometrica 83*(6), 2411–2451.

Adão, R., C. Arkolakis, and S. Ganapati (2020). Aggregate implications of firm heterogeneity: A nonparametric analysis of monopolistic competition trade models. Working paper 28081, National Bureau of Economic Research.

Alessandria, G. and H. Choi (2014). Establishment heterogeneity, exporter dynamics, and the effects of trade liberalization. *Journal of International Economics 94*(2), 207–223.

Amand, M. and F. Pelgrin (2016). Pareto Distribution in International Trade: Hard to Identify, Easy to Estimate. mimeo.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New Trade Models, Same Old Gains? *American Economic Review 102*(1), 94–130.

Asker, J., A. Collard-Wexler, and J. De Loecker (2017). Market Power, Production (Mis) Allocation, and OPEC. *Economic Research Initiatives at Duke (ERID) Working Paper* (256).

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science 293*(5536), 1818–1820.

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*, 1 – 19.

Bee, M. and S. Schiavo (2018). Powerless: gains from trade when firm productivity is not Pareto distributed. *Review of World Economics 154*(1), 15–45.

Berthou, A. and L. Fontagné (2016). Variable trade costs, composition effects and the intensive margin of trade. *The World Economy 39*(1), 54–71.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Carvalho, V. M. and B. Grassi (2019). Large firm dynamics and the business cycle. *American Economic Review 109*(4), 1375–1425.

Cernat, L. (2014). Towards "Trade Policy Analysis 2.0". From national comparative advantage to firm-level data. Chief Economist Note 4, European Central Bank.

Chaney, T. (2008, September). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Costinot, A. and A. Rodríguez-Clare (2014). Chapter 4 - trade theory with numbers: Quantifying the consequences of globalization. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 197–261. Elsevier.

Das, S., M. J. Roberts, and J. R. Tybout (2007). Market entry costs, producer heterogeneity, and export dynamics. *Econometrica 75*(3), 837–873.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica 79*(5), 1407–1451.

Dewitte, R., M. Dumont, G. Rayp, and P. Willemé (forthcoming). Unobserved Heterogeneity in the Productivity Distribution and Gains From Trade. *Canadian Journal of Economics*.

di Giovanni, J. and A. A. Levchenko (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy 120*(6), 1083–1132.

di Giovanni, J. and A. A. Levchenko (2013). Firm entry, trade, and welfare in zipf's world. *Journal of International Economics 89*(2), 283–296.

di Giovanni, J., A. A. Levchenko, and R. Rancière (2011). Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics 85*(1), 42–52.

Eaton, J., S. S. Kortum, and S. Sotelo (2012). International trade: Linking micro and macro. NBER Working Paper 17864, National bureau of economic research.

Egger, P. H., K. Erhardt, and S. Nigai (2020). Empirical productivity distributions and international trade. Discussion Paper DP15160, CEPR.

Fernandes, A. M., P. J. Klenow, S. Meleshchuk, M. D. Pierola, and A. Rodríguez-Clare (2018). The Intensive Margin in Trade: Moving Beyond Pareto. Policy Research working paper WPS 8625, World Bank Group.

Fitzgerald, D. and S. Haller (2018). Exporters and shocks. *Journal of International Economics 113*, 154–171.

Fontagné, L., P. Martin, and G. Orefice (2018). The international elasticity puzzle is worse than you think. *Journal of International Economics 115*, 115–129.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Halpern, L., M. Koren, and A. Szeidl (2015). Imported inputs and productivity. *American Economic Review 105*(12), 3660–3703.

Head, K., T. Mayer, and M. Thoenig (2014). Welfare and Trade without Pareto. *American Economic Review 104*(5), 310–316.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Impullitti, G., A. A. Irarrazabal, and L. D. Opromolla (2013). A theory of entry into and exit from export markets. *Journal of International Economics 90*(1), 75–90.

Javorcik, B. S. (2004). Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages. *The American Economic Review 94*(3), 605–627.

Kalemli-Ozcan, S., B. Sorensen, C. Villegas-Sanchez, V. Volosovych, and S. Yesiltas (2015). How to construct nationally representative firm level data from the orbis global database: New facts and aggregate implications. Working paper 21558, National Bureau of Economic Research.

Kasahara, H. and B. Lapham (2013). Productivity and the decision to import and export: Theory and evidence. *Journal of International Economics 89*(2), 297–316.

Klette, T. J. and Z. Griliches (1996). The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. *Journal of Applied Econometrics 11*(4), 343–61.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Melitz, M. J. and S. J. Redding (2014). Chapter 1 - Heterogeneous Firms and Trade. In E. H. Gita Gopinath and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 1–54. Elsevier.

Melitz, M. J. and S. J. Redding (2015). New Trade Models, New Welfare Implications. *American Economic Review 105*(3), 1105–46.

Merlevede, B., M. De Zwaan, K. Lenaerts, and V. Purice (2015). Multinational networks, foreign, and domestic firms in europe. Working Paper 900, Ghent University.

Mrázová, M., J. P. Neary, and M. Parenti (2021). Sales and markup dispersion: Theory and empirics. *Econometrica 89*(4), 1753–1788.

Nigai, S. (2017). A tale of two tails: Productivity distribution and the gains from trade. *Journal of International Economics 104*, 44–62.

Piveteau, P. and G. Smagghue (2019). Estimating firm product quality using trade data. *Journal of International Economics 118*, 217–232.

Rivers, D. A. (2010). Are Exporters More Productive than Non-Exporters? mimeo, University of Western Ontario.

Ruhl, K. J. and J. L. Willis (2017). New exporter dynamics. *International Economic Review 58*(3), 703–726.

Sager, E. and O. A. Timoshenko (2019). The double emg distribution and trade elasticities. *Canadian Journal of Economics 52*(4).

Sager, E. and O. A. Timoshenko (2020). Demand uncertainty, selection, and trade. mimeo.

Sickles, R. C. and V. Zelenyuk (2019). *Measurement of Productivity and Efficiency.* Cambridge University Press.

Tsionas, E. G. and S. C. Kumbhakar (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. *Journal of Applied Econometrics 29*(1), 110–132.

Van Biesebroeck, J. (2007). Robustness of productivity estimates. *The Journal of Industrial Economics 55*(3), 529–569.

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, and M. Larch (2016). *An advanced guide to trade policy analysis: The structural gravity model.* World Trade Organization.

Online Appendix to "Gains From Trade: Demand, Supply and Idiosyncratic Shocks"

Ruben Dewitte, Bruno Merlevede and Glenn Rayp*

12th May 2022

*Dewitte: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be, Merlevede: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, bruno.merlevede@ugent.be, Rayp: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, glenn.rayp@ugent.be. We thank Luca David Opromolla, Gonzague Vannoorenberghe, Gerdie Everaert, Marijn Verschelde, Angelos Theodorakopoulos and participants of the 2017 ETSG conference in Florence, the 2018 FEB research day in Ghent, the 2018 Compnet conference in Halle and the 2019 ETSG conference in Bern for their helpful comments and suggestions. Ruben Dewitte acknowledges financial support from Research Foundation Flanders (FWO) grant 12B8822N.

Appendix A Additional Figures and Tables

A.1 Figures

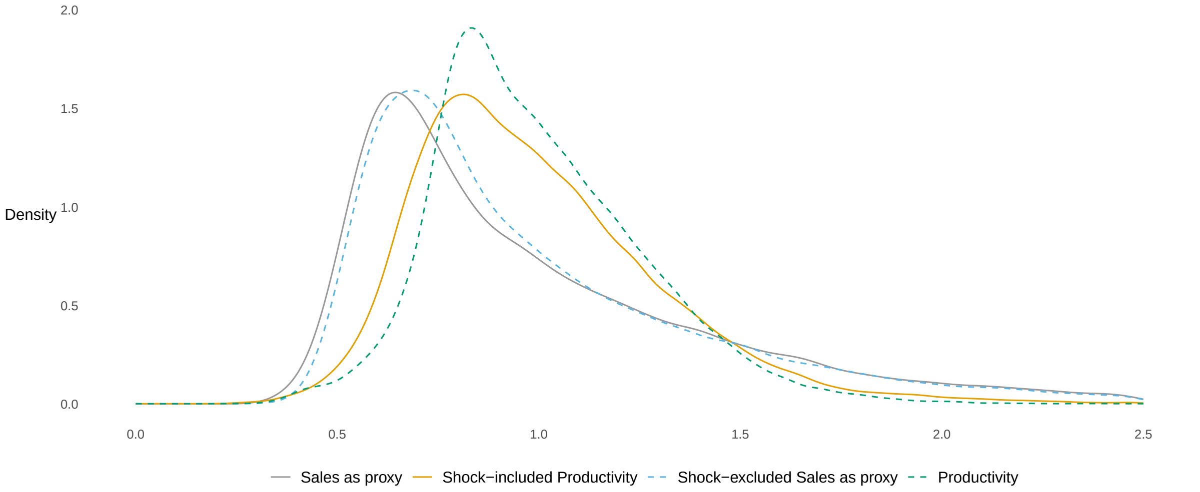

Figure 1: Nonparametric kernel density of domestic sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-included productivity and productivity in the year 2006.
Note: All variables are demeaned.

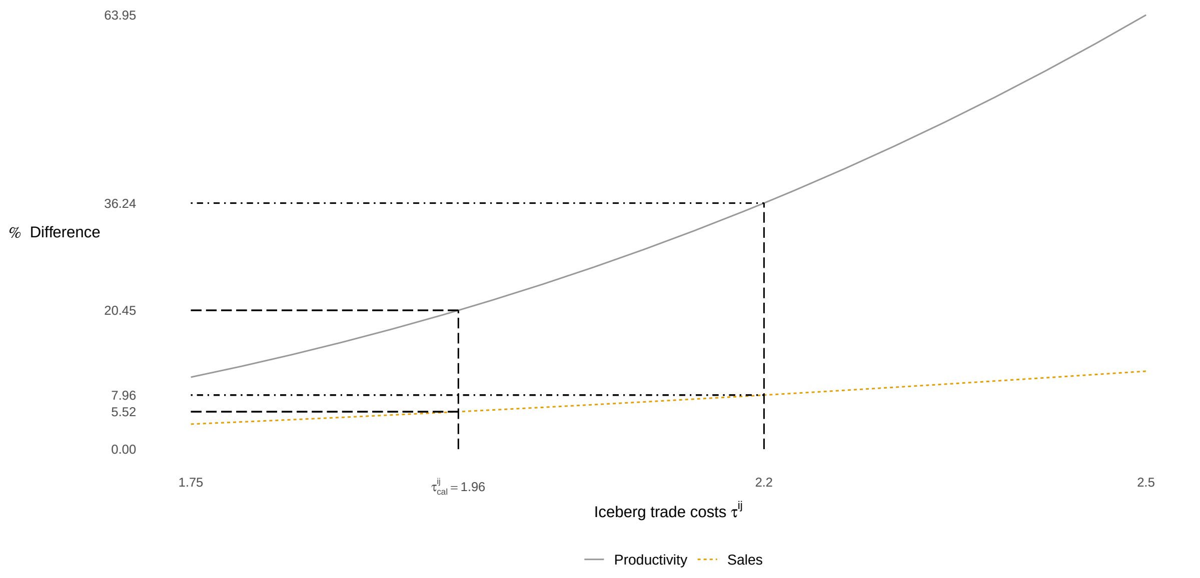

Figure 2: Idiosyncratic shocks-induced overestimation (in %) of welfare gains from a reduction in variable trade costs relative to Autarky ($\tau = 10 \rightarrow \tau'$)

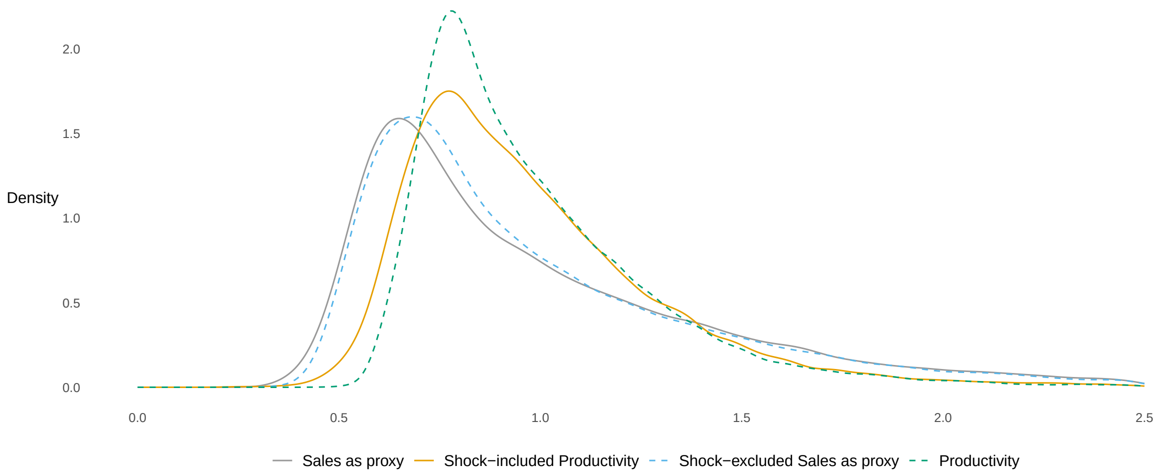

Figure 3: Nonparametric kernel density of domestic sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-included productivity and productivity in the year 2006 obtained from assuming a Translog production function.
Note: All variables are demeaned.

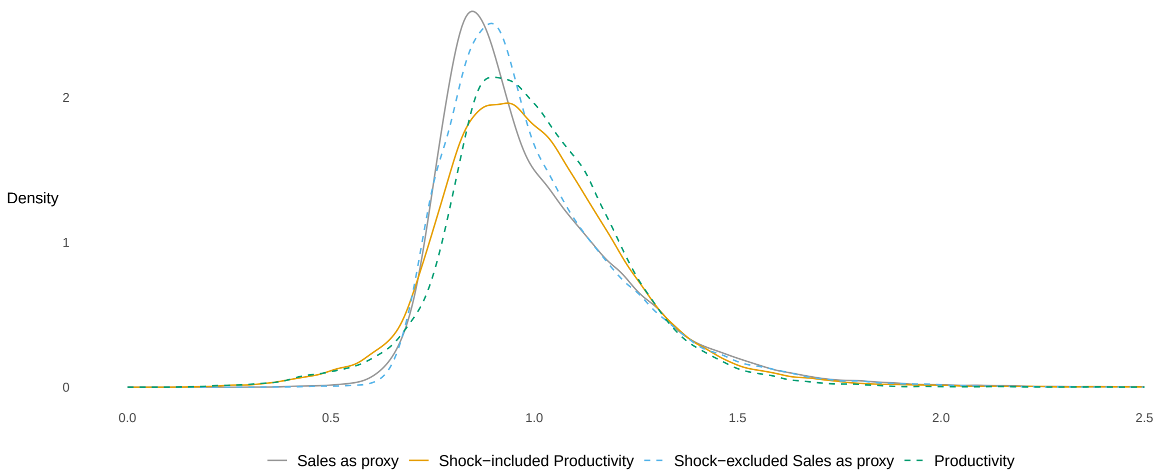

Figure 4: Nonparametric kernel density of domestic sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-included productivity and productivity in the year 2006 obtained from an uncleaned data sample.
Note: All variables are demeaned.

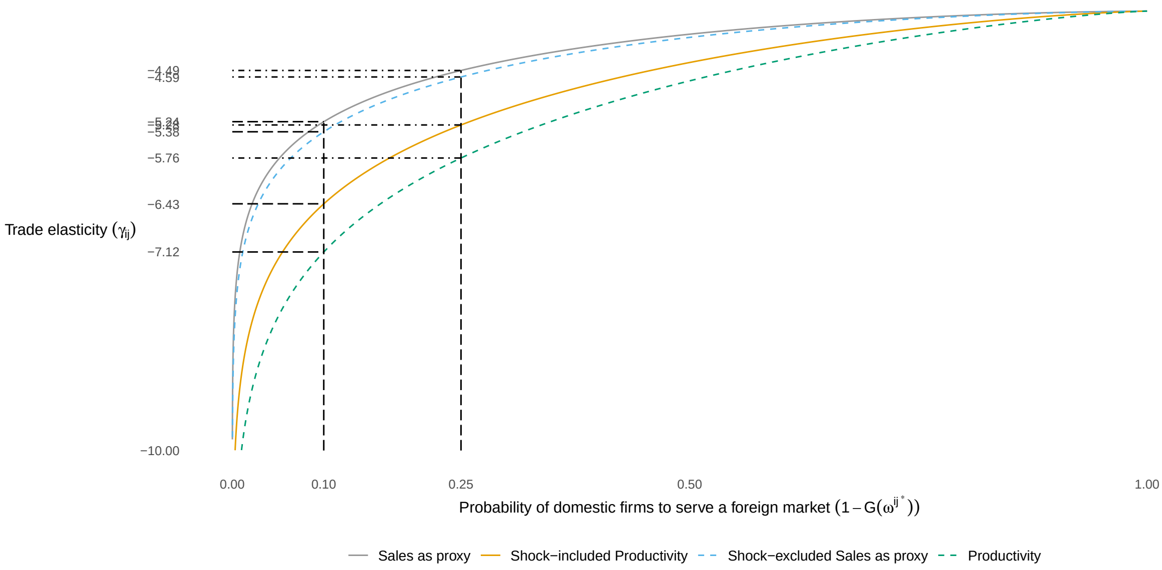

Figure 5: Trade elasticities for sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-included productivity and productivity in the year 2006 obtained from assuming a Translog production function.

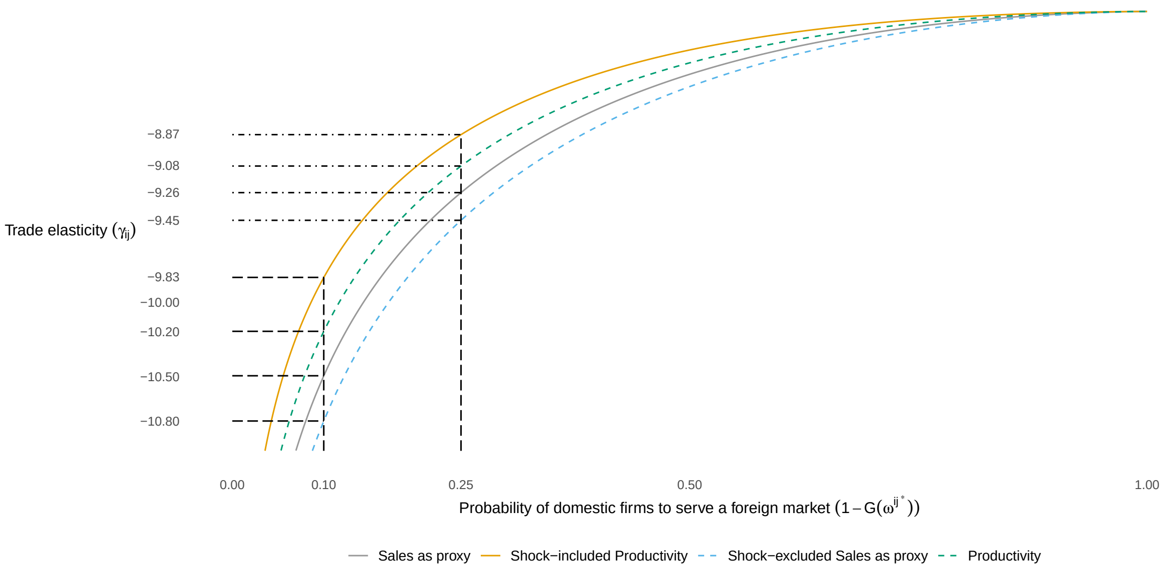

Figure 6: Trade elasticities for sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-included productivity and productivity in the year 2006 obtained from an uncleaned data sample.

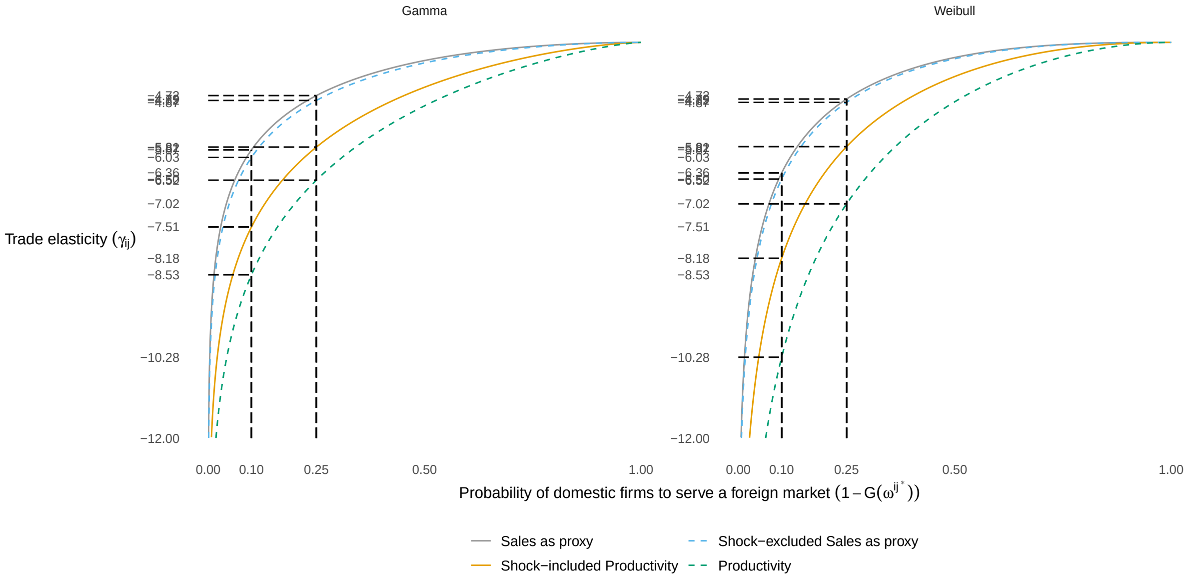

Figure 7: Trade elasticities for sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-included productivity and productivity in the year 2006 for different distributional assumptions.

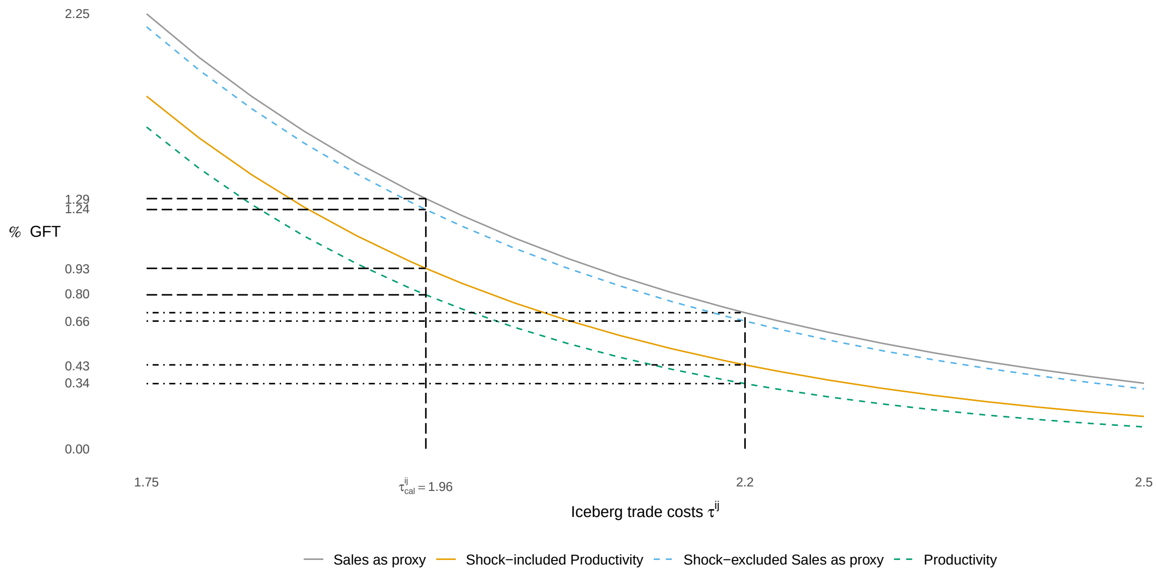

Figure 8: Percentage welfare gains om a reduction in variable trade costs relative to Autarky ($\tau = 10 \rightarrow \tau'$) assuming a Translog production function

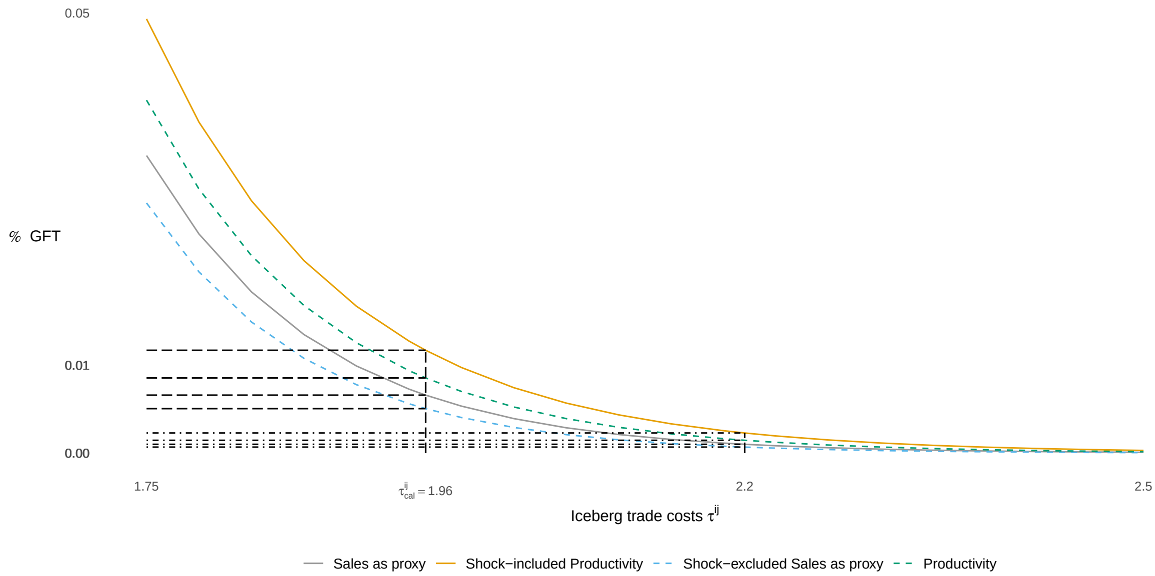

Figure 9: Percentage welfare gains om a reduction in variable trade costs relative to Autarky ($\tau = 10 \rightarrow \tau'$) for an uncleaned data sample

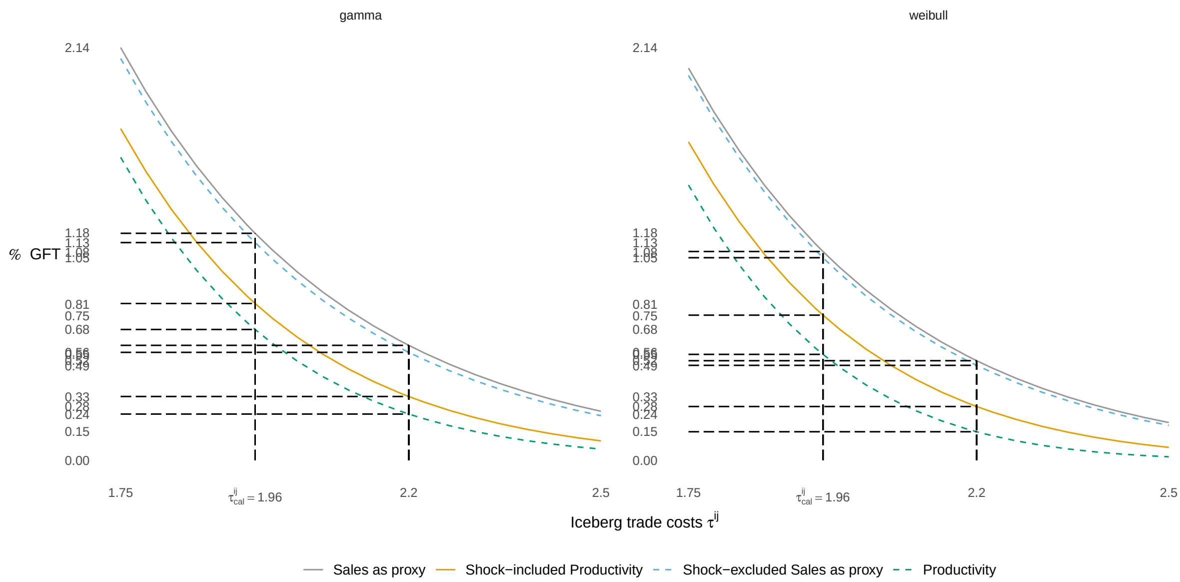

Figure 10: Percentage welfare gains om a reduction in variable trade costs relative to Autarky ($\tau = 10 \rightarrow \tau'$) for different distributional assumption

A.2 Tables

Table 1: Summary statistics

Variable (in logs)	Obs.	Nr. Firms	mean	sd	min	max
Sales	379765	86959	14.13	1.46	10.57	19.72
Domestic sales[a]	379765	86959	14.02	1.40	10.56	18.67
Value added[b]	379765	86959	13.66	1.39	4.35	19.64
Domestic value added[a,b]	379765	86959	13.54	1.33	4.35	18.50
Capital	379765	86959	11.37	1.85	6.91	17.09
Labor	379765	86959	2.50	1.21	0.69	6.46
Materials	379765	86959	12.85	1.82	7.41	18.40
Aggregate Domestic sales[c]	379765	86959	16.19	0.14	15.95	16.42

Notes: All financial variables are displayed in real terms. [a]Domestic variables are obtained as the difference between the total and the exporting share of those variables. [b]Value added is obtained as the difference between sales and material inputs. [c]Aggregate variables are constructed as the market-share weighted sum of the underlying variables.

Table 2: Translog production function estimation results.

	$\frac{\sigma-1}{\sigma}\boldsymbol{\beta}$			$\boldsymbol{\beta}$			
	Capital	**Labor**	**RTS**	**Capital**	**Labor**	**RTS**	**Elasticity of Substitution**
NLLS	0.151	0.871	1.022	0.183	1.051	1.233	5.836
	(0.002)	(0.007)	(0.009)	(0.002)	(0.007)	(0.008)	(0.180)
ACF	0.142	0.806	0.947	0.181	1.027	1.208	4.633
	(0.002)	(0.008)	(0.009)	(0.002)	(0.008)	(0.008)	(0.126)

Notes: Standard errors displayed between brackets are obtained from wild bootstrap clustered at the firm level with 99 repliations. Estimates obtained from French firm-level database over the years 1998–2006 with 379,765 observations from 86,959 firms.

Table 3: Production function estimation results from an uncleaned data sample.

	$\frac{\sigma-1}{\sigma}\boldsymbol{\beta}$			$\boldsymbol{\beta}$			
	Capital	**Labor**	**RTS**	**Capital**	**Labor**	**RTS**	**Elasticity of Substitution**
NLLS	0.162	0.838	0.999	0.187	0.969	1.156	7.364
	(0.002)	(0.006)	(0.007)	(0.002)	(0.005)	(0.006)	(0.243)
ACF	0.155	0.821	0.976	0.174	0.923	1.097	9.046
	(0.002)	(0.005)	(0.006)	(0.002)	(0.004)	(0.005)	(0.313)

Notes: Standard errors displayed between brackets are obtained from wild bootstrap clustered at the firm level with 99 repliations. Estimates obtained from French firm-level database over the years 1998–2006 with 474,044 observations from 99,465 firms.

Table 4: Variance and quantile ratios for sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-augmented productivity, and productivity in the year 2006 obtained from assuming a Translog production function.

Variable	Variance	75/25	90/10	95/5	99/1
Sales as proxy $\left(\left(x_{bt}^{ii}\right)^{\frac{1}{\hat{\sigma}-1}}\right)$	0.218	1.838	2.911	3.786	5.780
Shock-excluded Sales as proxy $\left(\left(x_{bt}^{ii}\right)^{\frac{1}{\hat{\sigma}-1}} e^{-\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$	0.208	1.767	2.811	3.627	5.410
Shock-included Productivity $\left(e^{\hat{\omega}_{bt}+\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$	0.161	1.505	2.135	2.656	4.598
Productivity $\left(e^{\hat{\omega}_{bt}}\right)$	0.122	1.447	1.953	2.377	3.653

Notes: Values obtained from sample of 34,339 French firms in 2006.

Table 5: Variance and quantile ratios for sales as a proxy for productivity, shock-excluded sales as a proxy for productivity, shock-augmented productivity, and productivity in the year 2006 from an uncleaned data sample.

Variable	Variance	75/25	90/10	95/5	99/1
Sales as proxy $\left(\left(x_{bt}^{ii}\right)^{\frac{1}{\hat{\sigma}-1}}\right)$	0.056	1.337	1.700	1.978	2.743
Shock-excluded Sales as proxy $\left(\left(x_{bt}^{ii}\right)^{\frac{1}{\hat{\sigma}-1}} e^{-\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$	0.051	1.312	1.688	1.953	2.580
Shock-included Productivity $\left(e^{\hat{\omega}_{bt}+\frac{\epsilon_{bt}^{T}}{\hat{\sigma}-1}}\right)$	0.059	1.333	1.727	2.088	3.567
Productivity $\left(e^{\hat{\omega}_{bt}}\right)$	0.046	1.301	1.646	1.993	3.402

Notes: Values obtained from sample of 44,151 French firms in 2006.

Appendix B Heterogeneous firms model

We specify an open economy model with persistent firm-level uncertainty in initial and future productivity as well as non-persistent firm-level uncertainty in the production process originating from demand and supply. The core elements are Melitz (2003) augmented with the stochastic evolution of firm productivity as in Hopenhayn (1992), while the introduction of uncertainty surrounding realized demand and supply is embedded in the structural estimation literature (see for instance Das et al. (2007); De Loecker (2011); Kasahara and Lapham (2013); Gandhi et al. (2020)). This will provide us with a general framework to evaluate empirical methods that deduce firm-level heterogeneity from sales data.

In a second stage, we reduce the productivity dynamics to certainty in future productivity (Melitz, 2003) while featuring uncertainty in realized supply and demand.[1] Eliminating the dynamics as such has the advantage of resulting in clear analytical expressions for the equilibrium variables. Moreover, it allows us to easily demonstrate the influence of transitory uncertainty on the trade elasticity and GFT compared to the predominant Melitz (2003)-model (see, for instance Head et al. (2014); Melitz and Redding (2015); Nigai (2017); Bee and Schiavo (2018)).

B.1 Setup

Demand Consumer preferences in country $j \in J$ are defined over a continuum of horizontally differentiated varieties originating from country $i \in I$ ($\varpi \in \Omega^i$) and are assumed to take the Constant Elasticity of Substitution (CES) utility (U) form at time t,

$$U_t^j = \left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} \, d\varpi \right)^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

with σ the elasticity of substitution between varieties and $e^{\nu_t(\varpi)}$ a variety-specific demand shock independently and identically distributed across varieties and time. Let the aggregate expenditure in country j be R^j, and the price of a good p_t^{ij}, then the utility maximixation problem is

$$\begin{aligned} \max_{y_t^{ij}(\varpi)} U_t^j &= \left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} \, d\varpi \right)^{\frac{\sigma}{\sigma-1}} \\ & \qquad s.t. \quad \sum_{i=1}^{I} \int_{\varpi \in \Omega^i} p_t^{ij}(\varpi) y_t^{ij}(\varpi) d\varpi \leq R^j. \end{aligned} \tag{2}$$

The Lagrangian is:

[1] For a discussion on the implications of productivity dynamics on the economy, see Impullitti et al. (2013); Ruhl and Willis (2017).

$$\mathcal{L} = \left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} d\varpi \right)^{\frac{\sigma}{\sigma-1}} - \lambda \left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} p_t^{ij}(\varpi) y_t^{ij}(\varpi) d\varpi - R^j \right), \tag{3}$$

and First Order Conditions (FOC) are:

1.

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial y_t^{ij}(\varpi)} = 0 \\ &\Leftrightarrow \quad \left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} d\varpi \right)^{\frac{1}{\sigma-1}} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{-1}{\sigma}} = \lambda p_t^{ij}(\varpi) \end{aligned} \tag{4}$$

2.

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{I} \int_{\varpi \in \Omega^i} p_t^{ij}(\varpi) y_t^{ij}(\varpi) d\varpi = R^j \tag{5}$$

Exponentiating the first FOC by $(1-\sigma)$ and aggregating, we obtain

$$\begin{aligned} &\left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} d\varpi \right)^{-1} e^{\frac{1-\sigma}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} = \lambda^{1-\sigma} p_t^{ij}(\varpi)^{1-\sigma} \\ &\Leftrightarrow \\ &\left(\sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} d\varpi \right)^{-1} \sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\frac{1}{\sigma}\nu_t(\varpi)} y_t^{ij}(\varpi)^{\frac{\sigma-1}{\sigma}} d\varpi = \lambda^{1-\sigma} \sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\nu_t(\varpi)} p_t^{ij}(\varpi)^{1-\sigma} d\varpi \\ &\Leftrightarrow \\ &(P_t^j)^{-1} = \lambda, \end{aligned} \tag{6}$$

where and P_t^j is the CES aggregate price index in country j at time t:

$$(P_t^j)^{1-\sigma} = \sum_{i=1}^{I} \int_{\varpi \in \Omega^i} e^{\nu_t(\varpi)} p_t^{ij}(\varpi)^{1-\sigma} d\varpi. \tag{7}$$

Plugging the expression for λ back in the FOC provides us with the optimal consumption and expenditure decisions over the individual varieties:

$$\frac{y_t^{ij}(\varpi)}{Y_t^j} = \left(\frac{p_t^{ij}(\varpi)}{P_t^j} \right)^{-\sigma} e^{\nu_t(\varpi)}, \tag{8}$$

where the set of varieties consumed is considered as an aggregate good $Y_t^j \equiv U_t^j$.

Supply There is a continuum of businesses ($b \in B$) which choose to supply a distinct horizontally-differentiated variety. They are heterogeneous in terms of their productivity $\omega_{bt} \in [0, \infty]$ drawn from the unconditional Cumulative Distribution Function (CDF) $G(\omega_{bt})$ after paying a fixed cost fe_t^i to enter the market.[2] The firm's productivity follows a Markov process independent across firms with conditional distribution $G(\omega_{bt+1}|\mathcal{I}_{bt})$ such that[3]

$$e^{\omega_{bt+1}} = \mathbb{E}_\omega \left[e^{\omega_{bt+1}}|\mathcal{I}_{bt}\right] e^{\eta_{bt+1}}. \tag{9}$$

Productivity at time $t+1$ is thus a function of the information set of the firm at time t, $\mathcal{I}_{bt}$ with $\{\omega_{bt}\} \in \mathcal{I}_{bt}$, and a productivity shock $\eta_{bt} + 1$ (see below for a summary of the timing assumptions of the model). The productivity distribution is known to firms and stochastically increasing in ω_{bt}. The information set $\mathcal{I}_{bt}$ is a set of random variables that contains all the information a firm relies on to solve its period t decisions problems.

Production relies a composite factor of production $\boldsymbol{A}_{bt}^{ij}(\boldsymbol{\beta})$ (Melitz and Redding, 2014) consisting of Z fixed (F) and V variable (L) factors of production respectively: $\{F_{1t}, \ldots, F_{Zt}, L_{1t}, \ldots L_{Vt}\}$. Variable production factors can be adjusted every time period after the observation set $\mathcal{I}_{bt-1}$ is observed. Fixed production factors, on the other hand, take one time period to adjust.[4] These production factors are combined under Cobb-Douglas Constant Returns to Scale technology with respective factor intensities β_{iz}, β_{iv}:

$$y_{bt}^{ij} = q_{bt}^{ij} e^{\epsilon_{bt}} = \prod_{z=1}^{Z} \prod_{v=1}^{V} (F_{bzt}^{ij})^{\beta_z^i} (L_{bvt}^{ij})^{\beta_v^i} e^{\omega_{bt} + \epsilon_{bt}}, \qquad \sum_{z=1}^{Z} \beta_{iz} + \sum_{v=1}^{V} \beta_{iv} = 1, \tag{10}$$

subject to shocks to the production function $e^{\epsilon_{bt}}$, which are independent and identically distributed across firms and time. Supply of the production factors to the individual firm is perfectly elastic, so that firms are effectively price (W_{zt}^i, W_{vt}^i) takers on the input markets. Firms from country i have to pay a fixed cost f_t^{ij} to produce goods destined for country j denominated in final goods,[5] and variable iceberg trade costs, $\tau_t^{ij} > 1$, denominated in units of labour of the origin country.

As production factors differ in their timing of adjustment, we differentiate between long- and short-run profit maximization. The long-run expected profit maximization optimizes the quantity of fixed production factors for time t based on the information provided in $t-1$, $\mathcal{I}_{bt-1}$:

[2]We follow Asker et al. (2017) in differentiating all fixed costs from factors of production. "In their financial statements, firms report overhead costs as Selling, General and Administrative Expenses (SG&A). These expenses are not directly related to production, and include sales, advertising, marketing, executive compensation, ... and can in part be interpreted as expenses on intangible capital." (Asker et al., 2017, p. 4). We assume all fixed cost expenses are equally distributed within the source market.

[3]$\mathbb{E}_x [\ldots] = \int \ldots f(x)dx$

[4]For simplicity, we assume fixed production factors have no dynamic implications.

[5]Similar to the fixed entry costs, fixed production costs are due in the domestic market in final goods rather than in production factors.

$$
\begin{aligned}
\max_{\boldsymbol{F}^{ij}_{bzt}} \mathbb{E}_{\eta,\nu,\epsilon}\left[\pi^{ij}_{bt}|\mathcal{I}_{bt-1}\right] &= \max_{\boldsymbol{F}^{ij}_{bzt}} \mathbb{E}_{\eta,\nu,\epsilon}\left[p^{ij}_{bt}y^{ij}_{bt} - f^{ij}_t - \tau^{ij}_t\left(\sum_{z=1}^{Z} F^{ij}_{bzt}W^i_{zt} - \sum_{v=1}^{V} L^{ij}_{bvt}W^i_{vt}\right)\middle|\mathcal{I}_{bt-1}\right] \\
&= \max_{\boldsymbol{F}^{ij}_{bzt}} \left(Y^j_t\right)^{\frac{1}{\sigma}} P^j_t \mathbb{E}_\eta\left[\left(q^{ij}_{bt}\right)^{\frac{\sigma-1}{\sigma}}\middle|\mathcal{I}_{bt-1}\right]\mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt-1}\right] \\
&\qquad - f^{ij}_t - \tau^{ij}_t\left(\sum_{z=1}^{Z} F^{ij}_{bzt}W^i_{zt} - \sum_{v=1}^{V} L^{ij}_{bvt}W^i_{vt}\right). \qquad (11)
\end{aligned}
$$

From the First-order conditions, we obtain the optimal quantity of fixed production factors:

$$
\begin{aligned}
0 &= \frac{\partial \mathbb{E}_{\eta,\nu,\epsilon}\left[\pi^{ij}_{bt}|\mathcal{I}_{bt-1}\right]}{\partial F^{ij}_{bzt}} \\
&= \frac{\sigma-1}{\sigma}\mathbb{E}_\eta\left[\left(q^{ij}_{bt}\right)^{-\frac{1}{\sigma}}\middle|\mathcal{I}_{bt-1}\right]\frac{\beta^i_z\mathbb{E}_\eta\left[q^{ij}_{bt}\middle|\mathcal{I}_{bt-1}\right]}{F^{ij}_{bzt}}\left(Y^j_t\right)^{\frac{1}{\sigma}} P^j_t\mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt-1}\right] - \tau^{ij}_t W^i_{zt} \\
&\Leftrightarrow
\end{aligned}
$$

$$
F^{ij}_{bzt} = \frac{\sigma-1}{\sigma}\mathbb{E}_\eta\left[\left(q^{ij}_{bt}\right)^{-\frac{1}{\sigma}}\middle|\mathcal{I}_{bt-1}\right]\frac{\beta^i_z\mathbb{E}_\eta\left[q^{ij}_{bt}\middle|\mathcal{I}_{bt-1}\right]}{\tau^{ij}_t W^i_{zt}}\left(Y^j_t\right)^{\frac{1}{\sigma}} P^j_t\mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt-1}\right]. \qquad (12)
$$

The short-run expected profit maximization, then, optimizes the quantity of variable production factors for time t given the fixed production factors ($\overline{F}^{ij}_{bvt}$) and based on the information provided in t, $\mathcal{I}_{bt}$:

$$
\begin{aligned}
\max_{\boldsymbol{L}^{ij}_{bvt}} \mathbb{E}_{\nu,\epsilon}\left[\pi^{ij}_{bt}|\mathcal{I}_{bt}\right] &= \max_{\boldsymbol{L}^{ij}_{bvt}} \mathbb{E}_{\nu,\epsilon}\left[p^{ij}_{bt}\overline{y}^{ij}_{bt} - f^{ij}_t - \tau^{ij}_t\left(\sum_{z=1}^{Z} \overline{F}^{ij}_{bzt}W^i_{zt} - \sum_{v=1}^{V} L^{ij}_{bvt}W^i_{vt}\right)\middle|\mathcal{I}_{bt}\right] \\
&= \max_{\boldsymbol{L}^{ij}_{bvt}} \left(\overline{q}^{ij}_{bt}\right)^{\frac{\sigma-1}{\sigma}}\left(Y^j_t\right)^{\frac{1}{\sigma}} P^j_t\mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt}\right] \\
&\qquad - f^{ij}_t - \tau^{ij}_t\left(\sum_{z=1}^{Z} \overline{F}^{ij}_{bzt}W^i_{zt} - \sum_{v=1}^{V} L^{ij}_{bvt}W^i_{vt}\right). \qquad (13)
\end{aligned}
$$

The first-order conditions allow us to deduce optimal quantity of variable production factors:

$$
\begin{aligned}
0 &= \frac{\partial \mathbb{E}_{\nu,\epsilon}\left[\pi_{bt}^{ij}|\mathcal{I}_{bt}\right]}{\partial L_{bvt}^{ij}} \\
&= \frac{\sigma-1}{\sigma}\left(\overline{q}_{bt}^{ij}\right)^{-\frac{1}{\sigma}}\frac{\beta_v^i \overline{q}_{bt}^{ij}}{L_{bvt}^{ij}}\left(Y_t^j\right)^{\frac{1}{\sigma}} P_t^j \mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt}\right] - \tau_t^{ij} W_{zt}^i \\
&\Leftrightarrow \\
L_{bvt}^{ij} &= \frac{\sigma-1}{\sigma}\left(\overline{q}_{bt}^{ij}\right)^{-\frac{1}{\sigma}}\frac{\beta_v^i \overline{q}_{bt}^{ij}}{\tau_t^{ij} W_{vt}^i}\left(Y_t^j\right)^{\frac{1}{\sigma}} P_t^j \mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt}\right]. \qquad (14)
\end{aligned}
$$

Completing the production function (eq. 10) with the optimal input mix allows us to obtain an expression for the optimal quantity:[6,7]

$$
\begin{aligned}
q_{bt}^{ij} &= \prod_{z=1}^{Z}\prod_{v=1}^{V}(F_{bzt}^{ij})^{\beta_z^i}(L_{bvt}^{ij})^{\beta_v^i} e^{\omega_{bt}} \\
&= \frac{\sigma-1}{\sigma}\left(Y_t^j\right)^{\frac{1}{\sigma}} P_t^j \frac{1}{\tau_t^{ij}}\left(\prod_{z=1}^{Z}\prod_{v=1}^{V}\frac{(\beta_v^i)^{\beta_v^i}(\beta_z^i)^{\beta_z^i}}{(W_{vt}^i)^{\beta_v^i}(W_{zt}^i)^{\beta_z^i}}\right) \\
&\quad \times \mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt-1}\right]^{\sum_{z=1}^{Z}\beta_z^i}\mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt}\right]^{\sum_{v=1}^{V}\beta_v^i} \\
&\quad \times \mathbb{E}_{\eta}\left[\left(q_{bt}^{ij}\right)^{\frac{\sigma-1}{\sigma}}\middle|\mathcal{I}_{bt-1}\right]^{\sum_{z=1}^{Z}\beta_z^i}\left[\left(q_{bt}^{ij}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\sum_{v=1}^{V}\beta_v^i} e^{\omega_{bt}} \\
&= \frac{\sigma-1}{\sigma}\left(Y_t^j\right)^{\frac{1}{\sigma}} P_t^j \frac{(\beta_v^i)^{\sum_{v=1}^{V}\beta_v^i}(\beta_z^i)^{\sum_{z=1}^{Z}\beta_z^i}}{\tau_t^{ij}(W_{vt}^i)^{\sum_{v=1}^{V}\beta_v^i}(W_{zt}^i)^{\sum_{z=1}^{Z}\beta_z^i}} \\
&\quad \times \mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt}\right] \\
&\quad \times \left(\frac{\mathbb{E}_{\eta}\left[e^{\eta_{bt}}\middle|\mathcal{I}_{bt-1}\right]}{e^{\eta_{bt}}}\right)^{\frac{\sigma-1}{\sigma}\sum_{z=1}^{Z}\beta_z^i}\left(q_{bt}^{ij}\right)^{\frac{\sigma-1}{\sigma}} e^{\omega_{bt}} \\
&= \left(\frac{\sigma-1}{\sigma}\frac{e^{\omega_{bt}}}{\tau_t^{ij}W_t^i}\left(Y_t^j\right)^{\frac{1}{\sigma}}(P_t^j)\mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_{bt}}{\sigma}+\frac{\sigma-1}{\sigma}\epsilon_{bt}}\middle|\mathcal{I}_{bt}\right]\right)^{\sigma}\left(\frac{\mathbb{E}_{\eta}\left[e^{\eta_{bt}}\middle|\mathcal{I}_{bt-1}\right]}{e^{\eta_{bt}}}\right)^{(\sigma-1)\sum_{z=1}^{Z}\beta_z^i} \\
&= \left(\frac{\sigma-1}{\sigma}\frac{e^{\omega_{bt}}}{\tau_t^{ij}W_t^i}\right)^{\sigma} Y_t^j (P_t^j)^{\sigma}\mathbb{E}_{\varepsilon^T}\left[e^{\varepsilon_{bt}^T}\middle|\mathcal{I}_{bt}\right]^{\sigma}\frac{\mathbb{E}_{\varepsilon^P}\left[e^{\varepsilon_{bt}^P}\middle|\mathcal{I}_{bt-1}\right]^{\sigma}}{e^{\sigma\varepsilon_{bt}^P}} \\
&= \left(\frac{\sigma-1}{\sigma}\frac{e^{\omega_{bt}}}{\tau_t^{ij}W_t^i}\right)^{\sigma} Y_t^j (P_t^j)^{\sigma}\mathbb{E}_{\varepsilon^T}\left[e^{\varepsilon_{bt}^T}\right]^{\sigma}\frac{\mathbb{E}_{\varepsilon^P}\left[e^{\varepsilon_{bt}^P}\right]^{\sigma}}{e^{\sigma\varepsilon_{bt}^P}}, \qquad (15)
\end{aligned}
$$

[6]Note that if all inputs would be variable, this optimum quantity would reduce to

$$
q_b^{ij} = \left(\frac{\sigma-1}{\sigma}\frac{e^{\omega}}{\tau^{ij}W^i}\right)^{\sigma} Y^j \left(P^j\right)^{\sigma}\mathbb{E}_{\nu,\epsilon}\left[e^{\epsilon_b^T}\right]^{\sigma}
$$

[7]We rely on the i.i.d. nature of uncertainty to reduce the expectation term to a constant $\mathbb{E}_{\nu,\epsilon}\left[e^{\varepsilon_{bt}^T}\middle|\mathcal{I}_{bt}\right] = \mathbb{E}_{\nu,\epsilon}\left[e^{\varepsilon_{bt}^T}\right]$, $\mathbb{E}_{\eta}\left[e^{\varepsilon_{bt}^P}\middle|\mathcal{I}_{bt}\right] = \mathbb{E}_{\eta}\left[e^{\varepsilon_{bt}^P}\right]$. See Gandhi et al. (2020) for a discussion on the empirical consequences of assuming full independence.

where the wages are summarized as $W_t^i = \prod_{z=1}^{Z} \prod_{v=1}^{V} \frac{(\beta_v^i)^{\beta_v^i} (\beta_z^i)^{\beta_z^i}}{(W_{vt}^i)^{\beta_v^i} (W_{zt}^i)^{\beta_z^i}}$. ε_{bt}^T gathers the transitory demand and supply shocks ($\varepsilon_{bt}^T = e^{\frac{\nu_{bt}}{\sigma} + \frac{\sigma-1}{\sigma}\epsilon_{bt}}$), while ε_{bt}^P gathers the productivity shocks with permanent implications $\left(\varepsilon_{bt}^P = \frac{\sigma-1}{\sigma}\left(\sum_{z=1}^{Z} \beta_z^i\right)\eta_{bt}\right)$.

For each firm b at time t, the timing assumptions of the model can be summarized as follows:

1. Observe the vector of state variables $\mathcal{I}_{bt}$, with $\omega_{bt} \in \mathcal{I}_{bt}$;
2. Choose freely adjustable inputs optimally for each market and start producing the optimal quantity q_{bt};
3. Observe deviations from expectations regarding supply (ϵ_{bt}) and realize final output y_{bt};
4. Observe deviations from expectations regarding demand (ν_{bt}) and sell at a market-clearing price determined by the demand function.
5. Decide optimally on next-period fixed production factors for each market.

The realized revenue expression for firms from country i selling in destination j at time t can be obtained as the product of output y_{bt}^{ij} with the market-clearing price p_{bt}^{ij} given by eq. 2:[8]

$$
\begin{aligned}
x_{bt}^{ij} = p_{bt}^{ij} y_{bt}^{ij} &= \left(y_{bt}^{ij}\right)^{\frac{\sigma-1}{\sigma}} \left(Y_t^j\right)^{\frac{1}{\sigma}} e^{\frac{\nu_{bt}}{\sigma}} P_t^j \\
&= \left[\left(\frac{\sigma-1}{\sigma} \frac{e^{\omega_{bt}}}{\tau_t^{ij} W_t^i} \right)^{\sigma} Y_t^j (P_t^j)^{\sigma} \mathbb{E}_{\varepsilon^P, \varepsilon^T} \left[e^{\varepsilon_{bt}^T + \varepsilon_{bt}^P} \right]^{\sigma} \frac{1}{e^{\sigma \varepsilon_{bt}^P}} \right]^{\frac{\sigma-1}{\sigma}} \left(Y_t^j\right)^{\frac{1}{\sigma}} P_t^j e^{\sigma \varepsilon_{bt}^T} \\
&= \left(\frac{\sigma}{\sigma-1} \tau_t^{ij} W_t^i \right)^{1-\sigma} Y_t^j \left(P_t^j\right)^{\sigma} \mathbb{E}_{\varepsilon^P, \varepsilon^T} \left[e^{\varepsilon_{bt}^T + \varepsilon_{bt}^P} \right]^{\sigma-1} e^{(\sigma-1)\omega_{bt} + \varepsilon_{bt}^T - (\sigma-1)\epsilon_{bt}^P}
\end{aligned}
\tag{16}
$$

B.2 Operating decisions and aggregation with certainty in future productivity

Going further, we reduce the dynamics of the model specifying $G(\omega_{bt}|\mathcal{I}_{bt-1})$ such that $\omega_{bt} = \omega_{bt-1} = \omega_b$. The model thus simplifies to a heterogeneous firms model which features certainty in future productivity after entry but uncertainty in the realized supply and demand. Eliminating the dynamics as such has the advantage of resulting in clear analytical expressions for the equilibrium variables. Moreover, it allows us to focus on the influence of transitory uncertainty when computing the trade elasticity and GFT compared to the predominant heterogeneous firms model without transitory uncertainty (see for instance Head et al. (2014); Melitz and Redding (2015); Nigai (2017); Bee and Schiavo (2018)). The assumption of certainty in future productivity does imply, however, that there is no role for permanent uncertainty in the model from here onwards.

[8] Note that if all inputs would be variable, the realized revenue expression would simplify to

$$x_{bt}^{ij} = \left(\frac{\sigma}{\sigma-1} \tau_t^{ij} W_t^i \right)^{1-\sigma} Y_t^j \left(P_t^j\right)^{\sigma} \mathbb{E}_{\varepsilon^T} \left[e^{\varepsilon_{bt}^T} \right]^{\sigma-1} e^{(\sigma-1)\omega_{bt} + \varepsilon_{bt}^T}$$

The productivity cutoffs for serving each market are determined by two equations. First, a firm decides whether to exit or enter/stay in a market based on its ability to generate a positive profit, resulting in a zero-profit condition:

$$0 = \mathbb{E}_{\varepsilon^T}\left[\pi^{ij}\left((\sigma-1)\omega^{ij*} + \varepsilon_b^T\right)\right]. \tag{17}$$

Second, a subset of active firms make positive profits net of the sunk entry cost. Free entry implies that in equilibrium, this expected measure of ex-ante profits (inclusive of the entry cost) must be equal to zero

$$\begin{aligned} fe^i &= \left[1 - G(\omega^{ii*})\right]\int_{\omega^{ii*}}^{\infty}\int_{-\infty}^{\infty}\sum_{j=1}^{J}\pi^{ij}\left((\sigma-1)\omega_b + \varepsilon_b^T\right)\frac{dG(\omega_b)}{1-G(\omega^{ii*})}dG(\epsilon_b^T) \\ &= \left[1 - G(\omega^{ii*})\right]\mathbb{E}_{\omega,\varepsilon^T}\left[\sum_{j=1}^{J}\pi^{ij}\left((\sigma-1)\omega_b + \varepsilon_b^T\right)\Big|\omega_b > \omega^{ii*}\right] \end{aligned} \tag{18}$$

With the productivity cutoffs determined, we can sum equation 8 across all active firms trading between i and j (M^{ij}) to obtain an expression for aggregate trade between country i and j:

$$\begin{aligned} x^{ij} &= \frac{M^{ij}}{1-G(\omega^{ij*})}\int_{\omega^{ij*}}^{\infty}\int_{-\infty}^{\infty}x_b^{ij}dG(\omega_b)dG(\varepsilon_b^T) \\ &= \frac{M^{ij}}{1-G(\omega^{ij*})}\left(\frac{\sigma}{\sigma-1}\tau^{ij}W^i\right)^{1-\sigma}Y^j\left(P^j\right)^{\sigma}\mathbb{E}_{\varepsilon^T}\left[e^{\varepsilon_b^T}\right]^{\sigma-1}\int_{\omega^{ij*}}^{\infty}\int_{-\infty}^{\infty}e^{(\sigma-1)\omega_b+\varepsilon_b^T}dG(\omega_b)dG(\varepsilon_b^T) \\ &= \frac{M^{ij}}{1-G(\omega^{ij*})}\left(\frac{\sigma}{\sigma-1}\tau^{ij}W^i\right)^{1-\sigma}Y^j\left(P^j\right)^{\sigma}\mathbb{E}_{\varepsilon^T}\left[e^{\varepsilon_b^T}\right]^{\sigma-1}\int_{\omega^{ij*}}^{\infty}e^{(\sigma-1)\omega_b}dG(\omega_b)\int_{-\infty}^{\infty}e^{\varepsilon_b^T}dG(\varepsilon_b^T) \\ &= \frac{M^{ij}}{1-G(\omega^{ij*})}\left(\frac{\sigma}{\sigma-1}\tau^{ij}W^i\right)^{1-\sigma}Y^j\left(P^j\right)^{\sigma}\mathbb{E}_{\varepsilon^T}\left[e^{\varepsilon_b^T}\right]^{\sigma}\int_{\omega^{ij*}}^{\infty}e^{(\sigma-1)\omega_b}dG(\omega_b). \end{aligned} \tag{19}$$

From the aggregate revenue expression (eq. 19), we observe that the heterogeneity from independent transitory shocks reduces to a constant. The partial sensitivity of aggregate trade to changes in variable trade costs, the aggregate trade elasticity, is then defined as (Chaney, 2008; Arkolakis et al., 2012; Melitz and Redding, 2014; Bas et al., 2017): [9]

[9] Aggregate trade elasticity is here defined as the direct response of aggregate trade to trade costs, keeping the indirect effect trough the price index via its impact on the domestic cutoff fixed (Melitz and Redding, 2015).

$$
\begin{aligned}
\gamma^{ij} \equiv \left.\frac{\partial ln x^{ij}}{\partial ln \tau^{ij}}\right|_{\omega^{ii*}} &= 1-\sigma+\frac{dln\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)}{dln\omega^{ij*}} \left.\frac{\partial ln\omega^{ij*}}{\partial ln\tau^{ij}}\right|_{\omega^{ii*}} \\
&= 1-\sigma+\frac{d\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)}{d\omega^{ij*}} \frac{e^{\omega^{ij*}}}{\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)} \left.\frac{\partial ln\omega^{ij*}}{\partial ln\tau^{ij}}\right|_{\omega^{ii*}} \\
&= 1-\sigma-\frac{e^{\sigma\omega^{ij*}} g(\omega^{ij*})}{\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)},
\end{aligned} \tag{20}
$$

which is independent of transitory shocks to demand and supply. Whereas the firms' expectations of transitory shocks to demand and supply affect aggregate trade costs levels, the independent nature of these shocks renders these expectations invariable to a change in variable trade costs. As such, the idiosyncratic shocks do not affect the changes in trade flows as the result of a change in variable trade costs.

The mass of firms is specified as the ratio of aggregate over average revenue:

$$
M^i = \left[1-G(\omega^{ii*})\right] M^{ie} = \frac{x^i}{\mathbb{E}_{\omega,\epsilon^T}\left[x^i\left((\sigma-1)\omega_b+\varepsilon_b^T\right)\right]}. \tag{21}
$$

We can rewrite the mass of firms using the free entry condition, goods and labor market clearing $\left(x^i = \tilde{W}^i L^i + M^{ie} fe^i + \sum_{j=1}^{J} M^{ij} f^{ij}\right)$, with $\tilde{W}_t^i = W_t^i \prod_{z=1}^{Z}\prod_{v=1}^{V}(\beta_v^i)^{\beta_v^i}(\beta_z^i)^{\beta_z^i}$, as a function of exogenous variables:

$$
\begin{aligned}
M^i &= \frac{\tilde{W}^i L^i + M^{ie} fe^i + \sum_{j=1}^{J} M^{ij} f^{ij}}{\sigma\left(\frac{fe^i}{1-G(\omega^{ii*})}+\sum_{j=1}^{J}\frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})} f^{ij}\right)} \\
&= \frac{\tilde{W}^i L^i + M^i \frac{fe^i}{1-G(\omega^{ii*})} + M^i \sum_{j=1}^{J}\frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})} f^{ij}}{\sigma\left(\frac{fe^i}{1-G(\omega^{ii*})}+\sum_{j=1}^{J}\frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})} f^{ij}\right)} \\
&= \frac{\sigma}{\sigma-1}\frac{\tilde{W}^i L^i}{\sigma\left(\frac{fe^i}{1-G(\omega^{ii*})}+\sum_{j=1}^{J}\frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})} f^{ij}\right)} \\
&= \frac{\tilde{W}^i L^i}{(\sigma-1)\left(\frac{fe^i}{1-G(\omega^{ii*})}+\sum_{j=1}^{J}\frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})} f^{ij}\right)}
\end{aligned} \tag{22}
$$

Assuming a two-country symmetric economy and setting the wage of the composite factor as the numeraire, welfare can be calculated as the inverse of the price index

$$\mathbb{W}^i = (P^i)^{-1}, \tag{23}$$

with

$$\begin{aligned}
(P^i)^{1-\sigma} &= \sum_{j=1}^{J} M^{ij} \mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_b}{\sigma} + \frac{\sigma-1}{\sigma}\epsilon_b} \middle| \mathcal{I}_b \right]^{\sigma-1} \mathbb{E}_{\nu,\epsilon}\left[e^{\frac{\nu_b}{\sigma} + \frac{\epsilon_b}{\sigma}} \right]^{1-\sigma} \int_{\omega^{ij*}}^{\infty} \left(\frac{\sigma}{\sigma-1} \frac{\tau^{ij}}{e^{\omega_b}} W^i \right)^{1-\sigma} dG(\omega_b) \\
&= M^i \sum_{j=1}^{J} \frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})} \mathbb{E}_{\varepsilon^T}\left[e^{(\sigma-1)\varepsilon_b^T} \right] \mathbb{E}_{\nu,\epsilon}\left[e^{\frac{1-\sigma}{\sigma}(\nu_b+\epsilon_b)} \right] \left(\frac{\sigma}{\sigma-1} \tau^{ij} W^i \right)^{1-\sigma} \int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)
\end{aligned} \tag{24}$$

The changes in welfare from a change in variable trade costs ($\boldsymbol{\tau} \to \boldsymbol{\tau}'$) are then simply a ratio of the aggregate price indices. Whereas the firms' expectations of transitory shocks to demand and supply affect aggregate price levels, the independent nature of these shocks renders these expectations invariable to a change in variable trade costs. As such, the idiosyncratic shocks do not affect the changes in the price indices as the result of a change in variable trade costs. Transitory shocks have no influence on the aggregate gains from trade:

$$\frac{(\mathbb{W}^i)'}{\mathbb{W}^i} = \frac{P^i}{(P^i)'}. \tag{25}$$

Appendix C Structural productivity estimation

This section describes in detail the structural productivity estimation technique relied upon in this paper for a value-added Cobb-Douglas production function. The estimation strategy consist of two stages.[10] The first stage relies on the Ackerberg et al. (2015) proxy-variable approach to separate transitory shocks (ε^T_{bt}) from our main estimation equation (eq. 13). In the second stage, we identify the parameters of interest building on moment conditions with respect to the stochastic shocks of productivity. As such, we avert endogeneity problems and obtain consistent parameter estimates.

C.1 First stage

The first stage consists of separating the transitory idiosyncratic shock (ε^T_{bt}) from the main estimating equation (eq. 13). We follow Ackerberg et al. (2015) in dividing the set of variable production factors $\boldsymbol{L}^i_{bt}$ into a factor decided upon at time t, the proxy variable $\left\{L^i_{1bt}\right\}$, and a set of the remaining variable production factors also decided upon at time t, but before the proxy variable is decided upon $\boldsymbol{L}^i_{bt} \setminus \left\{L^i_{1bt}\right\}$).[11] As a result, the proxy demand function can be written as a function of all state variables and variable production factors, including unobserved (for the researcher) productivity:

$$L^i_{b1t} = h\left(\mathcal{I}_{bt}, \boldsymbol{L}^i_{bt} \setminus \left\{L^i_{1bt}\right\}\right). \tag{26}$$

Based on this proxy input demand equation and assuming strict monotonicity between this input demand and productivity, we can write:

$$\omega_{bt} = h^{-1}\left(\mathcal{I}_{bt} \setminus \omega_{bt}, \boldsymbol{L}^i_{bt}\right). \tag{27}$$

This inverse of the proxy demand function forms the basis of a control function approach that allows us to estimate the main estimation equation (eq. 13) and identify the transitory shock:

$$\begin{aligned}
ln\hat{\phi}^{ii}_{bt} &= lnx^{ii}_{bt} - ln\theta^{ii}_{bt}L^i_{b1t} - \varepsilon^T_{bt} \\
&= \tilde{h}\left(\theta^{ii}_{bt}, F^i_{b1t}, \ldots, F^i_{bZt}, L^i_{b2t}, \ldots, L^i_{bVt}, Y^i_t, h^{-1}\left(\mathcal{I}_{bt} \setminus \omega_{bt}, \boldsymbol{L}^i_{bt}\right)\right), \\
&= \frac{\sigma-1}{\sigma} ln\theta^{ii}_{bt} + \frac{\sigma-1}{\sigma}\left(\sum_{z=1}^{Z}\beta^i_z lnF^i_{bzt} + \sum_{v=2}^{V}\beta^i_v lnL^i_{bvt}\right) + \frac{1}{\sigma}lnY^i_t + \frac{\sigma-1}{\sigma}\omega_{bt},
\end{aligned} \tag{28}$$

[10]Due to data availability, we do not control for the selection bias as defined by Olley and Pakes (1996). Our framework can easily be extended to include an extra estimation stage as specified in Olley and Pakes (1996). However, following on their results for unbalanced panels, this extra stage is expected to have little influence on the results.

[11]See Ackerberg et al. (2015) for a discussion on the possible Data Generating Processes that could generate such differences in timing decisions for factor inputs.

where $ln\phi_{bt}^{ii} = lnx_{bt}^{ii} - ln\theta_{bt}^{ii} L_{b1t}^{i}$ denotes domestic sales minus the domestic share of the proxy variable input factor with unit output elasticity.

C.2 Second stage

The second stage, then, aims to obtain consistent estimates for the parameters in equation 28. We parametrize the assumed Markov process for productivity as a first-order auto-regressive process

$$\omega_{bt} = \alpha_0 + \alpha_1 \omega_{bt-1} + \eta_{bt}, \tag{29}$$

where, by construction $E\left[\eta_{bt} | \mathcal{I}_{bt-1}\right] = 0$.

Consistent parameter estimates for equation 28 can then be obtained based on the following moment conditions:

$$E\left(\eta_{bt}\begin{bmatrix} ln\boldsymbol{F}_{t}^{i} \\ ln\boldsymbol{L}_{bt}^{i} \setminus \left\{L_{1bt}^{i}\right\} \\ lnY_{t-1}^{i} \end{bmatrix}\right) = 0. \tag{30}$$

Current state parameter variables are exogenous to productivity shocks if decided upon at time $t-1$, while those that are at a later time can be identified from lagged observations. The elasticity of substitution parameter is identified from lagged values of the aggregate demand shifter.[12] Consistent estimates of the productivity Markov process and revenue production function parameters at hand, we are also capable of backing out an estimate of persistent shocks to productivity $\left(\varepsilon_{bt}^{P} = \frac{\hat{\sigma}-1}{\hat{\sigma}}\left(\sum_{z=1}^{Z}\hat{\beta}_{z}^{i}\right)\eta_{bt}\right)$.

[12]Note that, as already mentioned in Klette and Griliches (1996), this methodology does not allow for a time trend or time fixed effects, as the inclusion of a time variable would leave little to no variation for the demand shifter to identify the elasticity of substitution.

Appendix References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. *Econometrica 83*(6), 2411–2451.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New Trade Models, Same Old Gains? *American Economic Review 102*(1), 94–130.

Asker, J., A. Collard-Wexler, and J. De Loecker (2017). Market Power, Production (Mis) Allocation, and OPEC. *Economic Research Initiatives at Duke (ERID) Working Paper* (256).

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*, 1 – 19.

Bee, M. and S. Schiavo (2018). Powerless: gains from trade when firm productivity is not Pareto distributed. *Review of World Economics 154*(1), 15–45.

Chaney, T. (2008, September). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Das, S., M. J. Roberts, and J. R. Tybout (2007). Market entry costs, producer heterogeneity, and export dynamics. *Econometrica 75*(3), 837–873.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica 79*(5), 1407–1451.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Head, K., T. Mayer, and M. Thoenig (2014). Welfare and Trade without Pareto. *American Economic Review 104*(5), 310–316.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

Impullitti, G., A. A. Irarrazabal, and L. D. Opromolla (2013). A theory of entry into and exit from export markets. *Journal of International Economics 90*(1), 75–90.

Kasahara, H. and B. Lapham (2013). Productivity and the decision to import and export: Theory and evidence. *Journal of International Economics 89*(2), 297–316.

Klette, T. J. and Z. Griliches (1996). The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. *Journal of Applied Econometrics 11*(4), 343–61.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Melitz, M. J. and S. J. Redding (2014). Chapter 1 - Heterogeneous Firms and Trade. In E. H. Gita Gopinath and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 1–54. Elsevier.

Melitz, M. J. and S. J. Redding (2015). New Trade Models, New Welfare Implications. *American Economic Review 105*(3), 1105–46.

Nigai, S. (2017). A tale of two tails: Productivity distribution and the gains from trade. *Journal of International Economics 104*, 44–62.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Ruhl, K. J. and J. L. Willis (2017). New exporter dynamics. *International Economic Review 58*(3), 703–726.