Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity*

by

Robert C. Feenstra
University of California, Davis, and NBER

Revised, October 23, 2017

Abstract

The monopolistic competition model in international trade offers three sources of gains from trade beyond that of traditional comparative advantage: an endogenous expansion in product variety; a pro-competitive reduction in the markups charged by firms; and the self-selection of more efficient firms into exporting. Recent literature on trade with heterogeneous firms has emphasized the third of these effects, while the first two effects are ruled out when using a Pareto distribution for productivity with a support that is unbounded above, and no fixed costs. The goal of this paper is to restore a theoretical role for product variety and pro-competitive gains from trade by using a bounded Pareto distribution for productivity, and to demonstrate their empirical importance. For the U.S. economy over 1992 – 2005, we find that product variety and the reduction in markups jointly contribute about 75% to the increase in welfare resulting from trade expansion, whereas an upper bound to the selection effect is that it contributes the remaining 25% to the increase in U.S. welfare.

* Thanks are due to Costas Arkolakis, Andrés Rodríguez-Clare, Kadee Russ, Ina Simonovska and seminar participants at Boston College, MIT, UCLA, and the NBER, along with the referees for helpful comments. Vladimir Tyazhelnikov provided excellent research assistance. Financial support from the National Science Foundation is gratefully acknowledged.

1. Introduction

The monopolistic competition model in international trade offers three sources of gains from trade beyond that of traditional comparative advantage.[1] First, opening to trade may lead to an endogenous expansion in product variety, as goods not available in autarky become imported.[2] This first source is emphasized in the earliest writings by Krugman (1979) and throughout Helpman and Krugman (1985). A second source of gains emphasized by these authors is that the pro-competitive effect of trade reduces the markups charged by firms, and therefore lowers consumer prices. In order for this fall in prices to translate into a *social* gain, and not just a re-distribution from firms to consumers, we need the assumption of zero profits due to free entry. Then the reduced ratio of price to marginal cost implies a reduced ratio of average to marginal costs, too, so that firms are taking greater advantage of economies of scale. In this way, the consumer gains due to reduced markups become social gains because of the accompanying expansion of firm scale.

The third source of gains arises in the more recent models of monopolistic competition and trade with heterogeneous firms, due to Melitz (2003). In this case, trade will lead to the self-selection of more efficient firms into exporting, while less efficient firms exit the market, leading to a rise in average productivity. This third source of gains has been the focus of recent literature. For example, if we add the assumption that firm productivity is unbounded above with a Pareto distribution, as in Chaney (2008), then it can be shown that the gains from trade in the Melitz (2003) model are *entirely* due to the selection of firms because there are no gains from variety,

[1] The third source of gains from trade discussed below – the selection of more efficient firms into exporting – has an analogue in Ricardian comparative advantage where importers purchase from the most efficient exporting countries, as in Eaton and Kortum (2002).

[2] In a competitive model, product variety will expand if foreign countries produce varieties not available at home, as in the Armington (1969) specification. In a monopolistic competition model, the change in product variety is more complex because new import varieties can drive out domestic varieties, so the expansion of overall variety is not guaranteed, as we shall discuss.

and of course, there is no change in markups due to constant elasticity of substitution (CES) preferences.[3] Even without the unbounded Pareto assumption, Melitz and Redding (2015) have recently argued that the change in average productivity due to firm selection and trade in the Melitz model is what distinguishes it most clearly from the homogeneous firm model of Krugman (1980). If we allow for non-CES preferences with heterogeneous firms so that in principle a pro-competitive effect could operate, then Arkolakis, Costinot, Donaldson and Rodriguez-Clare (ACDR, 2017) have recently shown that neither this effect nor product variety leads to any gains in the absence of fixed costs; so once again, the key source of gains from trade comes from the selection of firms.[4] That result in ACDR depends on the assumption of a Pareto distribution of productivity with a support that is unbounded above.

The goal of this paper is to restore a role for product variety and pro-competitive gains from trade with heterogeneous firms, by using a bounded (or truncated) Pareto distribution for productivity. The empirical relevance of this approach is indicated by Helpman, Melitz and Rubenstein (2008), who used the bounded Pareto to obtain a gravity equation in trade that is consistent with the many instances of zero trade flows between countries.[5] It is surprising, then, that the bounded Pareto has not received more theoretical attention (though it is consistent with Melitz, 2003, and it is explored by Melitz and Redding, 2015). One reason for the popularity of

[3] Arkolakis, Demidova, Klenow and Rodriguez-Clare (2008) show that if the fixed costs of exporting are paid with labor in the importing country, then the number of import plus domestic varieties can rise or fall due to a change in the fixed or variable costs of trade. However, the *welfare impacts* of the changes in import and domestic varieties sum to zero. Feenstra (2010; 2015, pp. 165-166) shows the same result when the fixed costs are paid with labor in the exporting country. In both cases, therefore, the gains from trade are due entirely to the selection of more productive firms into exporting.

[4] ACDR find that total gains are *reduced* by the pro-competitive effect – which becomes an "anti-competitive" effect – when tastes are non-homothetic. That result is obtained because the positive overall gains lead to an expansion of demand in favor of the higher-markup varieties, which worsens the distortion as compared to the first-best with constant markups (Dhingra and Morrow, 2016). In contrast, we shall assume homothetic preferences, so this anti-competitive effect does not occur.

[5] Another motivation for using bounded productivity comes from the theory of globalization put forth by John Sutton and summarized in his Clarendon Lectures (Sutton, 2012). Sutton uses three assumptions to derive the interaction of firms as globalization proceeds, the third of which is "you can't make something out of nothing" (Sutton, 2012, p. 55). That assumption is intended to rule out unbounded productivity.

the unbounded Pareto is that, like CES preferences, it leads to highly tractable solutions for trade and welfare. A secondary goal of this paper is to show that the bounded Pareto distribution still yields tractable solutions, even with a class of preferences allowing for non-constant markups.

Specifically, we will work with a class of preferences introduced by Diewert (1976) known as the quadratic mean of order *r* (QMOR) expenditure function. This is perhaps the most general parametric form for expenditure that is dual to homothetic preferences. While it is included within the class of preferences implicitly defined by ACDR, they do not take advantage of a key feature that is used here: these preferences give an explicit functional form for the expenditure needed for one unit of utility – that is, for the cost of living. The QMOR expenditure function is introduced in section 2 where, because we are dealing with a monopolistic competition model, we assume that demand is symmetric across varieties and also that it has a finite reservation price. Given these properties, we establish the sign pattern of the parameters needed to ensure that the QMOR expenditure function is globally well-behaved: a feature that has not been assured in the mainly empirical prior applications.

Our use of the QMOR expenditure function sets this paper apart from other recent, theoretical literature dealing with variable markups in international trade. A more common choice is to use the additively separable utility function introduced by Krugman (1979), possibly with an explicit functional form for the sub-utility from each variety.[6] Zhelobodko *et al* (2012), Kichko *et al* (2013) and Dhingra and Morrow (2016) consider a broader class of additively separable functions than Krugman (1979) by allowing the elasticity of demand to be increasing or decreasing in quantity. These authors argue for a pro-competitive effect of trade in the latter case only (as assumed by Krugman and holding here). An additively separable *indirect* utility

[6] Behrens and Murata (2007, 2012) use exponential functions and the latter paper includes pro-competitive effects, while Saure (2012) and Simonovska (2015) use a logarithmic function with displaced origin.

function has been introduced by Bertoletti, Etro and Simonovska (2016), who find a very strong role for the product variety gains from trade.

Another line of literature related to this paper assumes a finite number of firms, in which case markups are endogenous even with nested-CES preferences.[7] Initiated by Atkeson and Burstein (2008), this framework is used by Edmond, Midrigan and Xu (2015) to compute the pro-competitive gains from trade between the United States and Taiwan. Specializing to the case of Bertrand competition between firms, De Blas and Russ (2015) contrast the results obtained by Bernard, Eaton, Jensen, and Kortum (2003) using an infinite number of rivals to those obtained instead with a finite number of rivals; only in the latter case does a pro-competitive effect of trade operate. Our paper is most closely related to Holmes, Hsu and Lee (2014), who also use Bertrand competition and show that if and only if the distribution of productivities is unbounded Pareto, then trade leads to gains only through selection and not through markups. In these papers, Bertrand competition occurs between firms producing perfect substitutes, so there are no gains from product variety.

Before proceeding, we should give a brief intuition as to why the pro-competitive effect of trade vanishes with heterogeneous firms, homothetic preferences, and the unbounded Pareto distribution. Suppose that we measure markups by the ratio of price to marginal cost, minus unity. The most productive firm has zero cost, but a non-zero price, so its markup is infinite. The least productive surviving firm will have its marginal cost equal to the reservation price, so its markup is zero. This range of $[0,+\infty)$ for markups applies equally well to domestic and foreign firms, even if the latter face variable trade costs. Furthermore, the distribution of markups within this range is determined by the Pareto distribution of productivity. So changes in trade costs have

---

[7] Eaton, Kortum and Sotelo (2013) also consider a model with a finite but stochastic number of firms.

no impact at all on the *distribution* of markups, from either domestic or foreign firms, but still affect the mass (or extensive margin) of exporters.[8] The fixed distribution of markups no longer holds, however, when productivity and markups are bounded above, since then the highest foreign markup depends on trade costs (so trade costs also affect the intensive margin).

Our paper proceeds as follows. We show in sections 2 and 3 that the QMOR expenditure function allows us to decompose the cost of living – and therefore welfare – into components that correspond to product variety, the pro-competitive effect, the selection effect which is captured by average firm productivity, and an additional term reflecting the *spread* of prices. In the trade environments we shall consider, we are able to establish how these components change individually and jointly due to liberalization. This allows us to establish the gains comparing autarky to frictionless trade (section 4), and for small changes in variable trade costs (section 5). Importantly, we contrast the *source* of gains with unbounded versus bounded Pareto, and show that it is only in the bounded case where the product variety and pro-competitive gains apply.

Finally, we are able to compare the magnitude of total gains from trade using unbounded versus bounded Pareto, both in theory and in practice. Measured in relation to initial utility, we find that the proportionate rise in welfare due to trade liberalization is *largest* in the unbounded Pareto case, despite the fact that neither the product variety nor the pro-competitive channels operate in this case. Constraining the Pareto distribution to be bounded allows those extra sources of gains to operate, but reduces the gains due to firm selection, so that the total gains are lower. In section 6 we apply these results to the expansion of trade for the U.S. economy over 1992 and 2005, as studied in the translog case by Feenstra and Weinstein (2017). We find that product variety and the reduction in markups jointly contribute 75% to the increase in U.S.

---

[8] As discussed in note 3, with non-homothetic preferences the distortion caused by unequal markups can be amplified or diminished when demand for different varieties changes, even with a fixed distribution of markups.

welfare resulting from trade expansion, whereas an upper-bound to the selection effect is that it contributes the remaining 25%. Section 7 concludes and the proofs of propositions and other technical material are gathered in the Appendices.

2. Consumer Preferences

Expenditure Function

We shall adopt the quadratic mean of order *r* (QMOR) expenditure function, which is defined by Diewert (1976, p. 130) over a discrete number of goods with price vector **p** as:

$$e_r(\mathbf{p}) = \left[\sum_i \sum_j b_{ij} p_i^{r/2} p_j^{r/2} \right]^{1/r}, \quad r \neq 0,$$

where r and b_{ij} are parameters. We shall consider the symmetric case where $b_{ii} = \alpha$ and $b_{ij} = \beta$ for $i \neq j$, so that the QMOR function is re-expressed over a continuum of goods indexed by ω as:

$$e_r(\mathbf{p}) = \left[\alpha \int p_\omega^r d\omega + \beta \left(\int p_\omega^{r/2} d\omega \right)^2 \right]^{1/r}, \quad r \neq 0. \tag{1}$$

This function is the expenditure needed to obtain *one unit* of utility, or the cost of living. For specific values of the parameters r, α and β, this expenditure function takes on familiar forms. For $\alpha > 0$, $\beta = 0$ and $r = (1-\sigma)$, the expenditure function is CES, so that $r < 0$ for $\sigma > 1$. For $r = 2$, we obtain a quadratic expenditure function, but without the additively separable outside good used by Melitz and Ottaviano (2008). For $r = 1$, we obtain what Diewert (1971) calls a Generalized Leontief function (since the dual to a Leontief production function is linear in prices like the first term of (1) for $r = 1$, while the second term adds generality). And as shown in Appendix A, as $r \to 0$ then (1) approaches a translog function. So the quadratic mean of order r function nests the commonly used homothetic cases.

While the special cases of the quadratic mean of order r function have been applied empirically, it has not been applied in a monopolistic competition setting. To do so, we need to recognize that demand is positive if and only if prices are less than a reservation price p^*, equal across goods since the expenditure function is symmetric. In the CES case the reservation price is infinite, but we will focus here on finite reservation prices. Goods that are not available should have their prices in (1) replaced by p^*, because that is the economically relevant price to evaluate expenditure, demand and welfare. We show in Appendix B that the reservation price equals:

$$p^* = \left(\frac{N}{N - [\tilde{N} + (\alpha / \beta)]} \right)^{2/r} \left(\int_{\Omega} \tfrac{1}{N} p_{\omega}^{r/2} d\omega \right)^{2/r} . \tag{2}$$

The second term on the right of (2) is a mean of order $r/2$ of the prices p_{ω}, also called a *power mean*, which lies between the minimum and maximum values of p_{ω}. The reservation price is above this mean price if and only if the first term on the right of (2) is greater than unity. To ensure this and also rule out the CES case of an infinite reservation price, we assume:

Assumption 1

(a) If $r < 0$ then $\alpha > 0$, $\beta < 0$ and $[\tilde{N} + (\alpha / \beta)] < 0$;

(b) If $r > 0$ then $r \leq 2$, $\alpha < 0$, $\beta > 0$ and $0 < [\tilde{N} + (\alpha / \beta)] < N$;

(c) As $r \to 0$ then $\alpha = \left(\frac{1}{\tilde{N}} - \frac{2\gamma}{r} \right)$ and $\beta = \frac{2\gamma}{r\tilde{N}}$ for any $\gamma > 0$.

It is readily confirmed that parts (a) and (b) of Assumption 1 ensure that the first term on the right of (2) exceeds unity, so the reservation price exceeds the mean price. Part (c) is consistent with (a) and (b) in the sense that either set of inequalities hold for small r so the first term of (2) is again greater than unity. Furthermore, in this limit it is shown by Diewert (1980, p. 451) and in Appendix A that the expenditure function in (1) approaches the translog form,

$$\ln e_0(\mathbf{p}) = \frac{1}{\tilde{N}}\int \ln p_\omega d\omega - \frac{\gamma}{2\tilde{N}}\int\int \ln p_\omega(\ln p_\omega - \ln p_{\omega'})d\omega d\omega' . \tag{3}$$

While we have motivated Assumption 1 by the requirement that the reservation price in (2) exceeds the mean price, in Appendix B we further show that these conditions ensure that the QMOR expenditure function is globally well-behaved provided that $r \leq 2$. We also obtain several convenient expressions for demand and expenditure, as follows.

First, we show that the share of expenditure on demand for variety ω is:

$$s_\omega(\mathbf{p}) = \frac{f(p_\omega / p^*)}{\left[e_r(\mathbf{p}) / p^*\right]^r}, \text{ with } f(p_\omega / p^*) \equiv \alpha\left(\frac{p_\omega}{p^*}\right)^r\left[1 - \left(\frac{p^*}{p_\omega}\right)^{r/2}\right]. \tag{4}$$

Because the expenditure shares integrate to unity, $\int_\Omega s_\omega(\mathbf{p})d\omega = 1$, it is immediate from (4) that we can solve for a "reduced form" expression for expenditure:

$$e_r(\mathbf{p}) = p^* \times D(\mathbf{p})^{1/r}, \text{ with } D(\mathbf{p}) \equiv \int_\Omega f(p_\omega / p^*)d\omega, \tag{5}$$

and so the expenditure shares in (4) are simplified as:

$$s_\omega(\mathbf{p}) = \frac{f(p_\omega / p^*)}{D(\mathbf{p})} . \tag{6}$$

Second, we can evaluate the elasticity of demand by differentiating (4) with respect to $p_\omega < p^*$, holding the reservation price p^* and expenditure $e_r(\mathbf{p})$ constant, obtaining:

$$\eta_\omega \equiv 1 - \frac{\partial \ln s_\omega(\mathbf{p})}{\partial \ln p_\omega} = 1 - r + \underbrace{\frac{r}{2}\left\{\left(\frac{p^*}{p_\omega}\right)^{r/2} \Big/ \left[\left(\frac{p^*}{p_\omega}\right)^{r/2} - 1\right]\right\}}_{>0}. \tag{7}$$

The final term on the right goes to zero in the CES case as $\beta \to 0$ and $p^* \to \infty$, so $\eta_\omega \to 1 - r = \sigma$.

But with $\beta \neq 0$ under Assumption 1, this term is positive for $p_\omega < p^*$ and so $\eta_\omega > 1 - r$.

Finally, we show that the elasticity of demand is increasing in price. Thus, the conditions in Assumption 1 ensure that the demand system satisfies this property globally and rule out the alternative case where the elasticity is decreasing in price (or increasing in quantity), as allowed by Zhelobodko *et al* (2012), Kichko *et al* (2013) and Dhingra and Morrow (2016).

Welfare Gains

Having confirmed that the expenditure function is well behaved for $r \leq 2$, we next analyze welfare. Assume for convenience that labor is the only factor of production and each consumer has one unit, so that income equals the wage, *w*. Then $w = ue_r(\mathbf{p})$, so a drop in the expenditure function will indicate welfare gains. The appearance of a new good means a drop in its price from the reservation price, so that welfare automatically rises as there is increased variety for the consumer. Our goal here is to develop more general sufficient conditions for welfare to rise from one equilibrium to another. To achieve this, we need a characterization of the term *D* in (5). Feenstra and Weinstein (2017) show that in the translog case, a reduced-form expression for the unit-expenditure function – or the cost of living – includes the Herfindahl index evaluated over the product shares: as this Herfindahl index *falls,* indicating that there is less concentration in the product shares, the cost-of-living *rises.*[9] That counter-intuitive result is interpreted by Feenstra and Weinstein (2017) as reflecting “crowding” in product space: each new product variety increases welfare, but by a reduced amount as more products are available.

[9] In Appendix A we obtain this result by showing that the limiting value of $\ln D(\mathbf{p})^{1/r}$ as $r \to 0$ equals the Herfindahl index times a negative coefficient.

For other values of *r*, we can still obtain a type of Herfindahl index by defining the "adjusted" expenditure shares:

$$z_\omega(\mathbf{p}) \equiv \frac{s_\omega(\mathbf{p})(p^*/p_\omega)^{r/2}}{\int_\Omega s_{\omega'}(\mathbf{p})(p^*/p_{\omega'})^{r/2} d\omega'}. \tag{8}$$

For the translog case, $r = 0$, these adjusted shares equal the conventional shares, while for the quadratic case, $r = 2$, these adjusted shares equal the *quantity* share of each product. Then defining the Herfindahl index, $H \equiv \int_\Omega z_\omega(\mathbf{p})^2 d\omega$, it is shown in Appendix B that,

$$D(\mathbf{p})^{1/r} = \underbrace{\left[-\alpha\left(\tilde{N} + \frac{\alpha}{\beta}\right)\right]^{1/r}}_{\text{constant}>0} \underbrace{\left[1 - \left(\tilde{N} + \frac{\alpha}{\beta}\right)H\right]^{1/r}}_{\downarrow \text{ in } H}, \tag{9}$$

where the final term is decreasing in the Herfindahl index *H* (since $[\tilde{N} + (\alpha/\beta)]$ has the same sign as *r*, from Assumption 1).

Recall that expenditure is $e_r(\mathbf{p}) = p^* \times D(\mathbf{p})^{1/r}$, from (5). As the reservation price falls, so does expenditure and welfare rises. But that gain is offset if the Herfindahl index also falls. In the trade environments we shall consider, that result will be likely whenever variety increases: while there is not a one-to-one correspondence between changes in the mass of products *N* and the Herfindahl, in all cases that we examine an increase in *N* implies a lower Herfindahl, which in turn implies an increase in $D(\mathbf{p})^{1/r}$. It follows that if expenditure falls, then it falls by *less* than the reduction in the reservation price. The question is whether we achieve some bound to this offsetting effect on welfare due to crowding in product space. That question is answered in the affirmative, as shown by the following decomposition of expenditure:

Lemma 1

Under Assumption 1, the cost of living can be decomposed alternatively as:

$$
\begin{aligned}
e_r(\mathbf{p}) &= p^* \underbrace{\left[-\alpha\left(\tilde{N}+\frac{\alpha}{\beta}\right)\right]^{1/r}}_{\text{constant}>0} \underbrace{\left[1-\left(\tilde{N}+\frac{\alpha}{\beta}\right)H\right]^{1/r}}_{\downarrow \text{ in } H} \\
&= \sqrt{p^*}\underbrace{\left[-\alpha\left(\tilde{N}+\frac{\alpha}{\beta}\right)\right]^{1/r}}_{\text{constant}>0} \underbrace{\left[\int_{\Omega} s_\omega(\mathbf{p})\sqrt{p_\omega}^{\,-r} d\omega\right]^{-1/r}}_{\uparrow \text{ in prices } p_\omega \text{ and in their spread}} .
\end{aligned}
\tag{10}
$$

Sufficient conditions for a fall in the cost of living and rise in welfare are that: (i) the reservation price falls; and (ii) the Herfindahl index does not fall or the share-weighted power mean of prices $\sqrt{p_\omega}$ on the second line does not rise.

The first line of (10), obtained by substituting (9) into (5), has already been discussed. It shows that if the reservation price is falling, but the Herfindahl index is also falling due to an increase in variety, then the decline in the cost of living (and increase in welfare) will be *less than* the fall in p^*. Feenstra and Weinstein (2017) refer to this outcome as a “crowding” effect. Now suppose that the Herfindahl falls when comparing two equilibria, which will tend to increase the cost of living, but that the reservation price also falls. Can we easily determine whether welfare rises or falls? The second line of (10) gives an affirmative answer. Regardless of the change in the Herfindahl, the cost of living falls and welfare rises if, along with the fall in the reservation price, the share-weighted power mean of the prices $\sqrt{p_\omega}$ shown on the second line does not rise. If this power mean falls, then it follows that the decline in the cost of living and increase in welfare *exceeds* the fall in $\sqrt{p^*}$. So under these conditions, the changes in p^* and $\sqrt{p^*}$ effectively become bounds for the change in the cost of living.

In the next section we shall further decompose the reservation price p^* into terms reflecting (i) product variety, (ii) the markups charged by firms, and (iii) an average of firm costs, which reflect selection across firms. Using this decomposition in (10), we will obtain a decomposition of the cost of living into the three potential sources of gains from trade, together with an *additional* term: either the Herfindahl index or the share-weighted power mean of prices. To understand why these additional terms enter (10), consider the special case where all firms have the same costs and markups, so that in autarky all varieties sell for the same price $\bar{p}_\omega$. Then substituting for the reservation price from (2), the cost of living becomes:

$$e_r(\mathbf{p}) = \bar{p}_\omega \underbrace{\left(\frac{-\alpha N[\tilde{N} + (\alpha / \beta)]}{N - [\tilde{N} + (\alpha / \beta)]} \right)^{1/r}}_{\downarrow \text{ in variety } N} . \tag{11}$$

The final term in (11) is a precise expression for how product variety impacts the cost of living when all firms are identical.[10] Up to a constant, this expression is the *square root* of the variety term that appears in the reservation price p^*, i.e. the square root of the first term on the right of (2). So this explains why the cost of living on the second line of (10) uses the *square root* of the reservation price: that adjustment is needed to accurately measure the gains from variety. It also explains why the first line of (10), which has the reservation price *as a level*, must also include the Herfindahl index: that index is needed to offset the impact of variety on the reservation price, where it is greater than (i.e. the square of) its impact on welfare.

Also appearing in the cost of living in (11) is the common price $\bar{p}_\omega$, which would reflect the common costs and common markup of the homogeneous firms. With heterogeneous firms, however, the *entire distribution* of costs, markups and therefore prices influence welfare. Indeed,

[10] Actually, the final term in (11) is a precise expression for how product variety impacts the cost of living even with heterogeneous firms, as shown by (B6) in the proof of Lemma 1 in Appendix B.

on the second line of (10) we indicate that the share-weighted power mean of the prices is increasing in their spread. This result can be seen as follows. Suppose that we start with the common price $\overline{p}_{\omega}$, and then consider a spread of prices (possibly due to unequal markups, for example) that leaves the total quantity purchased over all varieties unchanged, with unchanged expenditure and reservation price.[11] Then we show in the proof of Lemma 1 (Appendix B) that for $r < 2$, the share-weighted mean of prices in (10) *rises*. It follows that the spread in prices *raises* the cost of living and also lowers the Herfindahl index on the first line of (10). So another reason for the Herfindahl index and share-weighted power mean of prices to enter (10) is that they are summary statistics for the entire distribution of expenditure shares and prices, with an increasing spread of prices (due to unequal markups, for example) leading to lower welfare.

3. Autarky Equilibrium

We have already assumed that labor is the only factor of production, and we will focus on symmetric equilibria across countries so we normalize the wages at unity. As in Melitz (2003), we assume that firms receive a random draw of productivity denoted by φ, so marginal costs are a/φ, where a is the labor need per unit of output for a firm with the lowest productivity of $\varphi = 1$. We will allow the Pareto distribution of productivity to have either an upper-bound in its support, as in Helpman, Melitz and Rubenstein (2008), or to be unbounded above:

Assumption 2

(a) The productivity distribution in every country is Pareto, $G(\varphi) = (1-\varphi^{-\theta})/(1-b^{-\theta})$, $1 \leq \varphi \leq b$, where the upper bound is $b \in (1, +\infty]$ and $\theta > \max\{0, -r\}$;

(b) There is a sunk cost F of obtaining a productivity draw, but no fixed cost of production.

[11] Under these assumptions, there is no change in profits when the spread in prices is due to unequal markups.

In part (a), we allow the Pareto distribution to be unbounded ($b = \infty$) or bounded ($1 < b < \infty$).[12] The restriction that $\theta > \max\{0, -r\}$ becomes $\theta > (\sigma - 1) > 0$ in the CES case, which is needed for certain first moments to converge in that case; this restriction is needed here for the same reason. The assumption that there is no fixed cost of production in (b) is made for convenience and follows Melitz and Ottaviano (2008).

The optimal price for a firm with productivity φ is $p = (a / \varphi)\eta / (\eta - 1)$, where now we drop the subscript ω for varieties and instead we (implicitly) index firms by their productivity φ. We let $\mu \equiv p / (a / \varphi)$ denote the ratio of price to marginal cost, while $v \equiv p^* / (a / \varphi)$ denotes the ratio of the reservation price to marginal cost. From (7), the elasticity $\eta(p / p^*)$ is a function of the price relative to the reservation price, so $\eta(\mu / v)$ and using this notation the markup is:

$$\mu = \frac{\eta(\mu / v)}{\eta(\mu / v) - 1} \Leftrightarrow \left[\left(\frac{\mu}{v}\right)^{r/2} - \frac{1}{2}\right]\left[r(\mu - 1) + 1\right] = \frac{1}{2},$$

where the second expression follows from (7) and is used to solve for $\mu(v)$.[13] Differentiating this expression, it is shown in the Appendix (Lemma B2) that the elasticity of the markup is $0 < v\mu'(v) / \mu < 1$, so that changes in marginal cost are only partially passed-through to prices.

We can now write the equilibrium conditions in autarky. A firm paying the sunk cost of F receives a draw of productivity φ with probability $g(\varphi) = G'(\varphi)$. We make a change of variables from φ to v. Since $v \equiv p^* / (a / \varphi)$ then $\varphi = av / p^*$, so using the Pareto distribution:

$$g(\varphi)d\varphi = \frac{\theta \varphi^{-\theta-1}}{1 - b^{-\theta}} d\varphi = \frac{\theta v^{-\theta-1}}{1 - b^{-\theta}} \left(\frac{p^*}{a}\right)^{\theta} dv = \left(\frac{p^*}{a}\right)^{\theta} g(v)dv \ . \tag{12}$$

---

[12] In section 5 we will allow b=1, so that firms are homogeneous in their productivities.

[13] The left side of the second expression can be evaluated at μ=1 and $\mu = v$ to show that it is above and below ½, so that a solution $\mu \in (1, v)$ where it equals ½ always exists. We benefit from an early version of ACDR (2017) in writing η as a function of μ/v and μ as a function of v.

This change of variables will considerably simplify our expressions. Because there are no fixed costs of production, the lowest-productivity firm that will continue production will have marginal costs equal to the reservation price, so $v = 1$ is the lower bound. The upper bound for v, denoted by v^*, is obtained when productivity is $b = av / p^*$, so that:

$$v^* = bp^*/a.$$

Starting with the expenditure share $f / D(\mathbf{p})$ from (6), we multiply that by expenditure L to obtain total demand, and then by $(\mu - 1)/\mu$ to obtain profits. Using the bounds $v \in [1, v^*]$, the expected profit from entering the market must equal the sunk costs of F in equilibrium, so that:

$$F = \frac{L\int_1^{v^*}\left[\frac{\mu(v)-1}{\mu(v)}\right] f\left(\frac{\mu(v)}{v}\right)\left(\frac{p^*}{a}\right)^{\theta} g(v)dv}{N_e\int_1^{v^*} f\left(\frac{\mu(v)}{v}\right)\left(\frac{p^*}{a}\right)^{\theta} g(v)dv}, \tag{13}$$

In the denominator, we substitute the expression for $D(\mathbf{p}) = \int_{\Omega} f(p_{\omega} / p^*)d\omega$. But rather than using the general notation Ω for the set of available products, with the change in variables in (12) we are now defining that set by the bounds for v and the mass of *entering* firms N_e. The mass of firms remaining after those with lowest productivity exit will be:

$$N = N_e\int_1^{v^*}\left(\frac{p^*}{a}\right)^{\theta} g(v)dv = N_e\left(\frac{p^*}{a}\right)^{\theta} G(v^*), \tag{14}$$

which also equals $N_e[1 - G(a/p^*)]$, where (a/p^*) is the productivity of the firm with marginal cost (a/φ) just equal to the reservation price. So as usual in the Melitz model, the mass of surviving firms N equals the mass of entering firms times the probability of survival, and this probability is equivalently written as $[1 - G(a/p^*)] = (p^*/a)^{\theta} G(v^*)$.

We have already used the condition that expenditure equals the workforce L, so that full-employment holds. The remaining equilibrium condition is obtained from the reservation price in (2), re-written slightly by dividing by p^* and using the Pareto distribution from (12):

$$N - [\tilde{N} + (\alpha / \beta)] = \left(N_e \int_1^{v^*} \left(\frac{\mu(v)}{v} \right)^{r/2} \left(\frac{p^*}{a} \right)^{\theta} g(v) dv \right). \tag{15}$$

The solution to the equilibrium conditions (13)-(15) is summarized in the following result:

<u>Proposition 1</u>

Under Assumptions 1 and 2: (a) the autarky equilibrium conditions (13)-(15) have a positive solution for p^*, N_e and N; (b) if and only if $b = \infty$, the solution for N_e is proportional to the country size L, while the solution for N is independent of country size L.

The existence result in (a) relies on $\theta > \max\{0, -r\}$ in Assumption 2, so that the integrals in (13) and (15) remain bounded even for $v^* \to \infty$. The results in part (b), when productivity is unbounded, are obtained by inspection of the equilibrium conditions. In that case the upper-limit of integration in (13) is infinite so the two integrals are constant and (13) becomes $F = \kappa_1 L / N_e$. It is immediate that the mass of entrants is proportional to country size in this case. This result is also obtained by ACDR, and follows from the "proportionality relation" between expected profits and expected revenue. While these two variables are proportional for every firm in the CES case (i.e. regardless of productivity), they are proportional in *expected terms* for the demand systems that we or ACDR adopt, provided that the Pareto distribution is unbounded above and there are no fixed costs of production. Those two assumptions ensure that the upper $(v=\infty)$ and lower $(v=1)$ limits of integration in (13) are exogenous.

Part (b) gives a second implication of unbounded productivity that is less well known than the linear relationship between entry and country size, and concerns the mass N of surviving firms. Substituting (14) into (15), the equilibrium condition for the reservation price becomes:

$$-\alpha\left(\tilde{N}+\frac{\alpha}{\beta}\right)=N\left\{\int_{1}^{v^*}\alpha\left[\left(\frac{\mu(v)}{v}\right)^{r/2}-1\right]\frac{g(v)}{G(v^*)}dv\right\}, \tag{16}$$

which equals a positive constant on the left, from Assumption 1. For $b=\infty$ and $v^*=bp^*/a=\infty$, it follows that (16) solves uniquely for N, independent of p^* and country size L. This surprising result is also found by Arkolakis, Costinot and Rodriguez-Clare (2010) for the translog case, and ACDR for their more general demand function. A related result has been shown by Baldwin and Forslid (2010), who consider the Helpman, Melitz and Rubenstein (2008) model with CES preferences and a bounded Pareto distribution for firm costs. They show that a reduction in variable trade costs has *no effect* on the variety of products consumed when the fixed costs of exporting and domestic production are equal: new import varieties reduce domestic varieties one-for-one in that case. In contrast, we have excluded the CES case in Assumption 1 and also exclude fixed costs, and find that product variety is independent of country size if and only if the Pareto distribution is unbounded, $b=\infty$. This finding will have strong implications for the sources of gains from trade, examined in the following section.

Before turning to that discussion, we use the firm-level structure introduced in this section to further decompose the cost of living. The reservation price in autarky can be written as the product of terms that reflect the average of firm markups and costs, as follows:

Lemma 2

The reservation price in the closed economy is:

$$p^* = \underbrace{\left(\frac{N}{N-[\tilde{N}+(\alpha/\beta)]}\right)^{2/r}}_{\downarrow \text{ in variety } N} \underbrace{\left[\int_1^{v^*} \mu(v)^{r/2} \frac{\tilde{g}(v)}{\tilde{G}(v^*)} dv\right]^{2/r}}_{\text{Average markup}} \underbrace{\left[\int_1^{v^*} \left(\frac{p^*}{v}\right)^{r/2} \frac{g(v)}{G(v^*)} dv\right]^{2/r}}_{\text{Average of costs}}, \qquad (17)$$

where $\tilde{g}(v) \equiv \left(1+\frac{r}{2\theta}\right) g(v) v^{-r/2}$ is a Pareto density with distribution $\tilde{G}(v^*) \equiv \int_1^{v^*} \tilde{g}(v) dv$.

The first term appearing on the right of (17) is the same variety term appearing in (2). The second term is a power mean of the markups $\mu(v)$, using the "adjusted" density $\tilde{g}(v)$, which is positive from Assumption 2. To interpret the last term, recall that $v \equiv p^*/(a/\varphi)$ is the ratio of the reservation price to marginal cost, so $p^*/v = a/\varphi$ is the marginal cost of a firm with productivity φ. The last term in (17) is therefore a power mean of the marginal costs of firms. If at least one of the three terms in Lemma 2 falls, and the others along with the share-weighted price term in Lemma 1 does not rise, then we are assured of a welfare gain. We now examine trade environments allowing for such welfare gains.

4. Frictionless Trade

We initially consider frictionless trade, where in addition to the assumption of no fixed costs of production or export, we also ignore variable costs of trade (while introducing such trade costs in the next section). We suppose that the expenditure function in (1) along with Assumptions 1 and 2 holds across countries. In this environment, moving from autarky to frictionless trade is equivalent to growth in the labor force *L*. We have already shown in Proposition 1 that with an unbounded Pareto distribution, product variety *N* does not change but N_e rises in proportion to *L*. It follows that the probability of survival is falling, so there is a *positive selection effect:* only firms with productivity above a higher cutoff level produce in the larger market, while smaller firms are crowded out. Furthermore, this selection effect is the *only* source of welfare gain in the larger market: variety N is independent of L and it will follow that the Herfindahl index in (10) does not change; the average markup in (17) does not change because the upper-limit of integration is $v^* \equiv p^*b/a \to \infty$ as $b \to \infty$; and it follows from Lemmas 1 and 2 that only the fall in firms' costs changes the reservation price and welfare.

When productivity is bounded, however, then we shall find that all three sources of gains from country growth operate: variety increases, the average markup falls, and there is a positive selection effect. To show this, we perform the comparative statics on (13)-(15). Differentiating (14)-(15) and simplifying, we obtain:

$$d\ln N_e = -\theta(1+A)d\ln p^*, \qquad A \equiv \frac{N_e}{[\tilde{N}+(\alpha/\beta)]}\left(\frac{b^{-\theta}}{1-b^{-\theta}}\right)\left[1-\left(\frac{\mu(v^*)}{v^*}\right)^{r/2}\right],$$

where $v^* = p^*b/a$. Re-express (13) by moving the denominator $D(\mathbf{p})$ to the left, obtaining:

$$0 = \int_1^{v^*}\left\{\left[\frac{\mu(v)-1}{\mu(v)}\right]L - FN_e\right\} f\left(\frac{\mu(v)}{v}\right)\left(\frac{p^*}{a}\right)^{\theta} g(v)dv.$$

Totally differentiating this condition, and substituting for $d\ln N_e$ from above, we obtain:

$$d\ln N_e = \left(\frac{1+A}{1+A+B}\right)d\ln L \qquad \text{and} \qquad d\ln p^* = \frac{-d\ln L}{\theta(1+A+B)}, \tag{18}$$

where $B \equiv \left[\frac{L}{F}\left(\frac{\mu(v^*)-1}{\mu(v^*)}\right) - N_e\right]\frac{f[\mu(v^*)/v^*]b^{-\theta}}{D(\mathbf{p})(1-b^{-\theta})}$.

To give the intuition for these results, consider the free entry condition (13). The markups appearing in the numerator of this expression are increasing as the reservation price rises or marginal cost falls, $\mu'(v) > 0$, and likewise for the Lerner index $[\mu(v) - 1] / \mu(v)$. The rising markup follows from the fact that the demand elasticity is increasing in price, as noted earlier. So as the reservation price *falls*, the expected markup in the numerator of (13) falls relative to the integral in the denominator. It follows that N_e rises *less than proportionately* with L. That is shown in the first result in (18), where $A > 0$ for $b < \infty$ from Assumption 1, while $B > 0$ for $b < \infty$ because the Lerner index takes on its highest value at the upper bound v^*, so that:

$$\frac{L}{F}\left(\frac{\mu(v^*)-1}{\mu(v^*)}\right) > \frac{L}{F}\int_1^{v^*}\left(\frac{\mu(v)-1}{\mu(v)}\right)\left[\frac{f\left(\frac{\mu(v)}{v}\right)g(v)}{\int_1^{v^*} f\left(\frac{\mu(v')}{v'}\right)g(v')dv'}\right]dv = N_e.$$

The inequality states that the highest Lerner index exceeds its average, and then the equality follows directly from the free entry condition (13) and ensures that $B > 0$ for $b < \infty$.

When productivity is unbounded and $b \to \infty$, then A, $B \to 0$, and so in that case we have $d\ln N_e = d\ln L$, as asserted in Proposition 1(b). For unbounded productivity we also have that the reservation price changes by $d\ln p^* = -d\ln L / \theta$ in (18) and, as discussed above, this change is purely due to the drop in the average of firm costs, i.e. the self-selection of more efficient firms in the larger market.

When productivity is bounded, however, then the reservation price falls by less than $d\ln L / \theta$, as shown by (18) with A, $B > 0$. That means that the increased selection of firms is offset. We conclude that country growth has two opposing effects on product variety N: entry of firms N_e rises less than proportionately with L; but also the reservation price falls by less, so the increased selection is offset. It turns out that this second effect dominates so that product variety

N rises with *L*. This result can be seen from the equilibrium condition for the reservation price in (16), where the left side is constant while the integral on the right is a weighted average of terms that are increasing in *v*, so the integral is too (as shown in Appendix B). It follows with $b < \infty$, a fall in $v^* = bp^*/a$ is associated with a rise in *N*. These various results are summarized as:

Proposition 2

Under Assumptions 1 and 2, an increase in country size *L* under frictionless trade leads to: (a) when $b = \infty$, then p^* falls only due to the drop in the average of firm costs, with the Herfindahl index *H* fixed; (b) when $b < \infty$, then variety *N* rises, the Herfindahl falls, and the average of firm costs, markups and the weighted-average price term in (10) all fall; (c) the proportional welfare gain when $b < \infty$ is *less than* that with $b = \infty$.

Part (a), with an unbounded Pareto distribution, has already been discussed above and the constant Herfindahl is shown in Appendix B. The rise in product variety with bounded Pareto, in part (b), has also been motivated by the comparative statics above and the falling Herfindahl is also shown in Appendix B. The fact that the average of firm costs and markups both fall follows from those terms in (17): as the reservation price falls then so does $v^* \equiv p^*b/a$, and so we are excluding the highest markup term $\mu(v^*)$ in (17); but because p^* appears explicitly within the integral of costs, we are also reducing the average of firm costs. The share-weighted power mean of prices appearing in (10) also falls, and as discussed after Lemma 1, that can reflect the reduced spread of markups and prices. For these various reasons, the consumer gains.

Part (c) shows that despite the fact that all three sources of gains from trade operate in the bounded Pareto case, the total proportional gains from trade are *smaller* with bounded than with unbounded Pareto. This result follows from our welfare decomposition in Lemma 1 and the comparative statics above. With $b = \infty$, we found that $d\ln p^* = -d\ln L/\theta$ and the Herfindahl index is fixed, so it follows immediately from the first line of (10) that the increase in welfare is $d\ln L/\theta$. But with $b < \infty$, we found above that $d\ln p^* < -d\ln L/\theta$, and we also confirm in Appendix B that the Herfindahl index is falling as variety increases. For both reasons, if follows that the frictionless trade leads to an increase in welfare that is *less than* $d\ln L/\theta$, obtained in the unbounded case.

The result in part (c) is related to that in Melitz and Redding (2015), who focus on the CES case only. They show that, provided there are fixed costs of exporting, then the gains from trade with heterogeneous firms exceed that with homogeneous firms as in the Krugman (1980) model. Homogeneous firms are an extreme case of bounded Pareto where there is a mass point at a single productivity. So Melitz and Redding (2015) are comparing any productivity distribution for heterogeneous firms (including bounded or unbounded Pareto) with a degenerate distribution with a single mass point. In comparison, we find that even without fixed costs of trade, the gains from frictionless trade with unbounded Pareto exceed those with bounded Pareto for the QMOR class of preferences. So we are comparing the unbounded Pareto case to any bounded Pareto (but not including the degenerate case $b = 1$, ruled out in Assumption 2).

With this difference in our comparisons understood, the spirit of our results is similar: having a greater spread of productivities leads to higher proportional gains from trade. That is an especially surprising result in our context because by restricting the range of productivities we give scope for additional sources of gains from trade – due to product variety and reduced markups – that do not operate with the unbounded Pareto. We have found that imposing an upper bound on productivity allows these additional sources of gains to operate, but necessarily reduces the self-selection of more efficient firms so much that the total gains from trade are lower.

Our results can be compared to the formula for welfare gains found by Arkolakis, Costinot and Rodriguez-Clare (ACR, 2012) and in the homothetic case by ACDR, which emphasizes the share of total consumption purchased from the domestic market. Denoting that fraction by λ, the domestic labor force by $\bar{L}$, and the world labor force by $L > \bar{L}$, then $\lambda = 1$ in autarky and $\lambda = \bar{L} / L$ with frictionless trade. With growth it follows that $d \ln \lambda = -d \ln L < 0$. Applying our results above that the welfare gain is $-d \ln p^* = d \ln L / \theta$ with unbounded Pareto, but *smaller* with bounded Pareto, we have proved:

Corollary 1

Under Assumptions 1 and 2, the gain from frictionless trade equals $-d \ln \lambda / \theta > 0$ with an unbounded Pareto distribution, but is strictly less than this amount with a bounded Pareto distribution for productivity.

Notice that this result allows for differences in country sizes, i.e. a country of size $\bar{L}$ can open trade with another country of size $L - \bar{L} > 0$. Assumption 2 and Corollary 1 still maintain symmetry of production and fixed costs across countries, however.

5. Variable Trade Costs

We now allow for variable costs of trade, and we suppose that the trading countries are symmetric in all respects except in their proximity to each other. We shall let $\tilde{C} \geq 2$ denote the number of (identical) countries in the world, but due to trade costs, each country does not necessarily trade with all others. We number countries by their proximity to an exporter, so $c = 1$ denotes the local market, $c = 2$ denotes the next closest market, etc. In equilibrium we allow for trade with whole countries or a fraction of a country (as explained below). We shall assume the following structure of trade costs:

<u>Assumption 3</u>

Numbering countries by their proximity to an exporter, delivering one unit to country c means that $\tau(c) = \tau_0 c^\rho \geq 1$ units must be sent, with $\tau_0 \geq 1$, $\rho \geq 0$ and $1 < c \leq \widetilde{C}$.

These costs apply onto to cross-border trade, while local sales (c = 1) have $\tau \equiv 1$. Notice that number of countries c that a nation is trading with plays the same role in Assumption 3 as "distance" does in an empirical specification of variable transport costs, while τ_0 plays the same role as a "border cost," i.e. the extra amount that must be sent regardless of distance. In our working paper (Feenstra, 2014), we provide a microstructure that justifies the trade costs described in Assumption 3.

With Assumption 3, we can readily solve for the number C of countries that each nation actually trades with in the symmetric equilibrium. The most efficient firm in any country has marginal labor costs of a/b to produce one unit of output. Normalizing the wage at unity in every country, in equilibrium the marginal cost of producing enough to deliver one unit to the most distant country C will just equal the reservation price in that country:

$$\tau_0 C^\rho \left(\frac{a}{b}\right) = p^*, \text{ for } 1 < C < \widetilde{C}. \tag{19}$$

This equilibrium condition provides a very simple relation between the border cost τ_0 and the equilibrium number of trading partners. Of course, changes in the trade costs τ_0 will also affect the reservation price in (19), so we will need to specify all the equilibrium conditions to account for the endogenous response of both C and p^*. Note that if the trade cost τ_0 is sufficiently close to unity and ρ is close enough to zero, or if the Pareto distribution is unbounded with $b \to \infty$, then $\tau_0 \widetilde{C}^\rho (a/b) < p^*$ and so the most efficient firm from each country sells to *every* market in which case $C = \widetilde{C}$.

To write the other equilibrium conditions with trade, we revisit the change in variables introduced for the autarky economy. When a firm is selling to a foreign country, we let $v \equiv p^*/(\tau a/\varphi)$ denote the ratio of the reservation price to the marginal costs *inclusive* of the variable trade costs. It follows that $\varphi = \tau a v / p^*$, so that from (12):

$$g(\varphi)d\varphi = \frac{\theta\varphi^{-\theta-1}}{1-b^{-\theta}}d\varphi = \frac{\theta v^{-\theta-1}}{1-b^{-\theta}}\left(\frac{p^*}{a\tau}\right)^{\theta} dv = \left(\frac{p^*}{a\tau}\right)^{\theta} g(v)dv \ . \tag{20}$$

From the final expression in (20), we see that higher trade costs τ implies a lower density of firms in any interval *dv*, which shows how the trade costs affect the *extensive margin* of exporting firms. But in contrast to the unbounded Pareto case, trade costs now also affect the *intensive margin* of exporters, and of the highest-productivity exporter in particular. The upper bound for *v* when selling to the domestic market is still denoted by $v^* = bp^*/a,$ and the upper bound when selling to a foreign country *c* is:

$$v^*/\tau(c) = bp^*/a\tau(c). \tag{21}$$

With unbounded productivity, $b \to \infty$, the ratio of reservation price to marginal costs for foreign firms – inclusive of the variable trade costs – is in the range $[1,+\infty)$, the same as for home firms. So there is no difference in the distribution of marginal costs and prices charged by home and foreign firms: both countries have firms with essentially zero costs, charging an infinite markup, and firms with marginal costs equal to the reservation price, with zero markup. But with bounded productivity, we see from (21) that the ratio of the reservation price to marginal cost is in the range $v/\tau \in [1,\ bp^*/a\tau\ )$, which depends on the reservation price and trade costs. Now the price of the highest productivity firm is affected by trade costs, and we refer to this as an impact on the *intensive margin* of the highest productivity firm.

We continue to let N denote the total mass of products available to the representative consumer in each country, so this notation from section 2 stands. But in section 3, dealing with the autarky economy, we previously let N_e denote the mass of entering firms, while N was the mass of surviving firms. With trade we need to introduce a new notation for the mass of firms in a single country, so we now let M_e denote the mass of entering firms in a single country, and M denote the mass of surviving firms. These are related to the equilibrium condition (14), re-written using this new notation as,

$$M = M_e \int_1^{v^*} \left(\frac{p^*}{a} \right)^{\theta} g(v) dv = M_e \left(\frac{p^*}{a} \right)^{\theta} G(v^*). \tag{22}$$

Conditional on selling at home, the probability of firms in the interval dv selling to country c is then obtained by dividing (20) by the final terms in (22):

$$\frac{[p^*/a\tau(c)]^{\theta} g(v)}{(p^*/a)^{\theta} G(v^*)} dv = \frac{\tau(c)^{-\theta} g(v)}{G(v^*)} dv \,.$$

The total mass of products N available within a country is obtained by starting with the mass M produced in each country, and then integrating over the conditional density above:

$$\begin{aligned} N &= M \left\{ 1 + \int_1^C \int_1^{v^*/\tau(c)} \frac{\tau(c)^{-\theta} g(v)}{G(v^*)} dv dc \right\} \\ &= M \left\{ 1 + \tau_0^{-\theta} \left[\frac{C^{(1-\rho\theta)}}{1-(v^*)^{-\theta}} \right] - \frac{(C-1)(v^*)^{-\theta}}{1-(v^*)^{-\theta}} \right\}, \end{aligned} \tag{23}$$

where $C^{(1-\rho\theta)} \equiv (C^{1-\rho\theta} - 1)/(1-\rho\theta)$ is the Box-Cox transformation of C.[14] We see from (23) that a reduction in trade costs has a positive impact on the mass of varieties available in a country

[14] The result in (23) is obtained by first integrating over v, obtaining $\tau(c)^{-\theta} G[v^*/\tau(c)]/G(v^*)$; then using the Pareto distribution and trade costs in Assumptions 2 and 3; and then integrating over trading partners c.

through $\tau_0^{-\theta}$, and also an indirect effect through the reservation price. In addition, changes in trade costs can impact N through two further effects: by changing the range of countries C that are exporting to each destination, and by changing the mass of domestic products M.

The equilibrium conditions so far are (19), (22) and (23). In Appendix B, we further develop the free entry condition in the open economy, analogous to (13) in the closed economy, and the expression to solve for the reservation price in the open economy, analogous to (15). We also develop a decomposition of the reservation price into three terms reflecting product variety, the average markup of domestic firms and foreign firms exporting to home, and the average of their costs. This decomposition (Lemma B3) is analogous to Lemma 2 but applies to the open economy, and in conjunction with Lemma 1, gives us a decomposition of welfare.

We first consider the case of unbounded Pareto, so that $b \to \infty$ and $v^* \to \infty$.

Proposition 3

Under Assumptions 1 – 3, a reduction in trade costs τ_0 in the unbounded Pareto case implies that p^* falls only due to the drop in the average of firm costs, reflecting a fall in the variety M produced in each country, with entry M_e, the mass of available products N, the average markup and the Herfindahl index all fixed.

This result is analogous to Proposition 2(a), and shows that with the unbounded Pareto there is no variety or pro-competitive effects; the reservation price falls only due to the reduction in the average of firm costs. The drop in the reservation price lowers the probability of being a successful firm, shown by the final terms on the right of (22), leading less efficient firms to exit. With M_e fixed, the mass of varieties M produced in each country declines as trade costs fall, which is what keeps the variety available to consumers N fixed in (23) even as the range of exported products expands. Note the range of trading partners obtained from (19) is at the corner solution $C = \widetilde{C}$ as $b \to \infty$, but each country exports a greater proportion of the varieties M that it produces as trade costs fall.

Next, we consider a reduction in trade costs in the bounded Pareto case, with $b < \infty$. It is difficult to perform the comparative statics on the equilibrium conditions in general, so we shall simplify the problem with the following assumption:

Assumption 4

a) Firms in all countries are homogeneous with marginal labor costs of a ; b) $\rho > 0$ or $\widetilde{C}$ are large enough so that there is an interior solution to the range of trading partners, $1 < C < \widetilde{C}$.

Under part a), there is no difference between entering and surviving firms, so that $M_e = M$. From part b), a small reduction in the border costs $d\tau_0$ in the neighborhood of $\tau_0 = 1$ will lead to an expansion in the range of trading partners C determined by (19).[15] Then for such a small reduction in trade costs τ_0, symmetric across countries, we obtain the following results:

Proposition 4

Under Assumptions 1, 3 and 4, a slight reduction in the border costs $d\tau_0 < 0$ in a neighborhood of $\tau_0 = 1$ leads to: an increase in the consumer variety N; reductions in the Herfindahl index, in the mass of varieties $M_e = M$ produced in each country, and in the average markup and costs over domestic and foreign firms selling to each destination.

---

[15] In our working paper (Feenstra, 2014), we performed the comparative statics of the equilibrium conditions with heterogeneous firms and bounded Pareto, but evaluated at the frictionless equilibrium with $\tau_0 = 1$ and $\rho = 0$, which meant that $C = \widetilde{C}$ even with a slight change in τ_0. In that case we found that a slight reduction in border costs τ_0 led to no change in H or M_e, and a fall in p^*, so the only source of gains from trade were through selection. In contrast, we now consider the case where $\rho > 0$ and this distance effect is large enough so that there is an internal solution to the range of exporting countries C, while simplifying the problem with homogeneous firms.

This result is analogous to Proposition 2(b), and shows that with the bounded Pareto a slight reduction in the border costs leads to a positive impact on consumer variety through N, and a pro-competitive effect through reducing the average markup. The fall in the average markup is notable because there is incomplete pass-through from marginal costs to prices, so that reduced trade costs are *less-than* fully passed through to import prices and the markups earned by foreign exporters *rise*.[16] The markup earned by domestic firms falls, however, and Proposition 4 shows the *average* markup that is relevant for welfare also falls. [17] So the variety and pro-competitive sources of the gains from trade both operate under bounded Pareto. In addition, even though the marginal *production* costs of firms are the same, the marginal costs inclusive of *transport* costs differ. As border costs are reduced then these costs fall for existing exporters, but in addition, exports occur from more distant countries, and these have higher costs inclusive of transport. Proposition 4 shows that there is a positive selection effect in that the average costs indeed fall.

In contrast, with the unbounded case in Proposition 3, *only* the positive selection effect is operating with heterogeneous firms. What remains to be investigated is the magnitude of the *total* gains from trade in the bounded versus the unbounded cases. For the case of frictionless trade and a growth in country size in the last section, we found that the gains were higher in the unbounded case (Corollary 1). Does this result continue to hold when considering reductions in the costs of trade?

We can answer that question in the affirmative, and for a somewhat more general setting than considered in Proposition 4. Let us return to the case of heterogeneous firms. With bounded Pareto a *very large* reduction in variable trade costs for *any* country, moving it from autarky to

[16] This important point is stressed by ACDR (2017).

[17] In Lemma B4 in the Appendix we develop a decomposition of the reservation price into terms reflecting product variety, the average markup of domestic and foreign exporting firms, and the average of their costs. In conjunction with Lemma 1, this gives us a decomposition of welfare into these three terms plus the Herfindahl index.

frictionless trade, must act like the increase in country size L as analyzed in section 4: from Proposition 2, variety N available to consumers rises and the Herfindahl index and reservation price both fall. Because entry N_e rises less than proportionately with L, then entry in *each country*, $M_e \propto N_e/L$, must fall. The same result holds if there is growth in the labor force of *any* country under frictionless trade, which will reduce entry in *all* countries.[18] Proposition 4 shows us that these results hold also for a slight reduction in border costs around $\tau_0 = 1$.[19] On the other hand, with unbounded Pareto we know from Proposition 3 that only the reservation price falls, while entry M_e, the mass of available product N and the Herfindahl index are all fixed.

With these results in hand, let us now generalize and consider *any reduction in home or foreign trade costs or increase in country sizes* that are consistent with the results of Propositions 2-4: $dH \le 0$, $dM_e \le 0$ and $dp^* < 0$. Under these conditions, we will show that the rise in home welfare due to the reduction in trade costs is *bounded above* by the ACR/ACDR formula, thereby extending Corollary 1 to allow for trade costs. To establish this result, we start with the share of expenditure coming from domestic production, or λ. Integrating over the shares in (6) gives:

$$\lambda = \frac{M_e \int_1^{v^*} f\left(\frac{\mu(v)}{v}\right)\left(\frac{p^*}{a}\right)^{\theta} g(v)dv}{D(\mathbf{p})}. \tag{24}$$

We do not necessarily assume symmetry in country sizes, but simply let the denominator of (24), $D(\mathbf{p}) \equiv \int_{\Omega} f(p_\omega / p^*)d\omega$, be evaluated over all domestic and imported products into the home

[18] The world labor force can be divided as $L = L_1+\ldots+L_C$, with entry in each country of M_{ei} with $N_e = M_{e1}+\ldots+M_{eC}$. With symmetry of costs across countries then $N_e/L = M_{ei}/L_i$, which falls with growth in *any* country's labor force L_i, because any growth in L leads to less-than-proportionate growth in N_e from Proposition 2.

[19] Proposition 4 assumes homogeneous firms, but from Feenstra (2014) summarized in note 14, we know that with heterogeneous firms a slight reduction in border costs around τ_0=1 leads to no change in H or M_e, and a fall in p^*.

country. In other words, for the term $D(\mathbf{p})$ we are returning to the general notation of section 2, including the result in (9) that $[D(\mathbf{p})]^{1/r}$ is declining in the Herfindahl H evaluated over all domestic and imported products.

Recall that expenditure equals $e_r(\mathbf{p}) = p^* \times D(\mathbf{p})^{1/r}$, from (5). We use (24) to solve for p^* in terms of λ and $D(\mathbf{p})$, and substitute this solution into (5) to obtain:

$$e_r(\mathbf{p}) = a\lambda^{\frac{1}{\theta}} D(\mathbf{p})^{\frac{1}{\theta}+\frac{1}{r}} \left[\int_1^{v^*} M_e f\left(\tfrac{\mu(v)}{v}\right) g(v)dv \right]^{-\frac{1}{\theta}}. \qquad (25)$$

We consider a comparative statics change with $dH \le 0,\ dM_e \le 0,\ dp^* < 0$, so there is (weakly) a fall in the Herfindahl index due to greater imported variety, with (weakly) reduced domestic entry M_e and a reduction in the reservation price p^*. Then differentiating (25), we obtain:

$$d\ln e_r(\mathbf{p}) = \frac{d\ln\lambda}{\theta} + \underbrace{\frac{(\theta+r)}{\theta}}_{>0} \underbrace{\frac{d\ln D(\mathbf{p})}{r}}_{\ge 0} - \frac{1}{\theta} \underbrace{d\ln\left[\int_1^{v^*} M_e f\left(\tfrac{\mu(v)}{v}\right) g(v)dv \right]}_{<0} > \frac{d\ln\lambda}{\theta}. \qquad (26)$$

The term $(\theta + r)/\theta$ on the right of (26) is positive because of Assumption 2 that $\theta > -r$; the next term is non-negative because of the result in (9) that $D(\mathbf{p})^{1/r}$ is declining in H and our assumption that $dH \le 0;$ and the final term is negative because of our remaining assumptions that $dM_e \le 0$ and $dp^* < 0.$ With unbounded Pareto, all terms except the first on the right of (26) are zero from Propositions 2 and 3 with $v^* \to \infty$, and so welfare changes by $-d\ln\lambda/\theta.$ With bounded Pareto, however, the change in welfare is *less than* this amount because $dH \le 0$, $dM_e \le 0$ and $dp^* < 0$ by assumption, and v^* is finite. We therefore see that that the fall in the expenditure function is not as large as the fall in $d\ln\lambda/\theta,$ as stated formally by:

<u>Corollary 2</u>

Under Assumptions 1 – 3, the gain from any reduction in home or foreign trade costs or increase in country sizes equals $-d\ln\lambda/\theta > 0$ with an unbounded Pareto distribution, but is strictly less than this amount provided that the support of the Pareto distribution is bounded and that $dH \leq 0$, $dM_e \leq 0$, and $dp^* < 0$.

As we have found for a large reduction in trade costs from autarky to frictionless trade, or with growth in country size, we also expect that product variety N is increasing and that the average markup is declining for a small reduction in trade costs. Nevertheless, with a bounded Pareto distribution the selection effect (i.e. fall in average costs) is offset enough so that the total gain is also reduced, as shown by Corollary 2. We now turn to an empirical application of this result to the U.S. economy.

6. Application to the U.S. Economy

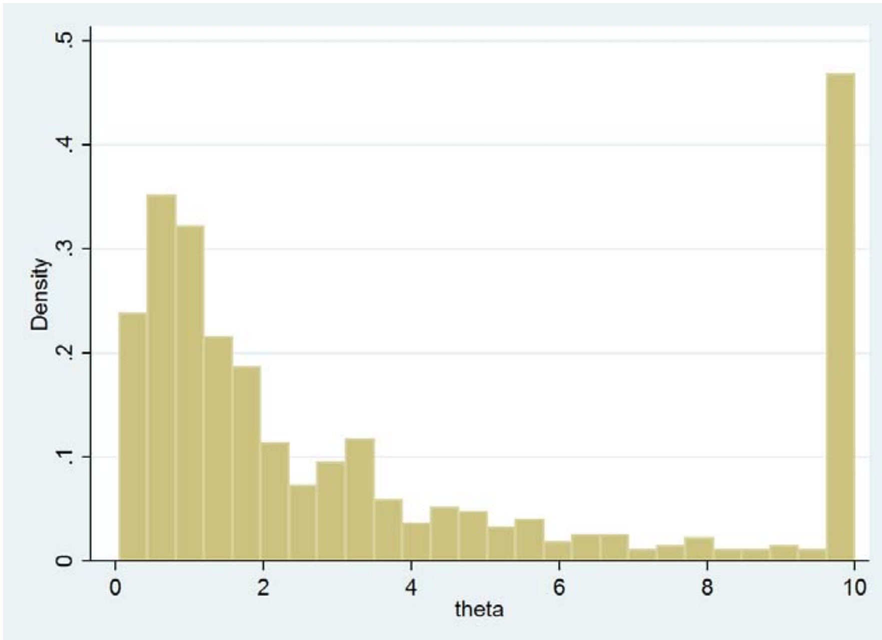
We consider the expansion in trade in the U.S. economy over 1992 to 2005, as studied by Feenstra and Weinstein (2017) for the translog case. The translog is a special case of the QMOR expenditure function with $r = 0$, so that Corollary 2 applies. Feenstra and Weinstein measure the pro-competitive effect, which we now denote by P, and the product variety gains, denoted by V, but they did not attempt to measure the gains due to firm selection and the reduction in average costs, now denoted by X. But Corollary 2 gives us a convenient short-cut to measure the selection gains, or at least an upper-bound to these gains. Specifically, we can simply use the formula $-\Delta\ln\lambda/\theta$ as an upper-bound to the total welfare gains, denoted by W, and then compute the selection gains as the residual $X = W - P - V$.

This exercise is an extension of the results in Feenstra and Weinstein (2017), who measured P and V at the 4-digit Harmonized System (HS) level, while concording the HS to the U.S. industry classification in order to measure import share relative to total U.S. consumption in each industry, or λ. The welfare gains are then aggregated across industries to achieve the total gains for merchandise (i.e. manufacturing, agriculture, and mining), and multiplying by the share 0.185, they obtain the gains relative to U.S. GDP. Feenstra and Weinstein perform this calculation separately over 1992-1997, and 1998-2005, because the U.S. classification was the Standard Industrial Classification up to 1997 and the North American Industry Classification system beginning in 1998. We likewise perform our calculation over these two sub-periods, and then sum the results to obtain the gains over 1992-2005.

The extra information that is needed here to perform this calculation is the Pareto parameter θ in each 4-digit HS industry. We estimate that parameter using firm-level export data from each country to the United States. This is the so-called Piers data used by Feenstra and Weinstein (2017), which covers sea shipments only. The Pareto productivity parameter θ is assumed to be constant across all source countries, so while it is estimated for sea shipments only, in calculating the formula $-\Delta \ln \lambda / \theta$ we are applying to imports by air and land, too. We estimate the productivity parameter using a maximum-likelihood technique applied to the firm-level exports. Generalizing Assumption 2 slightly, we allow for a bounded Pareto distribution with distinct upper-bounds on productivity for each exporting country. Then the parameter θ and the upper-bounds on the productivity for each exporter are chosen to maximize the likelihood function (see Appendix C and Tyazhelnikov, 2017, for a broader discussion).

The condition $\theta > \max\{0, -r\}$ in Assumption 2 becomes simply $\theta > 0$ for the translog function. Our maximum likelihood technique results in somewhat lower estimates of θ than found in existing literature, but satisfying $\theta > 0$ for 714 HS 4-digit industries. But if θ is very small, it will cause the upper bound to the welfare gains of $-\Delta \ln \lambda / \theta$ to blow up, i.e. that upper

Figure 1: Estimates of Theta (between 0.05 and 10)



Notes:
The estimates of the Pareto parameter θ are obtained as explained in Appendix C.

bound is no longer tight. Accordingly, we replace the estimates of θ below the lower tolerance of 0.05 (replacing 9 estimates of θ), or above an upper tolerance of 10 (replacing 127 estimates of θ), by those tolerances. The resulting histogram of θ estimates is shown in Figure 1 and has a median of 1.95 and mean of 3.61.

Using these estimates of the Pareto parameter, it is straightforward to construct the upper-bound to the welfare gains, $W = -\Delta \ln \lambda / \theta$, for each industry and compare that to the estimates of the variety gains V and pro-competitive gains P, while constructing $X = W - P - V$ as the residual gains due to the selection of firms. These estimates are then aggregated to all of merchandise and reported in Table 1 for the total period 1992-2005 as well as the two sub-periods, using several different lower tolerances for θ.

The results for 1992-2005 are shown in the first three rows of Table 1, and are the sum of that obtained over 1992-1997 and 1998-2005. In the first row we use the lower tolerance of $\theta \geq 0.05$ (which replaces 12 industry estimates with that value), increasing that tolerance to $\theta \geq 0.1$ (replacing 2 more estimates) and $\theta \geq 0.5$ (replacing 84 more estimates). The results are not very sensitive to that lower tolerance, and are also not sensitive to the upper tolerance of 10. In the

Table 1: Welfare Gains from Trade in the United States, 1992-2005

	Upper-bound to total welfare gain	Procom-petitive gain	Variety gain	Upper-bound to selection gain	Upper-bound to total welfare gain	Procom-petitive gain	Variety gain	Upper-bound to selection gain
Theta Bounds	***W***	***P***	***V***	***X***	***W***	***P***	***V***	***X***
	Percent of merchandise sector				*% GDP*	*Percent of W*		
				1992-2005				
[0.05,10]	6.90	2.20	2.78	1.91	1.28	32.0	40.3	27.8
[0.1, 10]	6.76	2.20	2.78	1.78	1.26	32.6	41.1	26.3
[0.5, 10]	6.37	2.20	2.78	1.38	1.18	34.6	43.6	21.7
				1992-1997				
[0.05,10]	2.08	1.10	3.85	-2.87	0.39	52.9	184.9	-137.8
[0.1, 10]	2.08	1.10	3.85	-2.86	0.39	52.8	184.7	-137.5
[0.5, 10]	2.00	1.10	3.85	-2.95	0.37	55.1	192.7	-147.7
				1998-2005				
[0.05,10]	4.82	1.10	-1.07	4.78	0.90	22.9	-22.2	99.2
[0.1, 10]	4.68	1.10	-1.07	4.64	0.87	23.6	-22.8	99.2
[0.5, 10]	4.37	1.10	-1.07	4.33	0.81	25.3	-24.4	99.2

Notes:
The total gain from trade W is obtained from the formula in Corollary 2 and is an upper bound. The pro-competitive gain P and the product variety gain V are obtained from Feenstra and Weinstein (2017), and then the selection gain X is computed as a residual, $X = W - P - V$, and is also an upper-bound The results for 1992-2005 in the first four rows equal the sum of the results for 1992-1997 and 1998-2005, while W in the fifth column multiplies the first column by 0.1858, reflecting the share of the merchandise sector (manufacturing, agriculture, and mining) in U.S. GDP. P, V, and X in the final three columns are obtained by dividing columns 2-4 by column 1 (times 100).

first row of results, we see that the welfare gain W equals 6.9% in the manufacturing sector, or 1.3% when measured relative to total GDP. Of that total, 72% of the gains come from the pro-competitive and product variety effects combined, with 28% coming from the residual selection effect (measured as an upper bound). When the lower tolerance for θ is increased to 0.1 and 0.5, the residual selection effect falls to 26% and 22%, respectively.

The results reported in Table 1 for the product variety and pro-competitive effects are very close to those in Feenstra and Weinstein (2017):[20] for the full 1992-2005 period, these two effects account for gains of 0.93%, or roughly one percent of total GDP, with both effects contributing substantially to this total. The results vary, however, in each of the sub-periods 1992-1997 and 1998-2005. As found in Feenstra and Weinstein and as reported in Table 1, the product variety gains are greatest in the 1992-1997 period but are *negative* in the 1998-2005 period. The variety gains are so large during 1992-1997 period that $P+V$ *exceeds* the upper-bound to the total gains, $W = -\Delta \ln \lambda / \theta$, resulting in a negative estimate for the residual selection gains, $X = W - P - V$. Conversely, the negative variety gains during 1998-2005 leads to an estimate of X that nearly equals the total gains W. How are we to interpret these unusual results obtained in the two sub-periods?

Our explanation, which goes beyond the theory presented here, is that domestic firms are *slow to exit the market* in response to import competition. With domestic varieties still available to consumers even as import varieties grow, the product variety gains are correspondingly increased, as we find over 1992-1997. But the continued presence of the domestic firms means that λ, the share of consumption coming from home firms in any industry, does not fall by as much as it would otherwise. Accordingly, our upper-bound estimate of the total gains from trade,

[20] This estimate differs slightly from gains of 0.86% of GDP reported in Feenstra and Weinstein (2017) because here we do not attempt to estimate θ in a small number of industries where there are less than 6 exporting firms in every source country. The results for merchandise are obtained by aggregating over the remaining industries.

$W = -\Delta \ln \lambda / \theta$, is not as large as it would be otherwise. This is seen in Table 1 for 1992-1997, where the gains *V* from product variety *exceed* the total gains *W*, leading to the negative residual estimate of the selection gains *X*.

As domestic firms exited during 1998-2005, more than commensurate with the rise in import varieties, we obtain a negative estimate of the variety gains *V* and a correspondingly large estimate of the residual selection gains *X*. We believe that these estimates reflect the delayed exit of domestic firms, leading to a much larger estimate of $W = -\Delta \ln \lambda / \theta$ in this later period than in the earlier period. Fortunately, by summing each source of the gains from trade over the two periods, we appear to be offsetting this effect of lagged exit of domestic firms. Our estimates over the entire 1992-2005 period neatly conform to our expectations that *W* exceeds $P+V$, resulting in a positive, upper-bound estimate for the selection gains *X*.

7. Conclusions

Our goal in this paper has been to evaluate the gains from trade when firm markups are endogenous. To achieve that, we have introduced a quite general class of preferences represented by the quadratic mean of order *r* expenditure function, due to Diewert (1976). Prior applications of this expenditure/cost function have been mainly empirical, i.e. estimating the function for specific values of the parameter *r*. In that case, the concavity and other properties of the function are checked at the estimated parameters. For theoretical purposes, we want to ensure that the function is globally well behaved. We have shown that this is the case for a symmetric function and the parameter values in Assumption 1.

Despite the general class of preferences we use, however, the crucial feature of the model allowing for multiple sources of gains from trade comes from the supply side of the model. With heterogeneous firms, the very simple result of Krugman (1979) linking a drop in the markup to (frictionless) trade no longer applies necessarily. Like in ACDR, we have shown that this result does not apply when the distribution of firm productivity is Pareto and the support is unbounded above, allowing for infinite productivity and zero costs. In the absence of fixed costs, we have found that this distribution rules out a pro-competitive gain from trade and also rules out any variety gain, suggesting that alternative forms for the productivity distribution should be investigated. Here we have focused on the Pareto distribution with a support that is bounded above. It is known from the work of Helpman, Melitz and Rubenstein (2008) that this distribution allows for a tractable gravity equation, at least in its empirical specification, and we have shown that it can also be used to obtain theoretical results.

We have investigated two cases of trade liberalization. The first, following Krugman (1979), was growth in country size. That exercise is meant to capture the movement from autarky to frictionless trade with another country. Equivalently, we can think of trade costs falling from some high level leading to autarky to zero, so the results for this case correspond to a *large* change in trade costs. We found in Proposition 2 that the product variety and pro-competitive gains from trade operate if and only if productivity is bounded above. In contrast, when productivity is unbounded as assumed by ACDR and when demand is homothetic as assumed here, then there are gains from trade due only to firm selection.[21] Despite that limitation, ACR/ACDR formula applies as an *upper bound* to the proportional gains obtained from frictionless trade when productivity is bounded and all three types of gains operate. When variety and pro-competitive gains operate (due to bounded productivity), then the selection effect is sufficiently offset that the total gains from trade are reduced. Indeed the simple formula $-d\ln\lambda/\theta$ from ACR and ACDR (with homothetic preferences), is an upper bound to the total gains (Corollary 1).

[21] These gains are offset by an "anticompetitive" effect when demand is not homothetic; see note 4.

The second case we have investigated is a small change in variable trade costs. To simplify the comparative statics in this case, we turned to the case of homogeneous firms, but with trade costs that depend on the distance to destination markets. Under plausible conditions, we still find that the ACR/ACDR formula is an upper bound to the total gains from trade (Corollary 2).[22] That result gives us the ability to solve for the selection gains as a residual from the ACR/ACDR formula, after subtracting the product variety and pro-competitive gains. Our application to the U.S. economy over 1992-2005, using translog results from Feenstra and Weinstein (2017), shows that the product variety and pro-competitive effects account for roughly 75% of the total gains due to trade expansion, while an upper-bound estimate of the selection effect is that it accounts for 25% of the total gains.

Several extensions should be pursued. First, we have not allowed for fixed costs of domestic production or exporting. The presence of such fixed costs would make the lowest profitable marginal cost endogenous (instead of unity as used here because we divide marginal cost by the reservation price). That change alone would re-introduce the product variety and pro-competitive channels as potential channels of gain. While it is challenging to allow for fixed costs and the QMOR demand system in general, its formulation is simplified in the translog case, which deserves more attention. Of course, with fixed costs then we should also allow these costs to fall as an alternative force leading to increased trade, and evaluate the gains.

Second, while the bounded Pareto gives us a particularly sharp comparison with the unbounded case, we should also allow for other distributions for firm productivity. Head, Mayer and Thoenig (2014) argue that the log-normal distribution offers a better approximation to actual firm sizes than the Pareto. Edmond, Midrigan and Xu (2015) allow for a "double" Pareto as a

[22] Bertoletti, Etro and Simonovska (2016) also find that the ACDR formula for gains is an upper-bound in a model allowing for non-homothetic preferences and with strong variety effects.

way to better approximate the actual distribution, and Holmes, Hsu and Lee (2014) also investigate distributions other than the Pareto. These distributions all deserve further attention, especially when investigating the sources of the gains from trade.

References

Arkolakis, Costas, Svetlana Demidova, Peter J. Klenow and Andrés Rodriguez-Clare, 2008, "Endogenous Variety and the Gains from Trade," *American Economic Review: Papers & Proceedings*, 98(2), 444-450.

Arkolakis, Costas, Arnaud Costinot and Andrés Rodriguez-Clare, 2010, "Gains From Trade under Monopolistic Competition: A Simple Example with Translog Expenditure Functions and Pareto Distributions of Firm-Level Productivity," available at (2/7/13): http://www.econ.yale.edu/~ka265/research/GainsfromTrade/Translog%2010-09-05.pdf

Arkolakis, Costas, Arnaud Costinot and Andrés Rodriguez-Clare, 2012, "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1), 94-130.

Arkolakis, Costas, Arnaud Costinot, David Donaldson and Andrés Rodriguez-Clare, 2017, "The Elusive Pro-Competitive Effect of Trade," http://economics.mit.edu/files/13415 .

Atkeson, Andrew, and Ariel Burstein, 2008, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98(5), 1998-2031.

Armington, Paul S., 1969, "A Theory of Demand for Products Distinguished by Place and Production." *IMF Staff Papers* 16: 159-178.

Baldwin, Richard E. and Rikard Forslid, 2010, "Trade Liberalization with Heterogeneous Firms," *Review of Development Economics*, 14(2), 161-176.

Behrens, Kristian and Yasusada Murata, 2007, "General Equilibrium Models of Monopolistic Competition: A New Approach," *Journal of Economic Theory*, 13, 776-787.

Behrens, Kristian and Yasusada Murata, 2012, "Globalization and Individual Gains from Trade," *Journal of Monetary Economics*, 59, 703-720.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen and Samuel Kortum, 2003, "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268-1290.

Bertoletti, Paolo, Federico Etro and Ina Simonovska, 2016, "International Trade with Indirect Additivity," NBER working paper no. 21984.

Broda, Christian and David E. Weinstein, 2006, "Globalization and the Gains from Variety," *Quarterly Journal of Economics,* May, 121(2), 541-585.

Chaney, Thomas, 2008, “Distorted Gravity: the Intensive and Extensive Margins of International Trade,” *American Economic Review*, 98(4), September, 1707-1721.

De Blas, Beatriz and Katheryn Niles Russ. 2015. “Understanding Markups in the Open Economy.” *American Economic Journals: Macroeconomics*, 7(2), 157-80.

Diewert, W. Erwin, 1971, “An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function,” *Journal of Political Economy*, 79(3), 481-507.

Diewert, W. Erwin, 1976, “Exact and Superlative Index Numbers,” *Journal of Econometrics,* 4, 115-145.

Diewert, W. Erwin, 1980, “Aggregation Problems in the Measurement of Capital,” in Dan Usher, ed., 1980, *The Measurement of Capital*, NBER and University of Chicago Press, 433 – 538.

Dhingra, Swati and John Morrow, 2016, “Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity,” Center for Economic Performance, London School of Economics.

Eaton, Jonathan and Samuel Kortum. 2002. “Technology, Geography and Trade.” *Econometrica,* 70(5): September, 1741-1780.

Eaton, Jonathan , Samuel S. Kortum, Sebastian Sotelo, 2013, “International Trade: Linking Micro and Macro,” in Daron Acemoglu, Manuel Arellano, and Eddie Dekel, eds., *Advances in Economics and Econometrics Tenth World Congress*, *Volume II: Applied Economics*, Cambridge University Press.

Edmond, Chris, Virgiliu Midrigan and Daniel Yi Xu, 2015, “Competition, Markups, and the Gains from International Trade,” *American Economic Review,* 105(10), 3183-3221.

Feenstra, Robert C., 2010, “Measuring the Gains from Trade under Monopolistic Competition,” *Canadian Journal of Economics,* February, 1-28.

Feenstra, Robert C., 2014, “Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms,” NBER Working Paper 19833.

Feenstra, Robert C., 2015, *Advanced International Trade,* 2nd edition, Princeton University Press.

Feenstra, Robert C. and David E. Weinstein, 2017, “Globalization, Markups and U.S. Welfare,” *Journal of Political Economy,* 125(4), 1041-1074.

Head, Keith, Thierry Mayer and Mathias Thoenig, 2014, “Welfare and Trade without Pareto,” *American Economic Review*, 104(5), May, 310-316.

Helpman, Elhanan and Paul R. Krugman, 1985, *Market Structure and Foreign Trade.* Cambridge: MIT Press.

Helpman, Elhanan, Marc J. Melitz and Yona Rubinstein, 2008, "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123, 441-487.

Holmes, Thomas J., Wen-Tai Hsu and Sanghoon Lee, 2014, "Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade," *Journal of International Economics*, 94(2), 195-206.

Kichko, Sergey, Sergey Kokovin, Evgeny Zhelobodko, 2013, "Trade Patterns and Export Pricing Under non-CES Preferences" http://ideas.repec.org/p/wsr/wpaper/y2013i124.html

Krugman, Paul K., 1979, "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 9, 469-479.

Krugman, Paul R., 1980, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review,* 70, December, 950-959.

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

Melitz, Marc J. and Gianmarco I. P. Ottaviano, 2008, "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75(1), 295-316.

Metliz, Marc J. and Stephen Redding, 2015, "New Trade Models, New Welfare Implications," *American Economic Review* 105(3), 1105-1146.

Saure, Philip, 2012, "Bounded Love of Variety and Patterns of Trade," *Open Economies Review* 23, 645-674.

Simonovska, Ina, 2015, "Income Differences and Prices of Tradables," *Review of Economic Studies,* 82 (4), 1612-1656.

Sutton, John, 2012, *Competing in Capabilities: The Globalization Process,* Clarendon Lectures in Economics, Oxford: Oxford University Press.

Tyazhelnikov, Vladimir, 2017, "The Usual Suspects: Pareto and Log-Normal Distributions of Firms' Productivity," University of California, Davis and the University of Sydney.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti and Jacques-François Thisse. 2012. "Monopolistic Competition: Beyond the Constant Elasticity of Substitution." *Econometrica,* 80(6): 2765-2784.