

NBER WORKING PAPER SERIES

NEW TRADE MODELS, NEW WELFARE IMPLICATIONS

Marc J. Melitz
Stephen J. Redding

Working Paper 18919
<http://www.nber.org/papers/w18919>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2013

We are grateful to Harvard and Princeton Universities for research support. This paper is a revised version of "Firm Heterogeneity and Aggregate Welfare," NBER Working Paper, 18919. We would like to thank the editor, four referees, conference and seminar participants at AEA, Harvard, Hong Kong, LSE, NBER, NYU, Paris, Princeton, UBC, Vancouver and Wharton, and Pol Antras, Costas Arkolakis, Ariel Burstein, Arnaud Costinot, Rob Feenstra, Gene Grossman, Gordon Hanson, Keith Head, Elhanan Helpman, Thierry Mayer, Gianmarco Ottaviano, Andres Rodriguez-Clare, Esteban Rossi-Hansberg, Jon Vogel and David Weinstein for helpful comments. We are also grateful to David Krisztian Nagy for research assistance. Responsibility for any results, opinions and errors is the authors' alone. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Marc J. Melitz and Stephen J. Redding. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

New Trade Models, New Welfare Implications
Marc J. Melitz and Stephen J. Redding
NBER Working Paper No. 18919
March 2013. "Tgxkgf "Hgdwtct { "4236
JEL No. F12,F15

ABSTRACT

We examine whether firm heterogeneity matters for the aggregate welfare implications of trade by comparing heterogeneous and homogeneous firm models. We show that the endogenous entry/ exit decisions of firms into/out of the domestic and export market provide a new channel for welfare gains that is absent from models without firm heterogeneity. We use a theoretical comparative static to isolate the additional component of welfare associated with this channel and to show that models of firm heterogeneity have new aggregate welfare implications. We express the welfare gains from trade in the heterogeneous and homogeneous firm models in terms of observable empirical moments. We show that in the general cases of these models the aggregate domestic trade share and trade elasticity are not sufficient statistics for welfare.

Marc J. Melitz
Department of Economics
Harvard University
215 Littauer Center
Cambridge, MA 02138
and NBER
mmelitz@harvard.edu

Stephen J. Redding
Department of Economics
and Woodrow Wilson School
Princeton University
Fisher Hall
Princeton, NJ 08544
and NBER
reddings@princeton.edu

An online appendix is available at:
<http://www.nber.org/data-appendix/w18919>

1 Introduction

Over the last decade, new theories of heterogeneous firms in differentiated product markets have been developed. These theories were designed to account for features of disaggregated trade data: only some firms export, exporters are more productive than non-exporters and trade liberalization induces intra-industry reallocations of resources between those different types of firms. These reallocations constituted a new potential channel for the gains from trade. However, the implications of these models for the aggregate gains from trade (combining together all welfare channels) were left unanswered.

In a recent paper, Arkolakis, Costinot and Rodriguez-Clare (2012, henceforth ACR) show that there exists a class of heterogeneous and homogeneous firm models in which a country's domestic trade share and the elasticity of trade with respect to variable trade costs are sufficient statistics for the aggregate welfare gains from trade. Therefore, if the different models within this class are calibrated to the same domestic trade share and the same trade elasticity, they imply the same welfare gains from trade. Based on this result, ACR summarizes the contribution of new theories of heterogeneous firms to the aggregate welfare implications of trade as "So far, not much."

In this paper, we examine whether firm heterogeneity matters for the aggregate welfare implications of trade by undertaking a theoretical comparative static in which we change the degree of firm heterogeneity holding all other structural parameters constant. We compare a heterogeneous firm model to a homogeneous firm model that is a special case with a degenerate productivity distribution. The heterogeneous firm model has an additional adjustment margin that is absent in the homogeneous firm model: the endogenous decisions of heterogeneous firms to enter and exit the domestic and export markets. Furthermore, adjustment along this margin is efficient, in the sense that the market equilibrium corresponds to the constrained efficient allocation chosen by a social planner. As a result, if the degenerate productivity distribution in the homogeneous firm model is chosen so that the two models have the same welfare for an initial value of trade costs, this additional adjustment margin implies that the heterogeneous firm model has higher welfare for all other values of trade costs. It follows that there are *larger welfare gains* from *reductions* in trade costs and *smaller welfare losses* from *increases* in trade costs in the heterogeneous firm model than in the homogeneous firm model.

We discuss how this comparative static isolates the additional welfare component associated with endogenous productivity responses via firm selection. Quantitatively, we find that this welfare component is substantial for a calibration of our heterogeneous firm model to U.S. data. Thus we conclude that new theories with heterogeneous firms that feature those endogenous productivity responses do have new aggregate welfare implications.

The class of models considered by ACR represents a subset of our homogeneous and heterogeneous firm models (associated with additional restrictions on the parameter space). Within this restricted subset, there is some tension between our theoretical comparative static (where firm heterogeneity matters for the gains from trade) and ACR's welfare evaluation based on an empirically measured trade share and elasticity (where firm heterogeneity is irrelevant for the gains from trade). However, we

show that this tension disappears as soon as we consider small (and empirically plausible) departures from that restricted subset of our models. Then, the trade share and trade elasticity are no longer sufficient statistics for welfare: even conditioning on these variables, changes in micro structure affect the welfare gains from trade.

We next develop several examples of trade liberalization scenarios where this additional impact of the micro structure on welfare can be substantial, even for small, empirically relevant, departures from those restrictions. These more general trade liberalization scenarios highlight that there is no single universal trade elasticity that can be applied to all contexts. Thus, evaluating a trade policy based on an estimated trade elasticity is subject to the Lucas Critique: This elasticity is not invariant with respect to changes in trade policy. Lastly, we show that assuming a constant trade elasticity when the true elasticity is variable can lead to substantial quantitative discrepancies between the predicted and true welfare effects of trade liberalization.

We extend the ACR approach of expressing the welfare gains from trade as a function of observable empirical moments (including the trade share and elasticity) to the more general cases of the homogeneous and heterogeneous firm models. We provide a framework for assessing whether the ACR formula provides a good approximation to the true welfare gains from trade liberalization. Using a trade elasticity estimated for a local change in trade costs in a similar context will reduce – but not eliminate – the discrepancy between the predicted and true welfare gains from trade. In addition to the two aggregate moments of the domestic trade share and trade elasticity, our more general welfare expression highlights differences in the hazard rate of the distribution of log firm size between the export and domestic markets and the response of firm entry to changes in trade costs, both of which can be examined empirically using firm-level data.

The remainder of the paper is organized as follows. In Section 2, we introduce the heterogeneous and homogeneous firm models. In Section 3, we develop our theoretical comparative static, and demonstrate the different aggregate welfare implications of changes in trade costs in the heterogeneous and homogeneous firm models. In Section 4, we extend the ACR approach of expressing welfare gains from trade as a function of observable empirical moments (including the trade share and elasticity) to the more general cases of the homogeneous and heterogeneous firm models. In Section 5, we provide several examples of trade liberalization scenarios where the additional impact of the model’s micro structure on welfare can be substantial. In Section 6, we examine the quantitative relevance of our theoretical results. Section 7 concludes.

2 Heterogeneous and Homogeneous Firm Models

We compare the canonical heterogeneous and homogeneous firm models of Melitz (2003) and Krugman (1980).¹ The homogeneous firm model is a special case of the heterogeneous firm model with a degenerate productivity distribution.

¹A web-based technical appendix contains the derivations of all expressions in the paper.

2.1 Closed Economy Heterogeneous Firm Model

The specification of preferences, production and entry is the same as in Melitz (2003).² There is a continuum of firms that are heterogeneous in terms of their productivity $\varphi \in (0, \varphi_{\max})$, which is drawn from a common cumulative distribution $G(\varphi)$ after incurring a sunk entry cost of f_e units of labor. We allow the upper bound of the support of the productivity distribution to be either finite ($\varphi_{\max} < \infty$) or infinite ($\varphi_{\max} = \infty$). Labor is the sole factor of production. Production involves a fixed production cost and a constant marginal cost that depends on firm productivity, so that $l(\varphi) = f_d + q(\varphi)/\varphi$ units of labor are required to supply $q(\varphi)$ units of output. Consumers have constant elasticity of substitution (CES) preferences with elasticity $\sigma > 1$ defined over the differentiated varieties supplied by firms. Profit maximization implies that variety prices are a constant mark-up over marginal cost that is determined by the elasticity σ . The revenue of a firm with productivity φ is then given by:

$$r_d(\varphi) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \varphi^{\sigma-1} R P^{\sigma-1} w^{1-\sigma}, \quad (1)$$

where R is aggregate revenue, P is the CES price index, and w is the wage. We use the subscript d to reference the domestic market.

We begin by considering the closed economy equilibrium, which can be summarized by the following three relationships, where we use the superscript A to denote the autarky equilibrium. First, fixed production costs imply a productivity cutoff (φ_d^A) below which firms exit. This cutoff is defined by a zero-profit condition equating operating profit to the fixed cost:

$$\frac{r_d(\varphi_d^A)}{\sigma} = \frac{R}{\sigma} \left(\frac{\sigma-1}{\sigma} \frac{P \varphi_d^A}{w}\right)^{\sigma-1} = w f_d. \quad (2)$$

Second, free entry requires that the probability of successful entry ($1 - G(\varphi_d^A)$) times average profits conditional on successful entry ($\bar{\pi}$) equals the sunk entry cost: $[1 - G(\varphi_d^A)] \bar{\pi} = w f_e$. Using the relationship linking relative firm revenue to relative firm productivity and the zero-profit cutoff condition above, this free entry condition can be expressed as:

$$f_d J(\varphi_d^A) = f_e, \quad (3)$$

where

$$J(\varphi_d^A) = \int_{\varphi_d^A}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_d^A}\right)^{\sigma-1} - 1 \right] dG(\varphi), \quad (4)$$

where $J(\varphi_d^A)$ is a monotonically decreasing function of the productivity cutoff. We can then write $J(\varphi_d^A)$ in terms of the ratio of average productivity to cutoff productivity:

$$J(\varphi_d^A) = [1 - G(\varphi_d^A)] \left[\left(\frac{\tilde{\varphi}_d^A}{\varphi_d^A}\right)^{\sigma-1} - 1 \right]. \quad (5)$$

²Following most of the subsequent international trade literature, including ACR, we consider a static version of Melitz (2003) in which there is zero probability of firm death.

Following Melitz (2003), we define $\tilde{\varphi}_d^A$ as a weighted average of firm productivity (corresponding to a harmonic mean weighted by output shares):

$$\tilde{\varphi}_d^A = \left[\int_{\varphi_d^A}^{\varphi_{\max}} \varphi^{\sigma-1} \frac{dG(\varphi)}{1 - G(\varphi_d^A)} \right]^{\frac{1}{\sigma-1}}. \quad (6)$$

Third, the mass of producing firms (M) equals the mass of entrants (M_e) times the probability of successful entry ($1 - G(\varphi_d^A)$). This mass of producing firms also equals aggregate revenue (R) divided by average firm revenue (\bar{r}). Using the relationship linking relative firm revenue to relative firm productivity and the free entry condition, the mass of producing firms can be written in terms of the economy's labor supply (L) relative to average fixed costs per firm:

$$M = [1 - G(\varphi_d^A)] M_e = \frac{R}{\bar{r}} = \frac{L}{\sigma F^A}. \quad (7)$$

In this derivation, we choose labor as the numeraire so that aggregate revenue R equals labor payments L , and we define F^A to represent the average fixed cost paid per surviving firm:

$$F^A = \frac{f_e}{1 - G(\varphi_d^A)} + f_d. \quad (8)$$

Using the CES price index and the mass of firms (7), closed economy welfare can be then written in terms of the mass of firms ($L/\sigma F^A$) and the weighted average productivity of these firms ($\tilde{\varphi}_d^A$):

$$\mathbb{W}_{\text{Het}}^A = \frac{w}{P} = \frac{\sigma - 1}{\sigma} \left\{ \frac{L}{\sigma F^A} (\tilde{\varphi}_d^A)^{\sigma-1} \right\}^{\frac{1}{\sigma-1}}. \quad (9)$$

2.2 Open Economy Heterogeneous Firm Model

We consider the case of trade between two symmetric countries. We use the subscript x to reference the export market and the superscript T to reference the open economy equilibrium. We assume that there is a fixed exporting cost of f_x units of labor in the source country and an iceberg variable trade cost, whereby $\tau > 1$ units of a variety must be shipped from one country in order for one unit to arrive in the other country. The open economy equilibrium is characterized by productivity cutoffs for serving the domestic market (φ_d^T) and export market (φ_x^T) that are defined by zero-profit conditions equating the operating profit in each market to the relevant fixed costs:

$$\frac{r_d(\varphi_d^T)}{\sigma} = \frac{R}{\sigma} \left(\frac{\sigma - 1}{\sigma} \frac{P \varphi_d^T}{w} \right)^{\sigma-1} = w f_d, \quad (10)$$

$$\frac{r_x(\varphi_x^T)}{\sigma} = \frac{R}{\sigma} \left(\frac{\sigma - 1}{\sigma} \frac{P \varphi_x^T}{\tau w} \right)^{\sigma-1} = w f_x. \quad (11)$$

The revenue functions $r_d(\varphi)$ and $r_x(\varphi)$ separate firm sales by destination market (domestic and export). Together these two zero-profit conditions imply that the export cutoff is a constant multiple of the domestic cutoff, where this multiple depends on the fixed and variable costs of trade:

$$\varphi_x^T = \tau \left(\frac{f_x}{f_d} \right)^{\frac{1}{\sigma-1}} \varphi_d^T. \quad (12)$$

For sufficiently high fixed and variable trade costs ($\tau(f_x/f_d)^{\frac{1}{\sigma-1}} > 1$), only the most productive firms export, consistent with an extensive empirical literature (see for example the review in Bernard, Jensen, Redding and Schott 2007).

The free entry condition again equates the expected value of entry to the sunk entry cost, $[1 - G(\varphi_d^T)] \bar{\pi} = wf_e$, and can be written as:

$$f_d J(\varphi_d^T) + f_x J(\varphi_x^T) = f_e, \quad (13)$$

where $J(\cdot)$ is defined in (4). Using the relationship between the productivity cutoffs (12), and noting that $J(\cdot)$ is a decreasing function, the free entry condition (13) determines a unique equilibrium value of the domestic cutoff (φ_d^T), which in turn determines the export cutoff (φ_x^T). Furthermore, the domestic cutoff in the open economy is strictly greater than the domestic cutoff in the closed economy ($\varphi_d^T > \varphi_d^A$) for positive values of fixed exporting costs.

As in the closed economy, the mass of producing firms (M) equals the mass of entrants (M_e) times the probability of successful entry ($1 - G(\varphi_d^T)$), and is determined by the economy's labor supply (L) relative to average fixed costs:

$$M = [1 - G(\varphi_d^T)] M_e = \frac{R}{\bar{r}} = \frac{L}{\sigma F^T}, \quad (14)$$

where F^T summarizes average fixed costs per surviving firm in the open economy:

$$F^T = \frac{f_e}{1 - G(\varphi_d^T)} + f_d + \chi f_x, \quad (15)$$

and $\chi = [1 - G(\varphi_x^T)] / [1 - G(\varphi_d^T)]$ is the proportion of exporting firms. In this derivation, we choose labor in one country as the numeraire and use country symmetry, which implies that aggregate revenue R still equals labor payments L in each country.

Using the CES price index and mass of firms (14), open economy welfare can be written in terms of the mass of varieties available for consumption ($L(1 + \chi)/\sigma F^T$) and the weighted average productivity of these varieties ($\tilde{\varphi}_t^T$):

$$\mathbb{W}_{\text{Het}}^T = \frac{w}{P} = \frac{\sigma - 1}{\sigma} \left\{ \frac{L(1 + \chi)}{\sigma F^T} (\tilde{\varphi}_t^T)^{\sigma-1} \right\}^{\frac{1}{\sigma-1}}. \quad (16)$$

This weighted average productivity ($\tilde{\varphi}_t^T$) in the open economy is constructed using the same weighting scheme (6) as we used for the closed economy. However the productivity of exporters is reduced by τ to account for the units "lost" in transit. Letting $\tilde{\varphi}_x^T$ denote the average productivity of exporters, defined as in (6), the overall productivity average for the open economy can be written:

$$(\tilde{\varphi}_t^T)^{\sigma-1} = \frac{1}{1 + \chi} \left[(\tilde{\varphi}_d^T)^{\sigma-1} + \chi (\tau^{-1} \tilde{\varphi}_x^T)^{\sigma-1} \right]. \quad (17)$$

Aggregate trade between the two countries is inversely related to the domestic trade share (the proportion of domestic sales in total sales):

$$\lambda_{\text{Het}} = \frac{\int_{\varphi_d^T}^{\varphi_{\text{max}}} r_d(\varphi) dG(\varphi)}{R} = \frac{1}{1 + \tau^{1-\sigma} \Lambda}, \quad (18)$$

where $\Lambda = \delta(\varphi_x^T)/\delta(\varphi_d^T) \leq 1$ is the market share of exporters in the domestic market and $\delta(\varphi_j) = \int_{\varphi_j}^{\varphi_j^{\max}} (\varphi)^{\sigma-1} dG(\varphi)$ is a function that depends only on $G(\cdot)$ and σ . The sensitivity of aggregate trade to changes in variable trade costs is captured by the full trade elasticity:

$$\theta_{\text{Het}} = -\frac{d \ln \left(\frac{1-\lambda_{\text{Het}}}{\lambda_{\text{Het}}} \right)}{d \ln \tau} = \begin{cases} (\sigma - 1) - \frac{d \ln \Lambda}{d \ln \tau} > 0 & \text{for } \tau (f_x/f_d)^{1/(\sigma-1)} > 1 \\ (\sigma - 1) > 0 & \text{for } \tau (f_x/f_d)^{1/(\sigma-1)} < 1 \end{cases}, \quad (19)$$

where $d \ln \Lambda / d \ln \tau < 0$. When trade costs are sufficiently low, all firms export and there is no extensive margin of trade. Given CES preferences, the elasticity of the intensive margin of trade is constant at $\sigma - 1$. When there is selection into the export market, the elasticity of trade θ_{Het} is the sum of the intensive margin elasticity $\sigma - 1$ and the extensive margin elasticity $-d \ln \Lambda / d \ln \tau$.

2.3 Closed Economy Homogeneous Firm Model

We construct a homogeneous firm model that is a special case of the heterogeneous firm model with a degenerate productivity distribution. Firms pay the same sunk entry cost of f_e units of labor and draw a productivity of either zero or $\bar{\varphi}_d$ with exogenous probabilities \bar{G}_d and $(1 - \bar{G}_d)$ respectively. Fixed production costs imply that only firms drawing a productivity of $\bar{\varphi}_d$ find it profitable to produce. Therefore producing firms are homogeneous and there is a degenerate productivity distribution conditional on production at $\bar{\varphi}_d$.

This homogeneous firm model is isomorphic to Krugman (1980), in which the representative firm's productivity is set equal to $\bar{\varphi}_d$ and the fixed production cost is scaled to incorporate the expected value of entry costs ($\bar{F}_d \equiv f_d + f_e / [1 - \bar{G}_d]$). These values for the representative firm's productivity and the fixed production cost are exogenous and held constant. To simplify the exposition, we adopt this Krugman (1980) interpretation. The representative firm's production technology is:

$$l = \frac{q}{\bar{\varphi}_d} + \bar{F}_d. \quad (20)$$

Consumers have the same CES preferences that we defined previously. Profit maximization implies that equilibrium prices are a constant markup over marginal cost. Profit maximization and free entry imply that equilibrium output and employment for the representative variety are proportional to the fixed production cost:

$$q = \bar{\varphi}_d \bar{F}_d (\sigma - 1), \quad l = \sigma \bar{F}_d.$$

Using equilibrium employment for the representative variety, the mass of firms can be determined from the labor market clearing condition:

$$M = \frac{L}{\sigma \bar{F}_d}. \quad (21)$$

Using the CES price index and the mass of firms (21), closed economy welfare can be written in terms of the mass of firms ($L/\sigma \bar{F}_d$) and productivity ($\bar{\varphi}_d$):

$$\mathbb{W}_{\text{Hom}}^A = \frac{w}{P} = \frac{\sigma - 1}{\sigma} \left\{ \frac{L}{\sigma \bar{F}_d} (\bar{\varphi}_d)^{\sigma-1} \right\}^{\frac{1}{\sigma-1}}, \quad (22)$$

where we again choose labor as the numeraire.

2.4 Open Economy Homogeneous Firm Model

We again consider trade between two symmetric countries and assume the same trading technology as in the heterogeneous firm model, so that there is a fixed exporting cost of f_x units of labor and an iceberg variable trade cost of $\tau > 1$ units of each variety.

In the homogeneous firm model, the probability of successful entry and productivity conditional on successful entry are exogenous and remain unchanged and equal to $[1 - \bar{G}_d]$ and $\bar{\varphi}_d$ respectively. For sufficiently high fixed and variable trade costs ($\tau^{\sigma-1} f_x / \bar{F}_d > 1$), the representative firm does not find it profitable to export. In contrast, for sufficiently low fixed and variable trade costs ($\tau^{\sigma-1} f_x / \bar{F}_d < 1$), the representative firm finds it profitable to export, and there is trade in both models.

Profit maximization again implies that equilibrium prices are a constant mark-up over marginal costs, with export prices a constant multiple of domestic prices due to the variable costs of trade. Profit maximization and free entry imply that equilibrium output and employment for the representative variety are proportional to fixed costs:

$$\begin{aligned} q &= \bar{\varphi}_d (\bar{F}_d + f_x) (\sigma - 1), \\ l &= \sigma (\bar{F}_d + f_x). \end{aligned}$$

Therefore both output and employment rise for the representative firm following the opening of trade to cover the additional fixed costs of exporting.

Using equilibrium employment for the representative variety, the mass of firms can be determined from the labor market clearing condition:

$$M = \frac{L}{\sigma (\bar{F}_d + f_x)}. \quad (23)$$

Using the CES price index and the mass of firms (23), open economy welfare can be written in terms of the mass of varieties available for consumption ($2L/\sigma(\bar{F}_d + f_x)$) and average productivity ($\bar{\varphi}_t$):

$$\mathbb{W}_{\text{Hom}}^T = \frac{w}{P} = \frac{\sigma - 1}{\sigma} \left\{ \frac{2L}{\sigma (\bar{F}_d + f_x)} (\bar{\varphi}_t)^{\sigma-1} \right\}^{\frac{1}{\sigma-1}}. \quad (24)$$

where average productivity ($\bar{\varphi}_t$) is constructed in the same way as in (17) for heterogeneous firms:³

$$(\bar{\varphi}_t)^{\sigma-1} = \frac{1}{2} \left[(\bar{\varphi}_d)^{\sigma-1} + (\tau^{-1} \bar{\varphi}_d)^{\sigma-1} \right]. \quad (25)$$

We again choose the wage in one of the symmetric countries as the numeraire.

In the case of homogeneous firms, the domestic trade share simplifies to:

$$\lambda_{\text{Hom}} = \frac{1}{1 + \tau^{1-\sigma}}. \quad (26)$$

There is no extensive margin of trade, so the trade elasticity is given by the constant elasticity for the intensive margin of trade (so long as there is some trade):

$$\theta_{\text{Hom}} = \begin{cases} \sigma - 1 & \text{for } \tau^{\sigma-1} f_x / \bar{F}_d < 1 \\ 0 & \text{otherwise} \end{cases}. \quad (27)$$

³With a representative firm, the average productivity of both domestic firms and exporters is $\bar{\varphi}_t$, and the proportion of exporting firms is one.

3 Theoretical Comparative Static

To examine the aggregate welfare properties of the two models, we first pick the parameters \bar{G}_d and $\bar{\varphi}_d$ for the degenerate productivity distribution with homogeneous firms such that welfare in an initial equilibrium is the same as in the equilibrium with heterogeneous firms. We next examine the effects of changes in trade costs from this initial equilibrium. This calibration enables us to undertake a theoretical comparative static in which we examine the effect of the productivity distribution on the model's welfare properties keeping all other structural parameters the same (same $f_d, f_e, f_x, \tau, L, \sigma$). We examine both the opening of the closed economy to trade and changes in trade costs in the open economy equilibrium.

3.1 Opening the Closed Economy to Trade

We begin by picking the parameters \bar{G}_d and $\bar{\varphi}_d$ of the degenerate productivity distribution with homogeneous firms such that the autarky equilibrium is isomorphic to that with heterogeneous firms, and examine the effect of opening the closed economy to trade.

Proposition 1 *Consider a homogeneous firm model that is a special case of the heterogeneous firm model with an exogenous probability of successful entry $[1 - \bar{G}_d] = [1 - G(\varphi_d^A)]$ and an exogenous degenerate distribution of productivity conditional on successful entry $\bar{\varphi}_d = \tilde{\varphi}_d^A$. Given the same value for all remaining parameters $\{f_d, f_e, L, \sigma\}$, all aggregate variables (welfare, wage, price index, mass of firms, and aggregate revenue) are the same in the closed economy equilibria of the two models.*

Proof. Comparing (9) and (22), equal welfare follows immediately from $\bar{\varphi}_d = \tilde{\varphi}_d^A$ and $[1 - \bar{G}_d] = [1 - G(\varphi_d^A)]$, which implies $\bar{F}_d = F^A$. This also implies equal price indices. Equal masses of firms follow immediately from equal price indices and $\bar{\varphi}_d = \tilde{\varphi}_d^A$. Equal aggregate revenue follows from $R = L$ in both models. ■

This first proposition reflects the aggregation properties of the heterogeneous firm model. All aggregate variables in this model take the same value as if there were a representative firm with productivity $\bar{\varphi}_d$ and fixed costs \bar{F}_d . The key difference between the heterogeneous firm model and such a representative firm model is that aggregate productivity in the heterogeneous firm model is endogenous and responds to changes in trade costs. From the expressions for open economy welfare in the two models ((16) and (24)), open economy welfare is higher in the heterogeneous firm model than in the homogeneous firm model if the following inequality is satisfied:

$$\frac{(\tilde{\varphi}_d^T)^{\sigma-1} + \chi\tau^{1-\sigma}(\tilde{\varphi}_x^T)^{\sigma-1}}{\frac{f_e}{1-G(\varphi_d^T)} + f_d + \chi f_x} > \frac{(1 + \tau^{1-\sigma})(\bar{\varphi}_d)^{\sigma-1}}{\bar{F}_d + f_x}.$$

From the free entry condition in the open economy equilibrium of the heterogeneous firm model, and noting that $\bar{F}_d = f_d + f_e/[1 - G(\varphi_d^A)]$ and $\bar{\varphi}_d = \tilde{\varphi}_d^A$, this inequality is necessarily satisfied for any open economy equilibrium of the heterogeneous firm model in which the productivity cutoffs are different in

the open and closed economies ($\varphi_d^T \neq \varphi_d^A$ and $\varphi_d^T \leq \varphi_x^T \leq \varphi_{\max}$). Since the two models have the same closed economy welfare, and open economy welfare is higher in the heterogeneous firm model than in the homogeneous firm model, it follows that the welfare gains from opening the closed economy to trade are larger in the heterogeneous firm model.

Proposition 2 *Choosing the degenerate productivity distribution in the homogeneous firm model so that the two models have the same closed economy welfare and the same structural parameters ($f_d, f_e, f_x, \tau, L, \sigma$), the proportional welfare gains from opening the closed economy to trade are strictly larger in the heterogeneous firm model than in the homogeneous firm model ($\mathbb{W}_{Het}^T/\mathbb{W}_{Het}^A > \mathbb{W}_{Hom}^T/\mathbb{W}_{Hom}^A$), except in the special case with no fixed exporting cost. In this special case, the proportional welfare gains from opening the closed economy to trade are the same in the two models.*

Proof. See the Appendix. ■

Since the proportional welfare gains from opening the closed economy to trade are strictly lower in the homogeneous firm model than in the heterogeneous firm model for positive fixed exporting costs, and since open economy welfare in the homogeneous firm model is monotonically decreasing in trade costs, we also obtain the following result.

Proposition 3 *Achieving the same proportional welfare gains from trade in the two models requires strictly lower trade costs (either lower f_x and/or lower τ) in the homogeneous firm model than in the heterogeneous firm model, except in the special case with no fixed exporting cost.*

Proof. The proposition follows immediately from $\mathbb{W}_{Het}^T/\mathbb{W}_{Het}^A > \mathbb{W}_{Hom}^T/\mathbb{W}_{Hom}^A$ in Proposition 2 and from $\frac{d\mathbb{W}_{Hom}^T}{df_x} < 0$ and $\frac{d\mathbb{W}_{Hom}^T}{d\tau} < 0$ in (24). ■

Although we chose the productivity of the representative firm ($\bar{\varphi}_d = \tilde{\varphi}_d^A$) to ensure the same closed economy welfare in both models, the ratio of open to closed economy welfare in the homogeneous firm model $\mathbb{W}_{Hom}^T/\mathbb{W}_{Hom}^A$ is independent of the representative firm's productivity (from (22) and (24)). It follows that both of the above propositions hold for any value of the representative firm's productivity. Both these propositions also hold for general continuous productivity distributions.

The intuition for these results involves revealed preference arguments of the kind commonly used in international trade.⁴ Our starting point is to note that, with CES preferences and monopolistic competition, the open economy equilibrium in the heterogeneous firm model is efficient. As shown in the web appendix, a welfare-maximizing social planner faced with the same production and entry technology would choose the same allocation as the market equilibrium. When the economy is opened to trade, the planner *could* choose not to adjust the set of firms selected for production and exports. Average productivity would then remain constant, and this outcome would replicate the opening to trade in the homogeneous firm case. The latter is thus a feasible allocation for the planner with the

⁴For the sake of parsimony, we focus on symmetric countries; however, this revealed preference argument applies more generally for asymmetric countries.

heterogeneous firm production and entry technology. However, efficiency implies that this planner chooses to replicate the market equilibrium of the heterogeneous firm model, which involves adjustments in the set of firms selected for production and exports – and an associated increase in average productivity. This induced allocation therefore yields higher welfare than any other feasible allocation – including the homogeneous firm outcome. Additionally, the planner’s objective function is strictly concave. Thus, welfare in the heterogeneous firm open economy must be strictly higher than in the homogeneous firm case whenever the cutoffs adjust to the opening to trade.⁵

The difference in aggregate welfare implications between the two models arises because of the additional efficient adjustment margin of firm entry and exit decisions in the heterogeneous firm model (for both the domestic and export markets). In the special case with no fixed exporting cost, the domestic productivity cutoff is unaffected by the opening of the closed economy to trade. As a result, the additional adjustment margin of firm entry and exit decisions is inoperative in the heterogeneous firm model, and the welfare effects of trade liberalization are the same in the two models. We consider this special case uninteresting, since firm productivity dispersion then plays no role in the heterogeneous firm model (the exit threshold and average productivity are the same in the closed and open economies). Furthermore, this special case stands at odds with an extensive body of empirical evidence that only some firms export, exporters are larger and more productive than non-exporters, and there are substantial fixed exporting costs.⁶

3.2 Changes in Trade Costs in the Open Economy Equilibrium

The role of the additional adjustment margin of firm entry and exit decisions for generating different aggregate welfare implications is not limited to the opening of the closed economy to trade and also holds for reductions in trade costs in the open economy equilibrium. To show this, we recast our heterogeneous and homogeneous firm models so that they have the same welfare in an initial open economy equilibrium. In order to ensure that the two models have the same initial welfare and only differ in their productivity distribution (keeping the same structural parameters $f_d, f_e, f_x, \tau, L, \sigma$), we extend the homogeneous firm model to allow for two types of firms: exporters and non-exporters. In this extension, firms again pay a sunk entry cost of f_e units of labor before observing their productivity. With probability $[1 - \bar{G}_x]$ a firm draws a productivity of $\bar{\varphi}_x$ and can export; with probability \bar{G}_{dx} the firm draws a productivity of $\bar{\varphi}_{dx}$ and cannot export; with probability $[\bar{G}_x - \bar{G}_{dx}]$ the firm draws a productivity of zero and does not find it profitable to produce.

We pick the parameters of this “extended” homogeneous firm model $(\bar{\varphi}_x, \bar{\varphi}_{dx}, \bar{G}_x, \bar{G}_{dx})$ such that the open economy equilibrium features the same aggregate variables as the initial open economy

⁵In contrast, if the elasticity of substitution between varieties is variable, the market equilibrium is not in general efficient (see Dixit and Stiglitz 1977 for the case of homogeneous firm models and Dhingra and Morrow 2012 for the case of heterogeneous firm models). Nevertheless, endogenous firm selection still has the potential to generate higher welfare in the heterogeneous firm model than in the homogeneous firm model, if adjustment along this additional margin is similar in both the market equilibrium and social optimum, as considered in the web appendix.

⁶For reviews of the extensive empirical literatures on firm export market participation, see Bernard, Jensen, Redding and Schott (2007) and Melitz and Redding (2013). For evidence of substantial fixed exporting costs, see Roberts and Tybout (1997) and Das, Roberts and Tybout (2007).

equilibrium with heterogeneous firms (same welfare, price index, mass of firms, aggregate revenue, and domestic trade share). This requires equating those parameters with their corresponding averages under firm heterogeneity:

$$\begin{aligned}\bar{G}_x &= G(\varphi_x^T), & \bar{G}_{dx} &= G(\varphi_x^T) - G(\varphi_d^T), \\ \bar{\varphi}_x &= \tilde{\varphi}_x^T, & \bar{\varphi}_{dx} &= \left\{ \frac{1}{G(\varphi_x^T) - G(\varphi_d^T)} \int_{\varphi_d^T}^{\varphi_x^T} \varphi^{\sigma-1} dG(\varphi) \right\}^{\frac{1}{\sigma-1}},\end{aligned}$$

where the last term represents the average productivity of non-exporters in the heterogeneous firm model. We keep the values of all the other structural parameters ($f_d, f_e, f_x, \tau, L, \sigma$) constant across the two models. This ensures that the aggregate statistics line-up across the two models in the initial open economy equilibrium.

Nevertheless these two models respond differently to changes in trade costs from this common initial equilibrium along a key dimension. In the heterogeneous firm model, the endogenous selection responses to trade costs lead to changes in the average productivity of exporting and non-exporting firms and in the proportion of exporting firms. In contrast, in the extended homogeneous firm model, the average productivity levels of exporters and non-exporters and the proportion of exporting firms remain constant.⁷ The presence of this additional adjustment margin in the heterogeneous firm model implies that welfare following the change in trade costs must be strictly higher than in the homogeneous firm model. This argument holds irrespective of whether trade costs decrease or increase. Therefore, welfare *gains* are *larger* in the heterogeneous firm model whenever trade costs fall, and welfare *losses* are *smaller* in the heterogeneous firm model whenever trade costs *increase*.

Proposition 4 *Starting from an initial open economy equilibrium with the same welfare and the same structural parameters in the two models ($f_d, f_e, f_x, \tau, L, \sigma$), a common decrease (increase) in variable or fixed trade costs generates larger welfare gains (smaller welfare losses) in the heterogeneous firm model than in the extended homogeneous firm model.*

Proof. See the Appendix. ■

Note that the extended homogeneous firm model is equivalent to a version of the heterogeneous firm model in which the domestic and export productivity cutoffs are held constant at their values in an initial open economy equilibrium. Put another way, consider a planner who is constrained to keep the same set of firms operating in both the domestic and export markets – i.e. the endogenous selection margin is inoperative. Under this constraint, the welfare-maximizing allocation coincides with the market equilibrium of the extended homogeneous firm model. In contrast, in the absence of this constraint, the welfare-maximizing allocation coincides with the market equilibrium of the heterogeneous firm model. Therefore, the welfare differential between the two models provides a

⁷Unless trade costs become sufficiently high that firms with productivity $\bar{\varphi}_x$ no longer find it profitable to export or firms with productivity $\bar{\varphi}_{dx}$ no longer find it profitable to produce. In both cases, the average productivity of the two types of firms remains constant at $\bar{\varphi}_x$ and $\bar{\varphi}_{dx}$

direct measure of the impact of selection on aggregate welfare. In other words, it isolates the *additional* contribution to aggregate welfare of the new endogenous selection/productivity channel highlighted by the heterogeneous firm model of trade – this represents the new welfare implications that we refer to in the title of this paper. Later in Section 6, we show that this additional welfare channel is quantitatively substantial for a model calibrated to U.S. aggregate and firm statistics.

Atkeson and Burstein (2010) considers this welfare differential from endogenous firm selection in a model with product and process innovation. They find that this welfare differential is of second-order. Proposition 4 is consistent with this result. As discussed above and shown formally in the web appendix, the initial equilibrium of the heterogeneous firm model is efficient. Therefore the envelope theorem implies that the changes in the productivity cutoffs in the heterogeneous firm model have only second-order effects on welfare. But, as we show later, these second-order welfare effects can be quite substantial for larger changes in trade costs.

3.3 Untruncated Pareto Distribution

Since the homogeneous firm model is a special case of the heterogeneous firm model, our above comparisons of the two models are equivalent to a discrete comparative static of moving from a non-degenerate to a degenerate productivity distribution within the heterogeneous firm model. In the special case of an untruncated Pareto productivity distribution, the degree of firm heterogeneity is summarized by a single parameter: the shape parameter k . Lower values of k correspond to greater firm heterogeneity and the homogeneous firm model corresponds to the limiting case in which $k \rightarrow \infty$. Therefore we can complement the above discrete comparative static with a continuous comparative static in the degree of firm heterogeneity.

Proposition 5 *Assuming that productivity in the heterogeneous firm model has an untruncated Pareto distribution ($g(\varphi) = k\varphi_{\min}^k \varphi^{-(k+1)}$, where $\varphi \geq \varphi_{\min} > 0$ and $k > \sigma - 1$) and fixed exporting costs are positive, greater dispersion of firm productivity (smaller k) implies: (a) larger welfare gains from opening the closed economy to trade (larger $\mathbb{W}_{Het}^T/\mathbb{W}_{Het}^A$), (b) larger (smaller) welfare gains (losses) from a decrease (increase) in variable trade costs in the open economy equilibrium.*

Proof. See the appendix. ■

A larger dispersion of firm productivity (smaller k) implies greater scope for adjustment along the margin of endogenous firm entry and exit decisions, which enables higher welfare to be achieved following a change in trade costs. As with our earlier results, the continuous comparative static in Proposition 5 changes the degree of firm heterogeneity (k) holding all other structural parameters constant. Therefore, in part (a), we compare closed and open economy equilibria in two heterogeneous firm models that have the same values of all parameters except for the degree of firm heterogeneity. Similarly, in part (b), we compare two open economy equilibria with different values of variable trade costs in two heterogeneous firm models that have the same values of all parameters except for the

degree of firm heterogeneity. In both cases, we determine the change in welfare in response to a change in trade costs as a function of the parameters of the model.

In this special case of an untruncated Pareto distribution, the heterogeneous firm model falls within the class considered by ACR. Therefore Proposition 5 confirms that the degree of firm heterogeneity affects the aggregate welfare implications of trade, in the sense of our theoretical comparative static, even within the class of models considered by ACR.

4 Welfare and Trade Policy Evaluation

Our theoretical comparative static in the previous section examines the impact of changes in the distribution of productivity holding other exogenous variables fixed across models. This exercise does not restrict the equilibrium values of the endogenous variables (in particular the domestic trade share λ and the trade elasticity θ) to be the same in the two models. Instead the equilibrium values for these endogenous variables differ systematically across the two models. In the appendix, we show that the heterogeneous firm model generates a higher trade elasticity than either the homogeneous firm model or its extension given the same value of the exogenous variables. On the one hand, moving from the closed economy to the open economy, there is *less* trade (higher λ) in the heterogeneous firm model than in the homogeneous firm model. On the other hand, starting from an open economy equilibrium, trade liberalization generates *more* trade (lower λ) in the heterogeneous firm model than in the extended homogeneous case.⁸

ACR show that there exists a restricted subset of our heterogeneous and homogeneous firm model (in terms of parameter space restrictions) in which these two variables – the trade elasticity and the trade share – become sufficient statistics for welfare: There is a unique mapping from changes in these variables to changes in welfare. Even then, the micro structure of the underlying model still matters for the welfare gains from trade, but only through its effect on the trade share and trade elasticity. Therefore, if aggregate data can be used to measure the trade elasticity independently of a model (the trade share, by definition, is directly observed from aggregate data) then these aggregate data can be used to accurately measure the welfare gains from trade; and this welfare computation will be independent of the micro structure of the underlying model. This discussion summarizes the crux of the differences between our theoretical comparative static and the ACR approach.

However, this tension between the two approaches only arises when the additional ACR restrictions on the parameter space (detailed below) are imposed. In this section, we extend the ACR approach of expressing the welfare gains from trade as a function of observable empirical moments to the more general cases of the homogeneous and heterogeneous firm models developed in Section 2. These empirical moments include the trade share and trade elasticity, but also additional ones that capture

⁸For sufficiently high trade costs, the domestic trade share is *higher* in the homogeneous firm model than in the heterogeneous firm model, because the representative firm does not find it profitable to export. As trade costs fall below the threshold at which the representative firm exports, the domestic trade share in the homogeneous firm model falls from one to a value *below* that in the heterogeneous firm case, and the trade elasticity jumps from zero to $\sigma - 1$ (less than the trade elasticity in the heterogeneous firm case).

micro structure (and differ between the two models). Thus, the sufficiency property of the moments generated from the aggregate data (the trade share and elasticity) is broken: even conditioning on those aggregate moments, micro structure influences the welfare gains from trade. In Section 5, we provide several examples of trade liberalization scenarios in which the additional impact of the micro structure on welfare can be substantial, even for small empirically relevant departures from the ACR restrictions. In Section 6, we quantitatively assess these differences in welfare predictions.

4.1 ACR Welfare Derivation

ACR show how the domestic trade share (λ) and trade elasticity (θ) are sufficient statistics for the welfare gains from trade in a large class of trade models (including special cases of our homogeneous and heterogeneous firm models), so long as three macro-level restrictions are satisfied: (R1) balanced trade; (R2) aggregate profits are a constant share of aggregate revenues; and (R3) a CES import demand system with a constant elasticity of trade with respect to variable trade costs. Under these restrictions, the welfare gains from trade regime T_0 to T_1 can be written:

$$\frac{\mathbb{W}^{T_1}}{\mathbb{W}^{T_0}} = \left(\frac{\lambda^{T_0}}{\lambda^{T_1}} \right)^{\frac{1}{\theta}}. \quad (28)$$

Thus, (28) will characterize the welfare gains from trade for both our homogeneous and heterogeneous firm models so long as (R1)-(R3) are satisfied. Trade is balanced in both of these models, so (R1) is always satisfied. However, the general versions of both models imply departures from a constant aggregate share of profits embodied in (R2). Given CES preferences, the constant aggregate trade elasticity restriction (R3) will be satisfied in all versions of our homogeneous firm model, whereas it will be violated along with (R2) in our general heterogeneous firm model.

4.2 Gains from Trade in the Homogeneous Firm Model

We consider a lowering of trade costs from τ_0 and f_{x0} (trade regime T_0) to τ_1 and f_{x1} (trade regime T_1). To simplify notation, we assume that τ_0 and f_{x0} may be high enough such that no trade is generated in T_0 . Let χ^{T_0} denote an indicator variable for positive trade. Then, using the expressions for welfare in the closed economy (22) and open economy (24) and the domestic trade share (26), we can write the welfare gains from trade in the heterogeneous firm model as:

$$\frac{\mathbb{W}^{T_1}}{\mathbb{W}^{T_0}} = \left[\frac{\lambda^{T_0} (\bar{F}_d + \chi^{T_0} f_{x0})}{\lambda^{T_1} (\bar{F}_d + f_{x1})} \right]^{\frac{1}{\sigma-1}}, \quad (29)$$

where $\theta = \sigma - 1$ is the elasticity of trade with respect to variable trade costs.

In this more general setting, the welfare gains depend on the same two aggregate moments (the domestic trade share and trade elasticity) as in ACR (28), but also on the change in firm size (captured in (29) by the total fixed costs paid by the representative firm). This change in firm size is an observable empirical moment, but one that characterizes a change in micro structure. Even after controlling for the two aggregate moments, these changes in micro structure will affect the welfare gains from trade.

Such changes in micro structure will occur whenever the fixed exporting cost changes in an open economy with trade ($\chi^{T_0} = 1$) or even in the presence of any positive fixed exporting costs in an economy that opens up to trade (from $\chi^{T_0} = 0$). These changes represent a violation of (R2) as the share of profits in revenue changes with firm size in the homogeneous firm model.

4.3 Gains from Trade in the Heterogeneous Firm Model

We now seek to express the welfare gains from trade liberalization in terms of observable empirical moments for the general case of our heterogeneous firm model. Since trade continuously drops to zero when trade costs increase, we can start from an open economy trade regime T without loss of generality. To simplify notation, we drop the T superscript. For now, we also assume that there is export market selection in this trade regime so that $\varphi_x > \varphi_d$.

From (19), the full trade elasticity with export market selection is $\theta_{\text{Het}} = (\sigma - 1) - d \ln \Lambda / d \ln \tau$, where $\Lambda = \delta(\varphi_x) / \delta(\varphi_d)$ represents the domestic market share of exporters (and hence changes in Λ capture changes in the extensive margin of trade). This full trade elasticity θ_{Het} incorporates the direct effect of τ on the extensive margin of trade via its impact on the export cutoff $\varphi_x = \tau(f_x/f_d)^{1/(\sigma-1)}\varphi_d$ (see (11)), as well as indirect effects through the price index via its impact on the domestic cutoff φ_d .

As argued by ACR, only the *partial* trade elasticity capturing the direct effect of τ is observed empirically, since it is estimated from a gravity equation with exporter and importer fixed effects. In the context of our symmetric country model, this partial elasticity can be derived from (18), which relates the domestic trade share to variable trade costs and the two productivity cutoffs ($\lambda = \lambda(\tau, \varphi_d, \varphi_x)$), and from (12), which relates the two productivity cutoffs to one another ($\varphi_x = \varphi_x(\tau, \varphi_d)$).⁹ Taking the partial derivative of the domestic trade share with respect to τ holding φ_d constant, we have:

$$\vartheta = - \left. \frac{\partial \ln \left(\frac{1-\lambda}{\lambda} \right)}{\partial \ln \tau} \right|_{\varphi_d} = (\sigma - 1) - \left. \frac{\partial \ln \Lambda}{\partial \ln \varphi_x} \frac{\partial \ln \varphi_x}{\partial \ln \tau} \right|_{\varphi_d},$$

where the relationship between the productivity cutoffs (12) implies $\partial \ln \varphi_x / \partial \ln \tau |_{\varphi_d} = 1$. Therefore the partial elasticity can be further written as:

$$\begin{aligned} \vartheta &= (\sigma - 1) - \left. \frac{\partial \ln \Lambda}{\partial \ln \varphi_x} \right|_{\varphi_d}, \\ &= (\sigma - 1) + \gamma(\varphi_x), \end{aligned} \tag{30}$$

where $\gamma(\varphi_j) = -d \ln \delta(\varphi_j) / d \ln \varphi_j$ is the elasticity of $\delta(\varphi_j)$ for market $j \in \{d, x\}$.

Note that $\delta(\varphi_j)$ is proportional to the cumulative market share (in any given market) of firms above any cutoff φ_j . Therefore $\gamma(\varphi_j)$ represents the hazard function for the distribution of log firm size within a market. If the distribution of productivity φ is an untruncated Pareto(k), then the distribution of firm size (in any given market) also will be an untruncated Pareto($k - \sigma + 1$) and the hazard function $\gamma(\cdot)$ will be constant at $k - (\sigma - 1)$. In this case, the partial and full trade elasticity

⁹In the web appendix, we show how a multi-country version of our model yields an expression for log bilateral trade that is linear in exporter and importer fixed effects and $\vartheta \ln \tau$.

are equal to one another and constant at k . Even a slight departure from an untruncated Pareto to a truncated Pareto implies that the partial and full trade elasticity are distinct from one another and variable. In this case, the hazard function $\gamma(\varphi_j)$ becomes:

$$\gamma(\varphi_j) = (k - (\sigma - 1)) \frac{\left(\frac{\varphi_{\min}}{\varphi_j}\right)^{k-(\sigma-1)}}{\left(\frac{\varphi_{\min}}{\varphi_j}\right)^{k-(\sigma-1)} - \left(\frac{\varphi_{\min}}{\varphi_{\max}}\right)^{k-(\sigma-1)}}, \quad (31)$$

where φ_{\max} is the upper bound to the support of the productivity distribution. As $\varphi_{\max} \rightarrow \infty$, the hazard function $\gamma(\varphi_j)$ converges to its constant value for an untruncated Pareto distribution: $\lim_{\varphi_{\max} \rightarrow \infty} \gamma(\varphi_j) = k - (\sigma - 1)$. More generally, for $\varphi_{\max} < \infty$, $\gamma(\varphi_j)$ takes a strictly higher value than for an untruncated Pareto productivity distribution and differs between the domestic and export market. The hazard function for each market is increasing in the productivity cutoff, attaining its minimum value as $\varphi_j \rightarrow \varphi_{\min}$, and converging towards infinity as $\varphi_j \rightarrow \varphi_{\max}$. Since higher variable trade costs reduce the domestic productivity cutoff and increase the export productivity cutoff, they imply a *lower* $\gamma(\varphi_d)$ and a *higher* $\gamma(\varphi_x)$.

Using welfare (16) and the trade share (18), welfare in the heterogeneous firm model can be written:

$$\mathbb{W}_{\text{Het}} = \frac{\sigma - 1}{\sigma} M_e^{\frac{1}{\sigma-1}} \left(\frac{\delta(\varphi_d)}{\lambda} \right)^{\frac{1}{\sigma-1}}, \quad (32)$$

Since welfare (16) also implies that changes in welfare are proportional to changes in the domestic productivity cutoff ($d \ln \mathbb{W} = d \ln \varphi_d$ and $d \ln \delta(\varphi_d) = -\gamma(\varphi_d) d \ln \varphi_d$), we can then write the welfare change using the (observable) partial trade elasticity from (30):

$$d \ln \mathbb{W} = \frac{1}{\vartheta + [\gamma(\varphi_d) - \gamma(\varphi_x)]} (d \ln M_e - d \ln \lambda). \quad (33)$$

As highlighted by ACR, restricting the distribution of productivity draws $G(\varphi)$ to be untruncated Pareto and assuming that there is export market selection ($\varphi_x^T > \varphi_d^T$), ensures that the macro restrictions (R1)-(R3) are satisfied. In this case, the hazard function is constant so that the difference $\gamma(\varphi_d) - \gamma(\varphi_x)$ is zero, and entry does not respond to changes in trade costs ($d \ln M_e = 0$). In this case, we recover the welfare gain derivation (28) from ACR. Since the trade elasticity is constant in this case, the welfare differential can be integrated to capture proportional welfare changes between any two trade regimes, so long as there is export market selection in both.

However, the welfare differential (33) highlights how, in the general case, the welfare gains from trade liberalization will change with the micro structure. Even after controlling for the trade share and trade elasticity, this micro structure matters for welfare through the hazard differential $\gamma(\varphi_d) - \gamma(\varphi_x)$. In Section 6, we show quantitatively how small changes in the shape of the distribution of firm productivity $G(\varphi)$ away from an untruncated Pareto distribution can lead to large changes in the hazard differential $\gamma(\varphi_d) - \gamma(\varphi_x)$. This issue is distinct from the challenge of measuring the “appropriate” trade elasticity ϑ in a world where this elasticity is variable (both across countries and within each country for different values of trade costs). In addition, the predicted welfare effects of

trade liberalization based on the ACR formula will diverge from the true welfare effects because of the variable nature of the partial trade elasticity ϑ .

The welfare differential (33) also shows that, in cases where trade liberalization leads to responses in firm entry ($d \ln M_e \neq 0$), then this change in micro structure will also affect the welfare gains from trade, even conditional on the domestic trade share and trade elasticity. In our setting with a general productivity distribution, this response in firm entry implies changes in product variety available for consumption. Another very special feature of an untruncated Pareto distribution is that the increase in product variety from imports (following trade liberalization) is exactly offset by a decrease in domestic product variety (associated with tougher selection). Feenstra (2013) shows that this special feature of the untruncated Pareto productivity distribution also extends to a general class of non-CES preferences – but that it is similarly broken by small departures away from an untruncated Pareto distribution (to a truncated Pareto distribution).¹⁰

Lastly, we briefly characterize the gains from trade (in terms of observable moments) when the trade costs are low enough that all surviving firms export. In other words, there is no export market selection and $\varphi_d^T = \varphi_x^T$. As we previously discussed, the equilibrium in this case will have identical aggregate properties to an equilibrium with homogeneous firms, in which all firms have a common productivity level $\tilde{\varphi}_d^T$ and face a fixed cost $F^T = (f_e / [1 - G(\varphi_d^T)]) + f_d + f_x$. Thus, the welfare gains associated with a transition from trade regime T_0 to T_1 can be measured using:

$$\frac{\mathbb{W}^{T_1}}{\mathbb{W}^{T_0}} = \left(\frac{\lambda^{T_0} F^{T_0}}{\lambda^{T_1} F^{T_1}} \right)^{\frac{1}{\sigma-1}}, \quad (34)$$

where in this case the full and partial trade elasticities are equal to one another: $\theta = \vartheta = \sigma - 1$.

As in the homogeneous firm case, we see that the welfare gains from trade depend on changes in average firm size (captured by F^T) as well as the domestic trade share and trade elasticity (which is now constant at $\sigma - 1$). Average firm size is now endogenous and varies with the domestic productivity cutoff (φ_d^T affects F^T). Any change in the fixed exporting costs between T_0 and T_1 will induce changes in average firm size - even when productivity has an untruncated Pareto distribution (as was the case in the homogeneous firm model, this situation represents a violation of ACR's macro restriction R2).

Taking the results of this subsection together, our generalization of the ACR welfare representation provides a way of quantitatively assessing whether predictions for the welfare gains from trade liberalization based on the domestic trade share and the assumption of a constant trade elasticity provide a good approximation to the true welfare gains. The success of this approximation depends on the hazard rate differential between the domestic and export markets and the degree to which firm entry responds to changes in trade costs. If firm-level data are available, these differences in hazard rates and the response of firm entry can be examined empirically. Admittedly, measuring the response of firm entry to changes in trade costs raises challenges. But these challenges are similar to those faced in estimating a partial trade elasticity and recovering the change in trade induced by a

¹⁰Similarly, Hsieh and Ossa (2011) show that this special feature of the untruncated Pareto productivity distribution extends to a multi-sector setting with asymmetric countries. See also Feenstra (2010).

change in variable trade costs alone. Head, Mayer and Thoenig (2013) propose a goodness of fit test of firm size distribution to the Pareto distribution that is similar to checking for changes in the hazard rate (which is constant under Pareto).

Even in cases where only aggregate trade data is available, one can still estimate differences in trade elasticities across country-partner pairs. Helpman, Melitz and Rubinstein (2008) and Novy (2013) both implement gravity estimation procedures that allow for variation in trade elasticities across country-partner pairs. Both papers find that this variation is substantial. Our generalized welfare derivations highlight that the discrepancy between the predicted and true welfare effects of trade liberalization will be minimized by choosing a trade elasticity for country-partners that most closely approximates the elasticity for a country's trade with itself. If the hazard rate function $\gamma(\varphi_j)$ is monotonic in the cutoff φ_j , then the hazard differential $\gamma(\varphi_d) - \gamma(\varphi_x)$ will be minimized when φ_x is closest to φ_d , which occurs for the trading partner with the highest share of exporting firms.

5 Trade Policy Evaluation

In the previous section, we showed how micro-structure affects the measurement of the welfare gains from trade – even when conditioning on a given trade elasticity and a given domestic trade share. This led to discrepancies between the true welfare effects of trade liberalization and those predicted by the ACR formula. We now illustrate more concretely how such discrepancies may arise when evaluating the welfare gains generated from a few specific trade liberalization scenarios. Our starting point is our heterogeneous firm model with two symmetric countries developed in Section 2. We consider the welfare gains from liberalizing trade first from trade regime $T_0(\tau_0, f_{x0})$ to $T_1(\tau_1, f_{x1})$, and then to $T_2(\tau_2, f_{x2})$. We contrast the true welfare gains from (16) with those measured by a policy analyst who applies the ACR formula (28). We also contrast the cases of *ex post* and *ex ante* policy evaluation using a similar approach to ACR and Costinot and Rodriguez-Clare (2013).

Specifically, we assume that trade liberalization from $T_0 \rightarrow T_1$ is evaluated *ex post* so that the domestic trade shares λ^{T_0} and λ^{T_1} are observed, and the (arc) trade elasticity therefore can be directly measured as:¹¹

$$\hat{\theta}_{01} = -\frac{\ln\left(\frac{1-\lambda^{T_1}}{\lambda^{T_1}}\right) - \ln\left(\frac{1-\lambda^{T_0}}{\lambda^{T_0}}\right)}{\ln \tau_1 - \ln \tau_0}. \quad (35)$$

The ACR predicted welfare gains from trade are then $\hat{\mathbb{W}}_{01} = (\lambda^{T_0}/\lambda^{T_1})^{1/\hat{\theta}_{01}}$. On the other hand, we assume that trade liberalization from $T_1 \rightarrow T_2$ is evaluated *ex ante*, so the domestic trade share λ^{T_2} is unobserved and is recovered from the model using the elasticity $\hat{\theta}_{01}$. That is, $\hat{\lambda}_{T_2}$ solves:

$$\hat{\theta}_{01} = -\frac{\ln\left(\frac{1-\hat{\lambda}_{T_2}}{\hat{\lambda}_{T_2}}\right) - \ln\left(\frac{1-\lambda^{T_1}}{\lambda^{T_1}}\right)}{\ln \tau_2 - \ln \tau_1}. \quad (36)$$

Ex ante, the ACR welfare derivation yields predicted welfare gains from trade given by $\hat{\mathbb{W}}_{12} =$

¹¹When the distribution of productivity draws $G(\varphi)$ is an untruncated Pareto – a necessary condition for the ACR macro restrictions to hold – there is no difference between the full and partial trade elasticities θ and ϑ .

$(\lambda^{T_1}/\hat{\lambda}^{T_2})^{1/\hat{\theta}_{01}}$. We assume that the trade costs in the trade regimes T_0 and T_1 are high enough to generate export market selection. However, we do not impose this restriction on the hypothetical trade regime T_2 : A policy analyst may be interested in evaluating the welfare gains from trade for scenarios that go most (or all) of the way to free trade.

5.1 Scenario 1: Untruncated Pareto Productivity Distribution

We assume that $G(\varphi)$ is distributed untruncated Pareto(k) and initially assume no change in the fixed export costs: $f_{x0} = f_{x1} = f_{x2}$. Then, the measured elasticity $\hat{\theta}_{01}$ will recover the constant elasticity k , and \hat{W}_{01} will exactly measure the “true” welfare gains from the *ex post* liberalization $T_0 \rightarrow T_1$. Also, if the hypothetical trade regime T_2 features export market selection (the trade costs τ_2 and f_{x2} are high enough), then $\hat{\theta}_{01} = k$ will also capture the trade elasticity between T_1 and T_2 , and the analyst would also correctly predict the attained domestic trade share in regime T_2 . Thus, the predicted welfare gain \hat{W}_{12} will again recover the “true” welfare gain from (16).

However, if the trade costs in regime T_2 are low enough – such that all firms export in T_2 – then the *ex ante* welfare evaluation will be incorrect. The true trade elasticity drops from k to $\sigma - 1$ once there is no export market selection, and this change will not be reflected in the elasticity $\hat{\theta}_{01}$. Consequently, the analyst will also incorrectly predict the attained domestic trade share in regime T_2 . This transition between the case of export market selection and no selection represents a violation of ACR’s macro restriction (R3). Yet, this transition occurs endogenously in our model; the only structural change is a reduction in trade costs.

We now consider the case where trade liberalization involves a change in the fixed exporting costs. Then, the measured trade elasticity $\hat{\theta}_{01}$ will over predict the true elasticity k , because the former confounds the effect of both the variable and fixed trade costs. In turn, this will generate discrepancies between the true and predicted welfare gains from trade liberalization for both the *ex post* and *ex ante* policy evaluations. This change in fixed costs does not represent any violation of ACR’s macro restrictions; it represents a measurement issue for the trade elasticity.¹²

5.2 Scenario 2: Truncated Pareto Productivity Distribution

We now assume that $G(\varphi)$ is distributed Pareto, but that there is a finite upper bound to the support of the productivity distribution ($\varphi_{\max} < \infty$). We return to our initial assumption of no change in the fixed exporting cost (so $f_{x0} = f_{x1} = f_{x2}$). As we highlighted in the previous section, this small departure from an untruncated Pareto distribution changes the derivation of the welfare gains from trade from the ACR formula (28) to (33). We abstract from measurement issues for the trade elasticity induced by differences between arc versus point elasticities, and between the full versus the partial trade elasticity. Thus, we assume that the measured elasticity $\hat{\theta}_{01}$ yields an accurate estimate for any

¹²ACR note that, in a multi-country world, the trade elasticity can be recovered from the estimation of a gravity equation when variations in bilateral tariffs are observed. This estimation also requires that the variation in *any* fixed trade cost that is correlated with the tariffs is also observed. To our knowledge, this data is not available.

point partial trade elasticity ϑ between τ_0 and τ_1 .¹³

Nevertheless, the analyst will obtain an incorrect measure of the *ex post* welfare gains from trade because both the difference in the hazard rates $\gamma(\varphi_d) - \gamma(\varphi_x)$ and the response of entry $\Delta \ln M_e$ will be non-zero. In our quantitative analysis in the next section, we find the former effect to be substantial while the latter effect is relatively small for our parametrization.

Lastly, we consider the evaluation of an *ex ante* hypothetical trade liberalization from T_1 to T_2 . The same discrepancies between the true and predicted welfare gains from trade liberalization as mentioned above for the *ex post* evaluation will also apply to the *ex ante* case. In addition, the measured elasticity $\hat{\theta}_{01}$ will no longer apply to changes in τ between τ_1 and τ_2 – even abstracting from the differences between arc versus point elasticities and between the full and partial elasticities: All of these elasticities are different for $\tau \in (\tau_1, \tau_2)$ relative to $\tau \in (\tau_0, \tau_1)$.¹⁴ This also means that the analyst will incorrectly predict the domestic trade share in regime T_2 .

In closing, we highlight that all the different trade liberalization scenarios that we have described in this section satisfy the assumptions for our open economy comparative static exercise described in Subsection 3.2. Thus, in all these cases, the change in aggregate productivity induced by endogenous selection (for both the domestic and export market) is directly linked to *new aggregate* welfare gains from trade relative to a case where endogenous selection was inoperative and aggregate productivity remained constant. In this sense, firm heterogeneity and the associated endogenous selection effects from trade generate new aggregate welfare implications.

We previously described the tension between this comparative static and the ACR result that – conditional on an observed domestic trade share and trade elasticity – micro structure is irrelevant for the measurement of the aggregate welfare gains from trade. In this section, we have shown that this tension disappears when one considers several different trade liberalization scenarios that are just as empirically relevant (we might even argue more empirically relevant) than one that exhibits the ACR result. In all the other scenarios, we find that micro structure matters for the aggregate welfare gains from trade – even after conditioning on the observed domestic trade share and trade elasticity.

6 Quantitative Relevance

In this section, we examine the quantitative relevance of our results. In subsection 6.1, we show that our theoretical comparative static is associated with quantitatively relevant differences in welfare between the heterogeneous and homogeneous firm models. In subsection 6.2, we show that the reduced-form nature of the trade elasticity is consequential for the practical evaluation of trade policies.

¹³In our calibration based on a truncated Pareto distribution in the next section, we find that differences between the arc partial elasticity and the arc full elasticity have a relatively small effect on the measured welfare gains from trade.

¹⁴A growing body of empirical research provides evidence of a variable trade elasticity, including Helpman, Melitz and Rubinstein (2008), Novy (2013), and Head, Mayer and Thoenig (2013). Simonovska and Waugh (2013) provides evidence that the estimated trade elasticity is model dependent. Imbs and Méjean (2009) and Ossa (2012) argue that aggregate trade elasticities are influenced by sectoral composition in multi-sector trade models.

6.1 Theoretical Comparative Static

In this subsection, we compare the welfare properties of the heterogeneous and homogeneous firm models holding all structural parameters other than the productivity distribution constant between the two models. We assume an untruncated Pareto distribution for productivity in the heterogeneous firm model; this satisfies the ACR macro restrictions so long as trade costs are high enough to generate export market selection. The homogeneous firm model satisfies the ACR restrictions so long as the fixed exporting cost remains constant. We choose standard values for the heterogeneous firm model's parameters based on central estimates from the existing empirical literature and moments of the U.S. data. Those same structural parameters then carry over to the homogeneous firm case, except that we set a degenerate productivity distribution as described below.

We set the elasticity of substitution between varieties $\sigma = 4$, which is consistent with the estimates using plant-level U.S. manufacturing data in Bernard, Eaton, Jensen and Kortum (2003). Since productivity has an untruncated Pareto distribution and firm revenue is a power function of firm productivity, firm revenue also has an untruncated Pareto distribution: $G(r) = 1 - \left(\frac{r_d}{r}\right)^{\frac{k}{\sigma-1}}$, where r_d is the revenue of the least productive firm. Existing empirical estimates suggest that the firm size distribution can be approximated by a Pareto distribution with a shape parameter close to one (see for example Axtell 2001). Therefore, we set the Pareto shape parameter for firm productivity $k = 4.25$, which ensures a Pareto shape parameter for firm revenue close to one and that log firm revenue has a finite mean ($\frac{k}{\sigma-1} = 1.42 > 1$). A choice for the Pareto scale parameter is equivalent to a choice of units in which to measure productivity, and hence we normalize $\varphi_{\min} = 1$.

We consider trade between two symmetric countries, and choose labor in one country as the numeraire ($w = 1$), which implies that the wage in both countries is equal to one. With an untruncated Pareto productivity distribution, scaling L and $\{f_e, f_d, f_x\}$ up or down by the same proportion leaves the productivity cutoffs $\{\varphi_d^T, \varphi_x^T\}$ and the mass of entrants unchanged (M_e), and merely scales average firm size (\bar{r}) up or down by the same proportion. Therefore we set L equal to the U.S. labor force and normalize f_d to one. With an untruncated Pareto productivity distribution, the sunk entry cost f_e affects the absolute levels of the productivity cutoffs and welfare but not their relative levels for different values of trade costs. As a result, the relative comparisons below are invariant to the choice of f_e , and hence we normalize f_e to one.

We calibrate τ to match the average fraction of exports in firm sales in U.S. manufacturing ($\frac{\tau^{1-\sigma}}{1+\tau^{1-\sigma}} = 0.14$, as reported in Bernard, Jensen, Redding and Schott 2007), which implies $\tau = 1.83$ (which is in line with the estimate of 1.7 in Anderson and van Wincoop 2004). Given our choice for the parameters $\{\sigma, k, \varphi_{\min}, f_d, f_e, \tau\}$, we choose f_x to ensure that the model is consistent with the average fraction of U.S. manufacturing firms that export (0.18, as reported in Bernard, Jensen, Redding and Schott 2007).

We choose the degenerate productivity distribution in the homogeneous firm model so that the two models generate the same aggregate variables in an initial equilibrium. In our baseline specification here, we do so for an initial open economy equilibrium using our calibrated values of trade costs

of $\tau = 1.83$ and $f_x = 0.545$. Thus we compare the heterogeneous firm model to the extended homogeneous firm model introduced in subsection 3.2. In the web appendix, we undertake a similar analysis for an initial closed economy equilibrium, as analyzed in subsection 3.1.

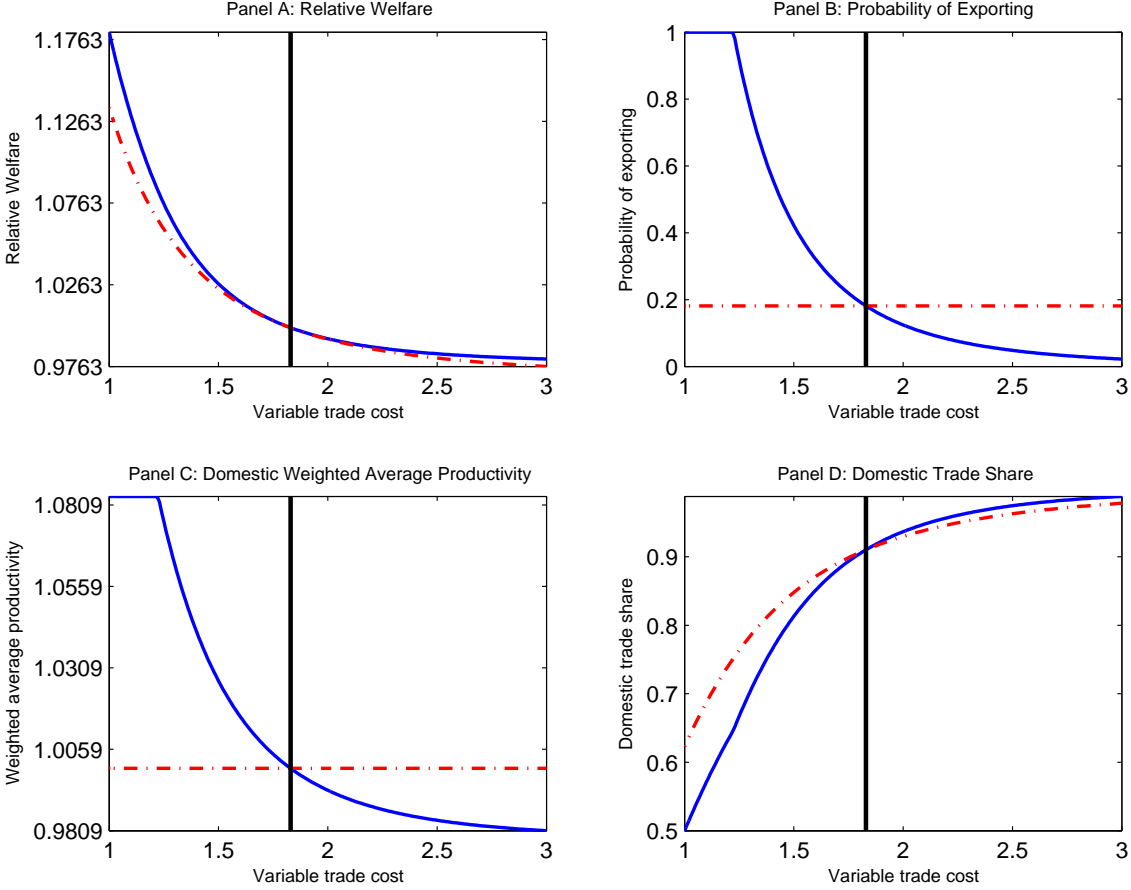
In our baseline specification here, we solve for the initial open economy equilibrium of the heterogeneous firm model, including the probability of successful firm entry $\left[1 - G\left(\varphi_d^{T_0}\right)\right]$, the proportion of exporting firms $\left[1 - G\left(\varphi_x^{T_0}\right)\right] / \left[1 - G\left(\varphi_d^{T_0}\right)\right]$, weighted average productivity in the export market $\left(\tilde{\varphi}_x^{T_0}\right)$, weighted average productivity in the domestic market $\left(\tilde{\varphi}_d^{T_0}\right)$, and the weighted average productivity of domestic firms $\left(\tilde{\varphi}_{dx}^{T_0}\right)$. In the extended homogeneous firm model, we choose the probabilities of entry and exporting and the weighted average productivities for domestic and exporting firms to equal to their values in the initial open economy equilibrium of the heterogeneous firm model: $\left[1 - \bar{G}_x\right] = \left[1 - G\left(\varphi_x^{T_0}\right)\right]$, $\bar{G}_{dx} = \left[G\left(\varphi_x^{T_0}\right) - G\left(\varphi_d^{T_0}\right)\right]$, $\bar{\varphi}_x = \tilde{\varphi}_x^{T_0}$, and $\bar{\varphi}_{dx} = \tilde{\varphi}_{dx}^{T_0}$. All parameters apart from the productivity distribution are held constant across the two models (same $f_d, f_e, f_x, \tau, L, \sigma$), which implies $\bar{F}_d = f_d + f_e / \left[1 - G\left(\varphi_d^{T_0}\right)\right]$. The key difference between the two models is that the market entry probabilities and weighted average productivities in each market respond to changes in trade costs in the heterogeneous firm model. In contrast, in the extended homogeneous firm model, these probabilities and weighted average productivities are parameters.

In Figure 1, we show the effects of adjusting variable trade costs from their calibrated value of $\tau^{T_0} = 1.83$ to values of $\tau^{T_1} \in [1, 3]$ (for which there is trade in both models).¹⁵ Panel A displays relative welfare, measured as welfare for each value of variable trade costs relative to welfare in the initial open economy equilibrium: $\mathbb{W}^{T_1} / \mathbb{W}^{T_0}$; Panel B shows the probability of exporting: $\chi^{T_1} = \left[1 - G\left(\varphi_x^{T_1}\right)\right] / \left[1 - G\left(\varphi_d^{T_1}\right)\right]$; Panel C displays weighted average productivity in the domestic market: $\tilde{\varphi}_d^{T_1} = \left[\frac{G\left(\varphi_x^{T_1}\right) - G\left(\varphi_d^{T_1}\right)}{1 - G\left(\varphi_d^{T_1}\right)}\right] \tilde{\varphi}_{dx}^{T_1} + \left[\frac{1 - G\left(\varphi_x^{T_1}\right)}{1 - G\left(\varphi_d^{T_1}\right)}\right] \tilde{\varphi}_x^{T_1}$; Panel D shows the domestic trade share: λ^{T_1} . The solid blue line corresponds to values in the heterogeneous firm model, while the red dashed line corresponds to values in the extended homogeneous firm model.

As shown in Panel A, welfare in the two models is the same for the calibrated value of variable trade costs, but is strictly higher in the heterogeneous firm model than in the extended homogeneous firm model for all other values of variable trade costs. Therefore the welfare gains (losses) from reductions (increases) in variable trade costs are greater (smaller) in the heterogeneous firm model than in the extended homogeneous firm model. The differences in welfare between the two models are quantitatively relevant for empirically plausible changes in variable trade costs. A reduction in variable trade costs from $\tau = 1.83$ to $\tau = 1$ generates welfare that is five percentage points of real GDP higher in the heterogeneous firm model than in the extended homogeneous firm (this represents almost a third of the overall welfare gains of 17 percentage points in the heterogeneous firm model).

As shown in Panels B and C, the source of these welfare differences is endogenous selection into the domestic and export markets in the heterogeneous firm model. For the calibrated value of variable trade costs of $\tau = 1.83$, the probability of exporting and weighted average productivity in the domestic

¹⁵For brevity, we concentrate on changes in variable trade costs, but we find a similar pattern of results for changes in fixed exporting costs, as shown in the calibration in the working paper version of this paper.



Note: Blue solid line heterogeneous firm model. Red dashed line extended homogeneous firm model. Same structural parameters.

Figure 1: Reductions in variable trade costs in the heterogeneous firm model with an untruncated Pareto distribution and in the extended homogeneous firm model

market are the same in the two models. As variable trade costs fall from their calibrated value to $\tau = 1$, the probability of exporting rises from 0.18 to 1 in the heterogeneous firm model, but remains constant in the extended homogeneous firm model (Panel B). Additionally, weighted average productivity in the domestic market rises by around 8 percent in the heterogeneous firm model, while remaining constant in the extended homogeneous firm model (Panel C).

As shown in Panel D, the domestic trade share is the same in the two models for the calibrated value of variable trade costs. Given the same structural parameters, the elasticity of trade with respect to variable trade costs is higher in the heterogeneous firm model with an untruncated Pareto distribution (k) than in the extended homogeneous firm model ($\sigma - 1$). Therefore, the domestic trade share is lower in the heterogeneous firm model for variable trade costs below the calibrated value, and is higher in the heterogeneous firm model for variable trade costs above the calibrated value.

Taken together, these results show that the different aggregate welfare predictions of the heterogeneous and homogeneous firm models in our theoretical comparative static are of quantitative relevance.

In the web appendix, we show that we find similar results when we calibrate the heterogeneous and homogeneous firm models to an initial closed economy equilibrium. Opening the closed economy to trade, we find that the difference in the welfare gains from trade between the two models can be as large as the welfare gains from trade in the homogeneous firm model.

6.2 Practical Evaluation of Trade Policies

We now examine the quantitative implications of the reduced-form nature of the trade elasticity for the practical evaluation of trade policies. We assume a truncated Pareto productivity distribution in the heterogeneous firm model (Scenario 2 from Section 5). We keep all of the other parameters in the heterogeneous firm model the same as for the untruncated Pareto distribution in the previous subsection. Given our choice of a lower bound to the support of the productivity distribution of $\varphi_{\min} = 1$, we now assume an upper bound to the support of the productivity distribution of $\varphi_{\max} = 4$.

In Figure 2, we examine each of the components of the proportional change in welfare (33) for this truncated Pareto productivity distribution. Panel A shows the partial trade elasticity (ϑ); Panel B displays the hazard differential between the domestic and export markets ($\gamma(\varphi_d) - \gamma(\varphi_x)$); Panel C shows the domestic trade share (λ); Panel D displays the mass of entrants (M_e). We change variable trade costs from their calibrated value of $\tau^{T_0} = 1.83$ to values of $\tau^{T_1} \in [1, 4]$ and examine the effect on each of these components of welfare.

In the special case in which the upper bound to the support of the productivity distribution converges to infinity ($\varphi_{\max} \rightarrow \infty$), the truncated Pareto distribution converges to an untruncated Pareto distribution. In this special case, the partial trade elasticity (ϑ) is constant and equal to the full trade elasticity (θ), which is equal to the Pareto shape parameter ($k = 4.25$). Furthermore, in this special case, the mass of entrants depends only on parameters and hence is constant.

In contrast, for a truncated Pareto distribution with a finite upper bound to the support of the productivity distribution ($\varphi_{\max} < \infty$), the partial trade elasticity (ϑ) is variable and differs from both the full trade elasticity (θ) and the Pareto shape parameter ($k = 4.25$). As we vary variable trade costs from one to four, the partial trade elasticity in Panel A ranges from less than five to more than sixteen.¹⁶ As variable trade costs increase, the export productivity cutoff (φ_x^T) rises, which increases the export hazard ($\gamma(\varphi_x^T)$) and hence in turn increases the partial trade elasticity (ϑ). As variable trade costs become sufficiently large that the export productivity cutoff approaches the upper bound to the support of the productivity distribution ($\varphi_x \rightarrow \varphi_{\max}$), the partial trade elasticity converges towards infinity ($\vartheta \rightarrow \infty$). As variable trade costs become sufficiently small that all firms export, the export and domestic productivity cutoffs become equal to one another ($\varphi_x^T = \varphi_d^T$) and independent of variable trade costs. For this range of variable trade costs for which all firms export, the partial trade elasticity (ϑ) is constant. Taken together, these results suggest that the partial trade elasticity can vary quite substantially from one context to another.

As shown in Panel B, these changes in variable trade costs have implications for the difference in

¹⁶Empirically, Novy (2013) estimates variable trade elasticities that range from less than five to more than twenty.

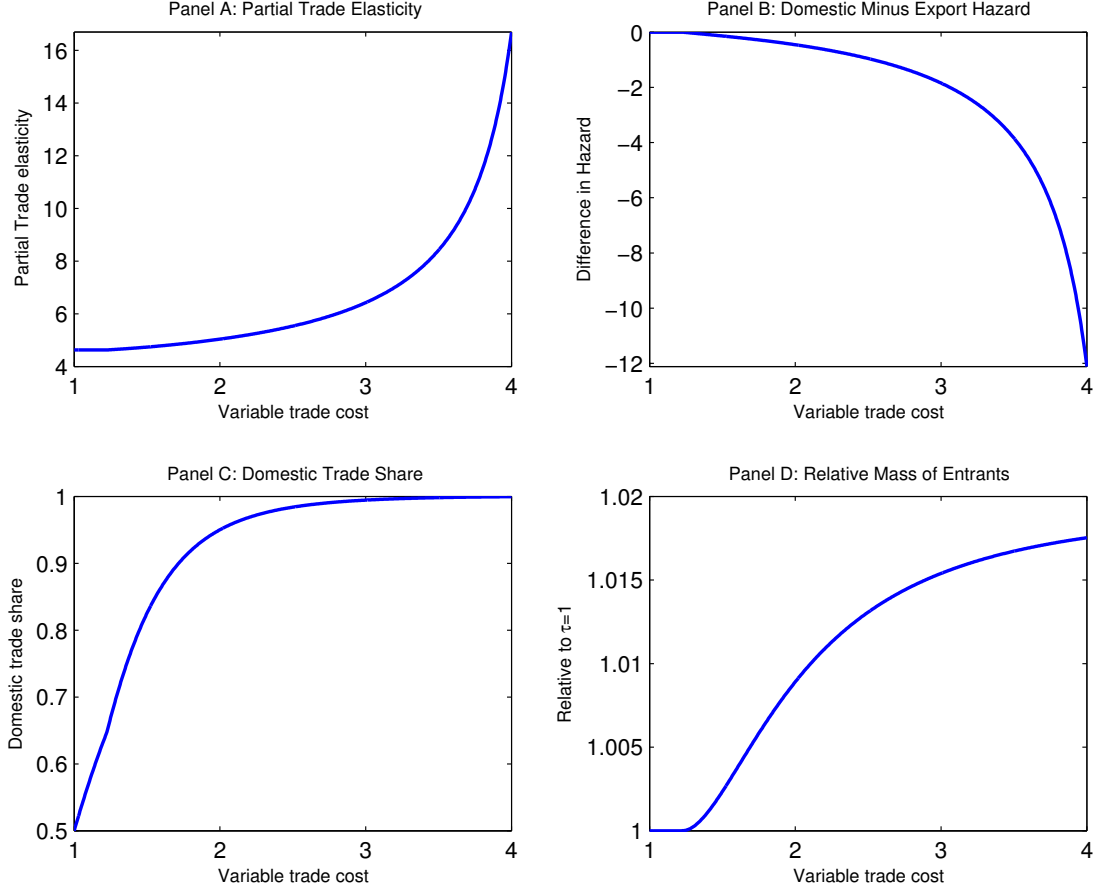


Figure 2: Trade elasticity, domestic trade share and mass of entrants in the heterogeneous firm model with a truncated Pareto productivity distribution

hazard functions between the domestic and export markets ($\gamma(\varphi_d^T) - \gamma(\varphi_x^T)$). As variable trade costs increase, the resulting rise in the export productivity cutoff (φ_x^T) *increases* the hazard function in the export market ($\gamma(\varphi_x^T)$), but the associated reduction in the domestic productivity cutoff (φ_d^T) *reduces* the hazard function in the domestic market ($\gamma(\varphi_d^T)$). As a result, as we vary variable trade costs from one to four, the hazard rate differential between the two markets ranges from zero to minus twelve. As variable trade costs become sufficiently large that the export productivity cutoff approaches the upper bound to the support of the productivity distribution ($\varphi_x \rightarrow \varphi_{\max}$), the difference between the two hazard functions converges towards minus infinity ($\gamma(\varphi_d^T) - \gamma(\varphi_x^T) \rightarrow -\infty$). As variable trade costs become sufficiently small that all firm export, the export and domestic productivity cutoffs become equal to one another ($\varphi_x^T = \varphi_d^T$), and the difference between the two hazard functions becomes equal to zero ($\gamma(\varphi_d^T) - \gamma(\varphi_x^T) = 0$). These results highlight the importance of controlling for differences across markets in the hazard function of the distribution of log firm size when computing the welfare gains from trade liberalization using (33).

As shown in Panel C, increases in variable trade costs raise the domestic trade share, which converges towards one as variable trade costs rise towards four, and converges towards a value of one half as variable trade costs fall towards one (reflecting country symmetry). As variable trade costs become sufficiently small that all firms export, the partial trade elasticity changes from $\vartheta = (\sigma - 1) + \gamma(\varphi_x^T)$ to $\vartheta = \sigma - 1$.

As shown in Panel D, increases in variable trade costs raise the mass of entrants, which is shown for each value of variable trade costs in the figure relative to its value for $\tau = 1$. With a truncated Pareto distribution, higher variable trade costs reduce average firm size conditional on successful entry. With a fixed labor endowment, this in turn leads to a larger mass of entrants. For the parameterization considered here, these changes in the mass of entrants are relatively small, with the mass of entrants increasing by less than 2 percent as variable trade costs increase from one to four. As variable trade costs become sufficiently small that all firms export, the export and domestic productivity cutoffs become equal to one another ($\varphi_x^T = \varphi_d^T$) and independent of variable trade costs. Therefore, for this range of variable trade costs, both average firm size and the mass of entrants are constant.

We now examine the quantitative implications of the above changes in micro structure for the evaluation of trade policies. Table 1 compares the true welfare gains from trade liberalization with a truncated Pareto distribution to the welfare gains that would be predicted by a policy analyst who falsely assumed a constant trade elasticity and applied the ACR formula. We examine trade liberalization from high variable trade costs for which the economy is relatively closed ($\tau = 4$ and $\lambda = 0.999$), through intermediate values of variable trade costs ($\tau = 1.5$ and $\lambda = 0.826$), and to low values of variable trade costs for which the economy is relatively open but still only some firms export ($\tau = 1.25$ and $\lambda = 0.669$).

In Column (1), we report the true relative change in welfare ($\mathbb{W}^{T_1}/\mathbb{W}^{T_0}$) in the heterogeneous firm model with a truncated Pareto distribution (as computed using (32)). Reducing variable trade costs from $\tau = 4$ to $\tau = 1.25$ increases welfare by 8.72 percent, which is broadly in line with estimates of the welfare gains from trade in recent quantitative trade models. Around half of these welfare gains are achieved from the reduction in variable trade costs from $\tau = 4$ to $\tau = 1.5$ (3.92 percent), with the remaining half realized from a further reduction in variable trade costs to $\tau = 1.25$ (4.61 percent). Since the variable partial trade elasticity is increasing in variable trade costs, larger welfare gains are

	(1)	(2)	(3)	(4)
Trade Liberalization	Actual (Truncated Pareto)	Predicted (ACR formula) ϑ^{start}	Predicted (ACR formula) $\vartheta^{\text{average}}$	Predicted (ACR formula) $\vartheta^{\text{start}}_{\text{end}}$
$\tau = 4$ to $\tau = 1.25$	108.72%	102.44%	106.15%	106.80%
$\tau = 4$ to $\tau = 1.5$	103.92%	101.15%	102.80%	103.05%
$\tau = 1.5$ to $\tau = 1.25$	104.61%	104.54%	104.59%	104.61%

Table 1: Actual and Predicted Welfare Gains from Trade Liberalization

generated from a given percentage reduction in variable trade costs when the economy is relatively open than when it is relatively closed. This property of a variable trade elasticity has important implications for the evaluation of future efforts at multilateral trade liberalization. Even if variable trade costs already have been reduced to relatively low levels, this does not necessarily mean that most of the welfare gains from reductions in variable trade costs already have been achieved.

In Column (2), we report the results of an *ex ante* policy evaluation under the (false) assumption of a constant trade elasticity. We consider a policy analyst who has access to estimates of the partial trade elasticity for an initial value of trade costs (ϑ^{start}). The policy analyst considers each of the reductions in variable trade costs (e.g. from $\tau = 4$ to $\tau = 1.5$) and uses the ACR formula to predict the welfare effects of these trade liberalizations based on the observed change in the domestic trade share and the assumption of a constant trade elasticity.

For trade liberalizations starting from high variable trade costs (the first and second rows), we find substantial discrepancies between the true and predicted welfare gains from trade liberalization. Reducing variable trade costs from $\tau = 4$ to $\tau = 1.25$ is predicted in Column (2) to increase welfare by 2.44 percent (a discrepancy of more than six percentage points or more than 70 percent). These discrepancies arise because the true trade elasticity is variable rather than constant and because the hazard function differs between the domestic and export markets. For high values of variable trade costs, the partial trade elasticity changes substantially across different values of trade costs (Panel A of Figure 2) and the difference in the hazard function between the domestic and export market is large (Panel B of Figure 2). In contrast, for reductions in variable trade costs from intermediate to low values (the third row), we find that the predicted and true welfare effects of trade liberalization are relatively close to one another (a discrepancy of less than one percentage point). At these lower values of variable trade costs, the partial trade elasticity is relatively stable (Panel A of Figure 2), and the difference in the hazard function between the domestic and export markets is small (Panel B of Figure 2), because the export and domestic productivity cutoffs are close together.

In Columns (3) and (4), we report the results of an *ex post* policy evaluation under the (false) assumption of a constant trade elasticity. We consider a policy analyst who has access to an estimate of the average trade elasticity in between the start and end values of variable trade costs. The policy analyst considers each of the reductions in variable trade costs (e.g. from $\tau = 4$ to $\tau = 1.5$) and uses the ACR formula to predict the welfare effects of these trade liberalizations based on the observed change in the domestic trade share and the estimated average trade elasticity. We consider two different estimates for the average trade elasticity. In Column (3), we compute an average partial trade elasticity by considering variable trade costs at intervals of 0.005, evaluating the partial trade elasticity at each of these points, and taking the arithmetic mean of the partial trade elasticities across these points ($\vartheta^{\text{average}}$). In Column (4), we compute an average full trade elasticity by evaluating the logarithmic percentage growth in trade between the start and end values of trade costs and dividing this by the logarithmic percentage reduction in variable trade costs ($\theta_{\text{end}}^{\text{start}}$). Although the estimated average trade elasticity in Column (4) is a full elasticity rather than a partial elasticity, in practice we

find similar results in both Columns (3) and (4).

For trade liberalization starting from high variable trade costs (the first and second rows), we continue to find quantitatively relevant discrepancies between the true and predicted welfare gains from trade liberalization. Reducing variable trade costs from $\tau = 4$ to $\tau = 1.25$ is predicted in Column (3) to increase welfare by 6.15 percent (a discrepancy of over 2.5 percentage points or around 29 percent of the true welfare gain from trade liberalization). In contrast, for reductions in variable trade costs from intermediate to low values (the third row), we find that the predicted and true welfare effects of trade liberalization are relatively close to one another (a discrepancy of less than one percentage point). Again this reflects the relative stability of the partial trade elasticity (Panel A of Figure 2) and the small difference between the domestic and export hazards (Panel B of Figure 2) at low values of variable trade costs. Unsurprisingly, the difference between the true and predicted welfare effects of trade liberalization is smaller using an average estimated trade elasticity in an *ex post* evaluation than using an initial estimated trade elasticity in an *ex ante* evaluation.

Key takeaways from this section are that both the partial trade elasticity and the hazard differential between the domestic and export markets can vary substantially across different values for variable trade costs (and hence in a multi-country world across relatively open and relatively closed economies). Taking a trade elasticity estimated from a relatively closed economy and applying this elasticity to a relatively open economy without controlling for the difference in hazard functions between the two markets can lead to quantitatively relevant discrepancies between the predicted and true welfare effects of trade liberalization in both *ex ante* and *ex post* evaluations. In contrast, taking a trade elasticity estimated from a relatively open economy and applying to another relatively open economy provides a much closer approximation to the true welfare effects of trade liberalization.

We focus our quantitative analysis in this subsection on the truncated Pareto distribution to highlight that only a small departure in assumptions from an untruncated Pareto distribution can induce substantial variation in partial trade elasticities and substantial differences in the hazard function between the domestic and export markets. But the point that the partial trade elasticity is variable and the hazard function differs across markets is much more general, and also applies for example with a log normal distribution, as examined in Head, Mayer and Thoenig (2013).

7 Conclusions

We examine the extent to which firm heterogeneity matters for the aggregate welfare gains associated with trade. We develop comparative static exercises that change the degree of firm heterogeneity while holding all other structural parameters constant. We discuss how the induced differences in aggregate welfare isolate the contribution of the endogenous response in aggregate productivity. This productivity response is driven by the endogenous entry/exit decisions of firms into/out of both the domestic and export market. This is a new channel for welfare gains that is absent in models without firm heterogeneity. We find that the contribution of this channel to welfare is always positive. Thus, the endogenous productivity adjustment generates higher welfare gains in response to trade liberaliza-

tion (and consequently also smaller welfare losses from increases in trade protection). Quantitatively, we find that the contribution of this additional welfare channel is substantial when we calibrate our model to fit U.S. firm-level and aggregate data.

ACR show that, if additional restrictions are imposed on our heterogeneous firm model, then this impact of firm heterogeneity on welfare is entirely captured by the response of two endogenous variables: The trade share and trade elasticity. Then, there is a unique mapping from changes in these aggregate variables to changes in welfare – which holds independently of the extent of firm heterogeneity (and the endogenous productivity and selection responses to trade). In this case, there is some tension between our theoretical comparative static (where firm heterogeneity matters for the gains from trade) and ACR’s welfare evaluation based on an empirically measured trade share and trade elasticity (where firm heterogeneity is irrelevant for the welfare gains from trade).

We show that this tension disappears as soon as we consider small and empirically plausible deviations from the ACR restrictions. Then, the trade share and trade elasticity are no longer sufficient statistics for welfare: even conditioning on these variables, changes in micro structure will affect the welfare gains from trade. In these generalizations, there is also no single universal trade elasticity that can be applied to all contexts. Thus, evaluating a trade policy based on an estimated trade elasticity is subject to the Lucas Critique: This elasticity is not invariant with respect to changes in trade policy. Quantitatively, we show that assuming a constant elasticity when the true elasticity is variable can lead to substantial discrepancies between the predicted and true welfare effects of trade liberalization. Using an elasticity estimated for a local change in a similar context helps to alleviate this problem – but not entirely – because empirical moments other than this elasticity (and observed trade share) will matter for the welfare gains.

On a more positive note, our generalization of the ACR welfare representation in terms of observable data moments provides a way of quantitatively assessing the extent to which the ACR formula provides a good approximation to the true welfare gains from trade. In section 4, we show how, conditional on a trade share and trade elasticity, the welfare gains from trade are also affected by differences in the hazard rate of the distribution of log firm size between the domestic and export markets, and by the response of firm entry to changes in trade costs. If firm-level data is available, then these two additional data moments can be examined empirically. Our generalized welfare derivations also point out that the discrepancy between the predicted and true welfare gains from trade liberalization can be reduced by choosing a trade elasticity that most closely approximates the elasticity for a country’s trade with itself.

Even in this more general framework with a variable trade elasticity and a hazard rate differential, our theoretical comparative static applies. The endogenous response of aggregate productivity to changes in trade costs – which is a key feature of models with firm heterogeneity and endogenous selection – is associated with higher aggregate welfare. We conclude that new theories with heterogeneous firms do indeed have new aggregate welfare implications.

A Appendix

A.1 Proof of Proposition 2

Proof. We establish the proposition for the various possible types of open economy equilibria depending on parameter values. **(I)** First, we consider parameter values for which the representative firm does not find it profitable to export in the homogeneous firm model ($\tau (f_x/\bar{F}_d)^{1/(\sigma-1)} > 1$). For these parameter values, the proposition follows immediately from the fact that the two models have the same closed economy welfare, there are welfare gains from trade, and trade only occurs in the heterogeneous firm model. **(II)** Second, we consider parameter values for which the representative firm exports in the homogeneous firm model and there is selection into export markets in the heterogeneous firm model ($0 < \tau (f_x/\bar{F}_d)^{1/(\sigma-1)} < 1 < \tau (f_x/f_d)^{1/(\sigma-1)}$). From (16) and (24), open economy welfare is higher in the heterogeneous firm model than in the homogeneous firm model if the following inequality is satisfied:

$$\frac{(\tilde{\varphi}_d^T)^{\sigma-1} + \chi\tau^{1-\sigma} (\tilde{\varphi}_x^T)^{\sigma-1}}{\frac{f_e}{1-G(\varphi_d^T)} + f_d + \chi f_x} > \frac{(1 + \tau^{1-\sigma}) (\tilde{\varphi}_d^A)^{\sigma-1}}{\bar{F}_d + f_x}. \quad (37)$$

To show that this inequality must be satisfied, we use the open economy free entry condition in the heterogeneous firm model, which implies:

$$\begin{aligned} f_d \int_{\varphi_d^T}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] dG(\varphi) + f_x \int_{\varphi_x^T}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_x^T} \right)^{\sigma-1} - 1 \right] dG(\varphi) &= f_e, \\ f_d [1 - G(\varphi_d^T)] \left[\left(\frac{\tilde{\varphi}_d^T}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] + f_x [1 - G(\varphi_x^T)] \left[\left(\frac{\tilde{\varphi}_x^T}{\varphi_x^T} \right)^{\sigma-1} - 1 \right] &= f_e, \\ f_d \left(\frac{\tilde{\varphi}_d^T}{\varphi_d^T} \right)^{\sigma-1} + f_x \frac{1 - G(\varphi_x^T)}{1 - G(\varphi_d^T)} \left(\frac{\tilde{\varphi}_x^T}{\varphi_x^T} \right)^{\sigma-1} &= \frac{f_e}{1 - G(\varphi_d^T)} + f_d + \chi f_x. \end{aligned}$$

Using $(\varphi_x^T)^{\sigma-1} = (\varphi_d^T)^{\sigma-1} \tau^{\sigma-1} f_x/f_d$, we obtain:

$$\frac{f_d}{(\varphi_d^T)^{\sigma-1}} \left[(\tilde{\varphi}_d^T)^{\sigma-1} + \chi\tau^{1-\sigma} (\tilde{\varphi}_x^T)^{\sigma-1} \right] = \frac{f_e}{1 - G(\varphi_d^T)} + f_d + \chi f_x. \quad (38)$$

Note that the open economy free entry condition in the heterogeneous firm model also implies:

$$f_d \int_{\varphi_d^A}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] dG(\varphi) + f_x \int_{\varphi_d^A}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_x^T} \right)^{\sigma-1} - 1 \right] dG(\varphi) < f_e, \quad (39)$$

since $\varphi_d^A < \varphi_d^T < \varphi_x^T$ and

$$\begin{aligned} \left[\left(\frac{\varphi}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] < 0, & \quad \text{for } \varphi < \varphi_d^T, \\ \left[\left(\frac{\varphi}{\varphi_x^T} \right)^{\sigma-1} - 1 \right] < 0 & \quad \text{for } \varphi < \varphi_x^T. \end{aligned}$$

Rewriting (39), we have:

$$f_d [1 - G(\varphi_d^A)] \left[\left(\frac{\tilde{\varphi}_d^A}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] + f_x [1 - G(\varphi_d^A)] \left[\left(\frac{\tilde{\varphi}_d^A}{\varphi_x^T} \right)^{\sigma-1} - 1 \right] < f_e,$$

$$f_d \left(\frac{\tilde{\varphi}_d^A}{\varphi_d^T} \right)^{\sigma-1} + f_x \left(\frac{\tilde{\varphi}_d^A}{\varphi_x^T} \right)^{\sigma-1} < \frac{f_e}{1 - G(\varphi_d^A)} + f_d + f_x.$$

Using $(\varphi_x^T)^{\sigma-1} = (\varphi_d^T)^{\sigma-1} \tau^{\sigma-1} f_x/f_d$, we obtain:

$$\frac{f_d}{(\varphi_d^T)^{\sigma-1}} (1 + \tau^{1-\sigma}) (\tilde{\varphi}_d^A)^{\sigma-1} < \frac{f_e}{1 - G(\varphi_d^A)} + f_d + f_x. \quad (40)$$

From (38) and (40), we have:

$$\frac{\frac{f_d}{(\varphi_d^T)^{\sigma-1}} \left[(\tilde{\varphi}_d^T)^{\sigma-1} + \chi \tau^{1-\sigma} (\tilde{\varphi}_x^T)^{\sigma-1} \right]}{\frac{f_e}{1 - G(\varphi_d^T)} + f_d + \chi f_x} = 1, \quad (41)$$

$$\frac{\frac{f_d}{(\varphi_d^T)^{\sigma-1}} (1 + \tau^{1-\sigma}) (\tilde{\varphi}_d^A)^{\sigma-1}}{\frac{f_e}{1 - G(\varphi_d^A)} + f_d + f_x} = \frac{\frac{f_d}{(\varphi_d^T)^{\sigma-1}} (1 + \tau^{1-\sigma}) (\tilde{\varphi}_d^A)^{\sigma-1}}{\bar{F}_d + f_x} < 1,$$

which establishes that inequality (37) is satisfied. In an open economy equilibrium of the heterogeneous firm model with export market selection, welfare also can be expressed as:

$$\mathbb{W}_{\text{Het}}^T = \frac{w}{P} = \left(\frac{L}{\sigma f_d} \right)^{\frac{1}{\sigma-1}} \frac{\sigma - 1}{\sigma} \varphi_d^T. \quad (42)$$

From (42) and (24), the condition for open economy welfare to be higher in the heterogeneous firm model with export market selection than in the homogeneous firm model can be also written as:

$$\left(\frac{1}{f_d} \right)^{\frac{1}{\sigma-1}} \varphi_d^T > \left(\frac{1 + \tau^{1-\sigma}}{\bar{F}_d + f_x} \right)^{\frac{1}{\sigma-1}} \tilde{\varphi}_d^A.$$

Using (37) and (41), this (equivalent) inequality is necessarily satisfied. Since closed economy welfare is the same in the two models, and open economy welfare is higher in the heterogeneous firm model than in the homogeneous firm model, it follows that the proportional welfare gains from trade are larger in the heterogeneous firm model ($\mathbb{W}_{\text{Het}}^T/\mathbb{W}_{\text{Het}}^A > \mathbb{W}_{\text{Hom}}^T/\mathbb{W}_{\text{Hom}}^A$). **(III)** Third, we consider parameter values for which the representative firm exports in the homogeneous firm model and all firms export in the heterogeneous firm model, but fixed exporting costs are still positive ($0 < \tau (f_x/\bar{F}_d)^{1/(\sigma-1)} < \tau (f_x/f_d)^{1/(\sigma-1)} \leq 1$). This is simply a special case of **(II)** in which $\varphi_x^T = \varphi_d^T$, $\tilde{\varphi}_x^T = \tilde{\varphi}_d^T$ and $\frac{1-G(\varphi_x^T)}{1-G(\varphi_d^T)} = 1$. Therefore the same line of reasoning as in **(II)** can be used to show that the inequality (37) is satisfied and hence open economy welfare is higher in the heterogeneous firm model than in the homogeneous firm model. In this special case in which all firms export, the free entry condition in the open economy equilibrium of the heterogeneous firm model implies:

$$(f_d + f_x) \int_{\varphi_d^T}^{\varphi_{\max}^T} \left[\left(\frac{\varphi}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] dG(\varphi) = f_e,$$

$$(f_d + f_x) \int_{\varphi_d^A}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] dG(\varphi) < f_e, \quad (43)$$

since $\varphi_d^A < \varphi_d^T$ and

$$\left[\left(\frac{\varphi}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] < 0, \quad \text{for } \varphi < \varphi_d^T.$$

Rewriting (43), we obtain:

$$(f_d + f_x) \left(\frac{\tilde{\varphi}_d^A}{\varphi_d^T} \right)^{\sigma-1} < \frac{f_e}{1 - G(\varphi_d^A)} + f_d + f_x = \bar{F}_d + f_x. \quad (44)$$

In an open economy equilibrium of the heterogeneous firm model in which all firms export, welfare can be also expressed as:

$$\mathbb{W}_{\text{Het}} = \frac{\sigma - 1}{\sigma} \left(\frac{(1 + \tau^{1-\sigma}) L}{\sigma (f_d + f_x)} \right)^{\frac{1}{\sigma-1}} \varphi_d. \quad (45)$$

Therefore the condition for open economy welfare in the heterogeneous firm model without export market selection (45) to be higher than open economy welfare in the homogeneous firm model (24) can be also written as:

$$\left(\frac{1}{f_d + f_x} \right)^{\frac{1}{\sigma-1}} \varphi_d^T > \left(\frac{1}{\bar{F}_d + f_x} \right)^{\frac{1}{\sigma-1}} \tilde{\varphi}_d^A.$$

From (44), this inequality is necessarily satisfied. Since closed economy welfare is the same in the two models, and open economy welfare is higher in the heterogeneous firm model than in the homogeneous firm model, it follows that the proportional welfare gains from trade are larger in the heterogeneous firm model ($\mathbb{W}_{\text{Het}}^T / \mathbb{W}_{\text{Het}}^A > \mathbb{W}_{\text{Hom}}^T / \mathbb{W}_{\text{Hom}}^A$). **(IV)** Finally, when fixed exporting costs are zero, we have $0 = \tau (f_x / \bar{F}_d)^{1/(\sigma-1)} = \tau (f_x / f_d)^{1/(\sigma-1)}$. This is a special case of **(III)** in which $\varphi_x^T = \varphi_d^T = \varphi_d^A$, $\tilde{\varphi}_x^T = \tilde{\varphi}_d^T = \tilde{\varphi}_d^A$ and $\frac{1-G(\varphi_x^T)}{1-G(\varphi_d^T)} = 1$. In this special case of zero fixed exporting costs, the free entry condition in the open economy equilibrium of the heterogeneous firm model implies:

$$(f_d + f_x) \int_{\varphi_d^A}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_d^T} \right)^{\sigma-1} - 1 \right] dG(\varphi) = f_e,$$

$$(f_d + f_x) \left(\frac{\tilde{\varphi}_d^A}{\varphi_d^T} \right)^{\sigma-1} = \frac{f_e}{1 - G(\varphi_d^A)} + f_d + f_x = \bar{F}_d + f_x,$$

where we have used $\varphi_d^A = \varphi_d^T$. From (45) and (24), it follows immediately that open economy welfare is the same in the two models when fixed exporting costs are equal to zero. ■

A.2 Proof of Proposition 4

Proof. In the initial open economy equilibrium before the change in trade costs, (16) implies that welfare in both the heterogeneous firm model and in the extended homogeneous firm model can be written as:

$$\left(\mathbb{W}_{\text{Het}}^{T_1} \right)^{\sigma-1} = \frac{L \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} \left[\left(\tilde{\varphi}_d^{T_1} \right)^{\sigma-1} + \chi_1 \tau_1^{1-\sigma} \left(\tilde{\varphi}_x^{T_1} \right)^{\sigma-1} \right]}{\sigma \left[\frac{f_e}{1-G(\varphi_d^{T_1})} + f_d + \chi_1 f_{x1} \right]}.$$

In the new open economy equilibrium after the change in trade costs, (16) implies that welfare in the heterogeneous firm model is:

$$\left(\mathbb{W}_{\text{Het}}^{T_2}\right)^{\sigma-1} = \frac{L \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \left[\left(\tilde{\varphi}_d^{T_2}\right)^{\sigma-1} + \chi_2 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_2}\right)^{\sigma-1} \right]}{\sigma \left[\frac{f_e}{1-G\left(\varphi_d^{T_2}\right)} + f_d + \chi_2 f_{x2} \right]}.$$

In contrast, in the new open economy equilibrium after the change in trade costs, welfare in the extended homogeneous firm model is:

$$\left(\mathbb{W}_{\text{Hom}}^{T_2}\right)^{\sigma-1} = \frac{L \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \left[\left(\tilde{\varphi}_d^{T_1}\right)^{\sigma-1} + \chi_1 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_1}\right)^{\sigma-1} \right]}{\sigma \left[\frac{f_e}{1-G\left(\varphi_d^{T_1}\right)} + f_d + \chi_1 f_{x2} \right]}.$$

To show that welfare in the new open economy equilibrium is higher in the heterogeneous firm model than in the homogeneous firm model, we need to show that:

$$\frac{\left(\tilde{\varphi}_d^{T_2}\right)^{\sigma-1} + \chi_2 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_2}\right)^{\sigma-1}}{\frac{f_e}{1-G\left(\varphi_d^{T_2}\right)} + f_d + \chi_2 f_{x2}} > \frac{\left(\tilde{\varphi}_d^{T_1}\right)^{\sigma-1} + \chi_1 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_1}\right)^{\sigma-1}}{\frac{f_e}{1-G\left(\varphi_d^{T_1}\right)} + f_d + \chi_1 f_{x2}}. \quad (46)$$

To establish this inequality, we use the free entry condition in the new open economy equilibrium of the heterogeneous firm model, which implies:

$$\begin{aligned} f_d \int_{\varphi_d^{T_2}}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_d^{T_2}} \right)^{\sigma-1} - 1 \right] dG(\varphi) + f_{x2} \int_{\varphi_x^{T_2}}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_x^{T_2}} \right)^{\sigma-1} - 1 \right] dG(\varphi) &= f_e, \\ f_d \left[1 - G\left(\varphi_d^{T_2}\right) \right] \left[\left(\frac{\tilde{\varphi}_d^{T_2}}{\varphi_d^{T_2}} \right)^{\sigma-1} - 1 \right] + f_{x2} \left[1 - G\left(\varphi_x^{T_2}\right) \right] \left[\left(\frac{\tilde{\varphi}_x^{T_2}}{\varphi_x^{T_2}} \right)^{\sigma-1} - 1 \right] &= f_e, \\ f_d \left(\frac{\tilde{\varphi}_d^{T_2}}{\varphi_d^{T_2}} \right)^{\sigma-1} + f_{x2} \frac{1 - G\left(\varphi_x^{T_2}\right)}{1 - G\left(\varphi_d^{T_2}\right)} \left(\frac{\tilde{\varphi}_x^{T_2}}{\varphi_x^{T_2}} \right)^{\sigma-1} &= \frac{f_e}{1 - G\left(\varphi_d^{T_2}\right)} + f_d + \frac{1 - G\left(\varphi_x^{T_2}\right)}{1 - G\left(\varphi_d^{T_2}\right)} f_{x2}. \end{aligned}$$

Using $\left(\varphi_x^{T_2}\right)^{\sigma-1} = \left(\varphi_d^{T_2}\right)^{\sigma-1} \tau_2^{\sigma-1} f_{x2}/f_d$, we obtain:

$$\frac{f_d}{\left(\varphi_d^{T_2}\right)^{\sigma-1}} \left[\left(\tilde{\varphi}_d^{T_2}\right)^{\sigma-1} + \chi_2 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_2}\right)^{\sigma-1} \right] = \frac{f_e}{1 - G\left(\varphi_d^{T_2}\right)} + f_d + \chi_2 f_{x2}. \quad (47)$$

Note that the free entry condition in the new open economy equilibrium of the heterogeneous firm model also implies:

$$f_d \int_{\varphi_d^{T_1}}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_d^{T_1}} \right)^{\sigma-1} - 1 \right] dG(\varphi) + f_{x2} \int_{\varphi_x^{T_1}}^{\varphi_{\max}} \left[\left(\frac{\varphi}{\varphi_x^{T_1}} \right)^{\sigma-1} - 1 \right] dG(\varphi) < f_e, \quad (48)$$

since $\varphi_d^{T_1} < \varphi_d^{T_2}$ and $\varphi_x^{T_1} > \varphi_x^{T_2}$ and

$$\left[\left(\frac{\varphi}{\varphi_d^{T_2}} \right)^{\sigma-1} - 1 \right] < 0, \quad \text{for } \varphi < \varphi_d^{T_2},$$

$$\left[\left(\frac{\varphi}{\varphi_x^{T_2}} \right)^{\sigma-1} - 1 \right] > 0 \quad \text{for } \varphi_x^{T_2} < \varphi < \varphi_x^{T_1}.$$

Rewriting (48), we have:

$$f_d \left[1 - G(\varphi_d^{T_1}) \right] \left[\left(\frac{\tilde{\varphi}_d^{T_1}}{\varphi_d^{T_2}} \right)^{\sigma-1} - 1 \right] + f_{x2} \left[1 - G(\varphi_x^{T_1}) \right] \left[\left(\frac{\tilde{\varphi}_x^{T_1}}{\varphi_x^{T_2}} \right)^{\sigma-1} - 1 \right] < f_e,$$

$$f_d \left(\frac{\tilde{\varphi}_d^{T_1}}{\varphi_d^{T_2}} \right)^{\sigma-1} + f_{x2} \frac{1 - G(\varphi_x^{T_1})}{1 - G(\varphi_d^{T_1})} \left(\frac{\tilde{\varphi}_x^{T_1}}{\varphi_x^{T_2}} \right)^{\sigma-1} < \frac{f_e}{1 - G(\varphi_d^{T_1})} + f_d + \frac{1 - G(\varphi_x^{T_1})}{1 - G(\varphi_d^{T_1})} f_{x2}.$$

Using $(\varphi_x^{T_2})^{\sigma-1} = (\varphi_d^{T_2})^{\sigma-1} \tau_2^{\sigma-1} f_{x2}/f_d$, we obtain:

$$\frac{f_d}{(\varphi_d^{T_2})^{\sigma-1}} \left[\left(\tilde{\varphi}_d^{T_1} \right)^{\sigma-1} + \chi_1 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_1} \right)^{\sigma-1} \right] < \frac{f_e}{1 - G(\varphi_d^{T_1})} + f_d + \chi_1 f_{x2}. \quad (49)$$

From (47) and (49), we have:

$$\frac{\frac{f_d}{(\varphi_d^{T_2})^{\sigma-1}} \left[\left(\tilde{\varphi}_d^{T_2} \right)^{\sigma-1} + \chi_2 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_2} \right)^{\sigma-1} \right]}{\frac{f_e}{1 - G(\varphi_d^{T_2})} + f_d + \chi_2 f_{x2}} = 1,$$

$$\frac{\frac{f_d}{(\varphi_d^{T_2})^{\sigma-1}} \left[\left(\tilde{\varphi}_d^{T_1} \right)^{\sigma-1} + \chi_1 \tau_2^{1-\sigma} \left(\tilde{\varphi}_x^{T_1} \right)^{\sigma-1} \right]}{\frac{f_e}{1 - G(\varphi_d^{T_1})} + f_d + \chi_1 f_{x2}} < 1,$$

which establishes the inequality (46). ■

A.3 Proof of Proposition 5

Proof. (a) First, consider parameter values for which the representative firm exports in the homogeneous firm model and there is selection into export markets in the heterogeneous firm model ($0 < \tau (f_x/\bar{F}_d)^{1/(\sigma-1)} < 1 < \tau (f_x/f_d)^{1/(\sigma-1)}$). From (42), we have:

$$\frac{\mathbb{W}_{\text{Het}}^T}{\mathbb{W}_{\text{Het}}^A} = \frac{\varphi_d^T}{\varphi_d^A}. \quad (50)$$

In the special case of an untruncated Pareto productivity distribution and for these parameter values for which there is selection into export markets in the heterogeneous firm model, we have:

$$\frac{\varphi_d^T}{\varphi_d^A} = \left[1 + \left(\frac{1}{\tau (f_x/f_d)^{1/(\sigma-1)}} \right)^k \frac{f_x}{f_d} \right]^{1/k},$$

which can be written as:

$$\ln\left(\frac{\varphi_d^T}{\varphi_d^A}\right) = k^{-1} \ln\left[1 + \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k \frac{f_x}{f_d}\right].$$

Note that

$$\frac{d \ln(\varphi_d^T/\varphi_d^A)}{dk} = -k^{-2} \ln\left[1 + \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k \frac{f_x}{f_d}\right] - \frac{k^{-1} \ln(\tau(f_x/f_d)^{1/(\sigma-1)}) \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k \frac{f_x}{f_d}}{\left[1 + \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k \frac{f_x}{f_d}\right]} < 0, \quad (51)$$

where we have used $d(a^x)/dx = (\ln a)a^x$. Since a smaller value of k corresponds to greater productivity dispersion, it follows that greater productivity dispersion implies larger φ_d^T/φ_d^A . Second, consider parameter values for which the representative firm exports in the homogeneous firm model and all firms export in the heterogeneous firm model, but fixed exporting costs are still positive ($0 < \tau(f_x/\bar{f}_d)^{1/(\sigma-1)} < \tau(f_x/f_d)^{1/(\sigma-1)} \leq 1$). From (42) and (45), we have:

$$\frac{\mathbb{W}_{\text{Het}}^T}{\mathbb{W}_{\text{Het}}^A} = \left(\frac{(1 + \tau^{1-\sigma})f_d}{f_d + f_x}\right)^{\frac{1}{\sigma-1}} \frac{\varphi_d^T}{\varphi_d^A}.$$

In the special case of an untruncated Pareto productivity distribution and for these parameter values for which all firms export in the heterogeneous firm model, we have:

$$\frac{\varphi_d^T}{\varphi_d^A} = \left[1 + \frac{f_x}{f_d}\right]^{1/k},$$

which can be written as:

$$\ln\left(\frac{\varphi_d^T}{\varphi_d^A}\right) = k^{-1} \ln\left[1 + \frac{f_x}{f_d}\right].$$

Note that

$$\frac{d \ln(\varphi_d^T/\varphi_d^A)}{dk} = -k^{-2} \ln\left[1 + \frac{f_x}{f_d}\right] < 0. \quad (52)$$

Since a smaller value of k corresponds to greater productivity dispersion, it follows that greater productivity dispersion again implies larger φ_d^T/φ_d^A . Taking (51) and (52) together and using (50), it follows that greater dispersion of firm productivity (smaller k) implies larger proportional welfare gains from opening the closed economy to trade. **(b)** Consider parameter values for which there is selection into export markets in the open economy equilibrium of the heterogeneous firm model ($\tau(f_x/f_d)^{1/(\sigma-1)} > 1$). In the special case of an untruncated Pareto productivity distribution, we have:

$$\varphi_d^T = \left(\frac{\sigma-1}{k-(\sigma-1)}\right)^{1/k} \left[\frac{f_d + \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k f_x}{f_e}\right]^{1/k} \varphi_{\min}.$$

Therefore:

$$\begin{aligned} \frac{d\varphi_d^T}{d\tau} \frac{\tau}{\varphi_d^T} d\tau &= -\frac{\left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k f_x}{f_d + \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k f_x} d\tau \\ &= -\xi d\tau. \end{aligned}$$

Hence:

$$\frac{d\left(\frac{d\varphi_d^T}{d\tau} \frac{\tau}{\varphi_d^T} d\tau\right)}{dk} = \frac{\ln\left(\tau (f_x/f_d)^{1/(\sigma-1)}\right) \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k f_x}{f_d + \left(\frac{1}{\tau(f_x/f_d)^{1/(\sigma-1)}}\right)^k f_x} (1 - \xi) d\tau,$$

which implies:

$$\begin{aligned} \frac{d\left(\frac{d\varphi_d^T}{d\tau} \frac{\tau}{\varphi_d^T} d\tau\right)}{dk} &< 0 \quad \text{for} \quad d\tau < 0, \\ \frac{d\left(\frac{d\varphi_d^T}{d\tau} \frac{\tau}{\varphi_d^T} d\tau\right)}{dk} &> 0 \quad \text{for} \quad d\tau > 0. \end{aligned}$$

Therefore greater dispersion of firm productivity (smaller k) implies a larger elasticity of the domestic productivity cutoff with respect to reductions in variable trade costs, which from (42) implies greater proportional welfare gains from reductions in variable trade costs. By the same reasoning, greater dispersion of firm productivity (smaller k) implies a smaller elasticity of the domestic productivity cutoff with respect to increases in variable trade costs, which from (42) implies smaller proportional welfare costs from increases in variable trade costs. ■

References

- [1] Anderson, James E. and Eric van Wincoop (2004) "Trade Costs," *Journal of Economic Literature*, 42(3), 691-751.
- [2] Arkolakis, Costas, Arnaud Costinot and Andrés Rodríguez-Clare (2012) "New Trade Models, Same Old Gains," *American Economic Review*, 102(1), 94-130.
- [3] Atkeson Andrew and Ariel Burstein (2010) "Innovation, Firm Dynamics and International Trade," *Journal of Political Economy*, 118(3), 433-484.
- [4] Axtell, R. L. (2001) "Zipf Distribution of US Firm Sizes," *Science*, 293, 1818-20.
- [5] Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel S. Kortum (2003) "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268-1290.
- [6] Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding and Peter K. Schott (2007) "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105-130.
- [7] Costinot, Arnaud and Andrés Rodríguez-Clare (2013) "Trade Theory with Numbers: Quantifying the Consequences of Globalization," *Handbook of International Economics*, forthcoming.
- [8] Das, Mita, Mark J. Roberts and James R. Tybout (2007) "Market Entry Costs, Producer Heterogeneity and Export Dynamics," *Econometrica*, 75(3), 837-73.
- [9] Dhingra, Swati and John Morrow (2012) "Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity," London School of Economics, mimeograph.
- [10] Dixit, Avinash K. and Joseph E. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3), 297-308.
- [11] Eaton, Jonathan and Kortum, Samuel (2002) "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741-1779.
- [12] Feenstra, Robert C. (2010) "Measuring the Gains from Trade under Monopolistic Competition," *Canadian Journal of Economics*, February, 1-28.
- [13] Feenstra, Robert C. (2013) "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," University of California, Davis, mimeograph.
- [14] Head, Keith, Thierry Mayer and Mathias Thoenig (2013) "Welfare and Trade Without Pareto," Sciences Po, Paris, mimeograph.
- [15] Helpman Elhanan, Melitz Marc J. and Yona Rubinstein (2008) "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123(2), 441-87.

- [16] Hsieh, Chang-Tai and Ralph Ossa (2011) “A Global View of Productivity Growth in China,” *NBER Working Paper*, 16778.
- [17] Imbs, Jean and Isabelle Méjean (2009) “Elasticity Optimism,” *CEPR Discussion Paper*, 7177.
- [18] Krugman, Paul. R. (1980) “Scale Economies, Product Differentiation, and the Pattern of Trade,” *American Economic Review*, 70(5), 950-959.
- [19] Melitz, Marc J. (2003) “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71(6), 1695-1725.
- [20] Melitz, Marc J. and Stephen J. Redding (2013) “Heterogeneous Firms and Trade,” *Handbook of International Economics*, Volume 4, forthcoming.
- [21] Novy, Dennis (2013) “International Trade without CES: Estimating Translog Gravity,” *Journal of International Economics*, 89(2), 271-282.
- [22] Ossa, Ralph (2012) “Why Trade Matters After All,” *NBER Working Paper*, 18113.
- [23] Roberts, Mark J. and James R. Tybout (1997) “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs,” *American Economic Review*, 87(4), 545-564.
- [24] Simonovska, Ina and Michael E. Waugh (2013) “Different Trade Models, Different Trade Elasticities,” University of California at Davis, mimeograph.