

NBER WORKING PAPER SERIES

THE ELUSIVE PRO-COMPETITIVE EFFECTS OF TRADE

Costas Arkolakis  
Arnaud Costinot  
Dave Donaldson  
Andrés Rodríguez-Clare

Working Paper 21370  
<http://www.nber.org/papers/w21370>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
July 2015

We thank numerous colleagues, discussants, and seminar participants for helpful comments. Federico Esposito, Cory Smith, and Anthony Tokman provided superb research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w21370.ack>

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Costas Arkolakis, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 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.

The Elusive Pro-Competitive Effects of Trade

Costas Arkolakis, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare

NBER Working Paper No. 21370

July 2015

JEL No. F1,F12

**ABSTRACT**

We study the gains from trade liberalization in models with monopolistic competition, firm-level heterogeneity, and variable markups. For a large class of demand functions used in the international macro and trade literature, we derive a parsimonious generalization of the welfare formula in Arkolakis, Costinot, and Rodríguez-Clare (2012). We then use micro-level trade data to quantify the implications of this new formula. Our main finding is that gains from trade liberalization predicted by models with variable markups are slightly lower than those predicted by models with constant markups. In this sense, pro-competitive effects of trade are elusive.

Costas Arkolakis  
Department of Economics  
Yale University, 28 Hillhouse Avenue  
P.O. Box 208268  
New Haven, CT 06520-8268  
and NBER  
costas.arkolakis@yale.edu

Dave Donaldson  
Department of Economics  
Stanford University  
579 Serra Mall  
Stanford, CA 94305  
and NBER  
ddonald@stanford.edu

Arnaud Costinot  
Department of Economics, E17-232  
MIT  
77 Massachusetts Avenue  
Cambridge MA 02139  
and NBER  
costinot@mit.edu

Andrés Rodríguez-Clare  
University of California at Berkeley  
Department of Economics  
Berkeley, CA 94720-3880  
and NBER  
andres1000@gmail.com

# 1 Introduction

How large are the gains from trade liberalization? Does the fact that trade liberalization affects firm-level markups, as documented in many micro-level studies, make these gains larger or smaller?

There are no simple answers to these questions. On the one hand, gains from trade liberalization may be larger in the presence of variable markups if opening up to trade reduces distortions on the domestic market. In the words of [Helpman and Krugman \(1989\)](#): “The idea that international trade increases competition [...] goes back to Adam Smith, and it has long been one of the reasons that economists give for believing that the gains from trade and the costs from protection are larger than their own models seem to suggest.” On the other hand, gains may be smaller if opening up to trade leads foreign firms to increase their markups. Again in the words of [Helpman and Krugman \(1989\)](#): “An occasionally popular argument about tariffs is that they will be largely absorbed through a decline in foreign markups rather than passed onto consumers—the foreigner pays the tariff.” If so, when trade costs go down, foreigners get their money back.

To make progress on these questions, we propose the following experiment. We start from a class of trade models that have been calibrated such that any model within that class: (i) is consistent with the same cross-section of bilateral trade flows and (ii) predicts the same change in bilateral trade flows for any change in trade costs. Broadly speaking, all these models agree on the macro data. However, they disagree on the micro data: some trade models within that class feature markups that vary across firms, whereas others do not. We then ask whether a researcher with access to micro data that observes variable markups at the firm-level would predict different welfare gains from trade liberalization than a researcher without access to such data that wrongly assumes constant markups across firms. The difference between the predictions of these two hypothetical researchers is what we will refer to as the pro-competitive effects of trade.

Our main finding is that gains from trade liberalization predicted by models with variable markups are slightly lower than those predicted by models with constant markups. Because a decline in trade costs indirectly lowers the residual demand for domestic goods, the former class of models predicts that domestic markups go down after trade liberalization, which reduces distortions and increases welfare. Yet, this indirect effect is dominated by the direct effect of a change in trade costs on foreign markups, which leads to lower welfare gains from trade liberalization overall. In short, pro-competitive effects of trade are elusive.

The scope of our analysis is restricted primarily by the nature of our experiment. Since

we want to focus on trade models that agree on the macro data—e.g., we do not want different welfare conclusions to be driven by the fact that a model predicts more trade than what we actually observe in commonly available datasets—we have chosen to analyze models according to which bilateral trade flows satisfy a gravity equation, which we will simply refer to as gravity models.

The benefit of focusing on such models is twofold. First, gravity models are very successful empirically and the workhorse models for quantitative work in the field; see e.g. [Head and Mayer \(2013\)](#) and [Costinot and Rodríguez-Clare \(2014\)](#). Second, welfare gains from trade liberalization in gravity models with CES utility and constant markups take a very simple form. [Arkolakis, Costinot, and Rodríguez-Clare \(2012\)](#), ACR hereafter, have shown that these gains are pinned down by two statistics: (i) the share of expenditure on domestic goods,  $\lambda$ ; and (ii) an elasticity of imports with respect to variable trade costs,  $\varepsilon$ , which we refer to as the trade elasticity. If a small change in variable trade costs raises trade openness in some country,  $d \ln \lambda < 0$ , then the associated welfare gain is given by

$$d \ln W = -d \ln \lambda / \varepsilon,$$

where  $d \ln W$  is the equivalent variation associated with the shock expressed as a percentage of the income of the representative agent.

In this paper we introduce a new set of gravity models that features monopolistic competition with firm-level heterogeneity—a standard market structure in the field—but departs from CES utility to generate variable markups at the firm-level. Our main theoretical finding is that for a general demand system that encompasses prominent alternatives to CES, the welfare effect of a small trade shock is given by

$$d \ln W = - (1 - \eta) d \ln \lambda / \varepsilon,$$

where  $\eta$  is a constant that summarizes the effects of various structural parameters, including the average elasticity of markups with respect to firm productivity. Thus the only endogenous variable that one needs to keep track of for welfare analysis remains the share of expenditure on domestic goods. The net welfare implications of changes in domestic and foreign markups boils down to a single new statistic,  $\eta$ , the sign of which determines whether or not there are pro-competitive effects of trade.

While the value of  $\eta$  is ultimately an empirical matter, it is interesting to note that under common alternatives to CES utility, such as those considered in [Krugman \(1979\)](#) and [Feenstra \(2003\)](#),  $\eta$  is weakly positive. This implies that gains from trade liberalization predicted by these new gravity models must be weakly lower than those predicted by

models with CES utility. By how much depends both on the extent to which lower trade costs get (incompletely) passed-through to domestic consumers and the extent to which domestic misallocations get alleviated. When preferences are homothetic, we show that these two forces exactly compensate each other so that gains from trade are identical to those in ACR, and hence  $\eta = 0$ . When preferences are non-homothetic, however, standard assumptions imply that the first (negative) force dominates the second (positive) force, and hence  $\eta > 0$ .

In the last part of our paper we develop a simple empirical strategy to estimate  $\eta$ . We focus on a parsimonious generalization of CES utility under which the sign of  $\eta$  depends only on one new demand parameter. Using micro-level U.S. trade data to estimate this alternative demand system, we find that  $\eta \simeq 0.04$ . Since  $d \ln \lambda$  and  $\varepsilon$  are the same in these models as in gravity models with constant markups, this positive value for  $\eta$  implies slightly lower gains from trade liberalization. This finding is robust to a number of departures from our baseline assumptions.

Our findings are related to, and have implications for, a large number of theoretical and empirical papers in the international trade literature. Many authors have studied the empirical relationship between international trade and firm-level markups; see e.g. [Levinsohn \(1993\)](#), [Harrison \(1994\)](#), [Krishna and Mitra \(1998\)](#), [Konings, Van Cayseele, and Warzynski \(2001\)](#), [Chen, Imbs, and Scott \(2009\)](#), [Loecker and Warzynski \(2012\)](#), and [Goldberg, Loecker, Khandelwal, and Pavcnik \(2012\)](#). Methodologies, data sources, and conclusions vary, but a common feature of the aforementioned papers is their exclusive focus on domestic producers. A key message from our analysis is that focusing on domestic producers may provide a misleading picture of the so-called pro-competitive effects of trade. Here we find that a decrease in trade costs reduces the markups of domestic producers. Yet, because it also increases the markups of foreign producers, gains from trade liberalization are actually lower than those predicted by standard models with CES utility.

A recent empirical paper by [Feenstra and Weinstein \(2010\)](#) is closely related to our analysis. The authors estimate a translog demand system—which is one of the demand systems covered by our analysis—to measure the contribution of new varieties and variable markups on the change in the U.S. consumer price index between 1992 and 2005. Using the fact that markups should be proportional to sales under translog, they conclude that the contribution of these two margins is of the same order of magnitude as the contribution of new varieties estimated by [Broda and Weinstein \(2006\)](#) under the assumption of CES utility. Our theoretical results show that in the translog case—like with any other homothetic preferences—the overall gains from a hypothetical decline in trade costs are exactly the same as under CES, which resonates well with [Feenstra and Weinstein \(2010\)](#)'s

empirical findings. It should be clear, however, that the two exercises are very different. [Feenstra and Weinstein \(2010\)](#) is a measurement exercise that uses a translog demand system to infer changes in particular components of the U.S. price index from observed changes in trade flows. This exercise is agnostic about the origins of changes in trade flows—whether it is driven by U.S. or foreign shocks—as well as their overall welfare implications. In contrast, the present paper is a counterfactual exercise that focuses on the welfare effect of trade liberalization, which we model as a change in variable trade costs. Our approach has both advantages and limitations: it allows us to take all general equilibrium effects into account when computing the exact welfare changes caused by trade liberalization, but it relies on the full structure of the model, not just the properties of the demand system.

The idea that gains from international trade may be higher or lower in the presence of distortions, in general, and variable markups, in particular, is an old one in the field; see e.g. [Bhagwati \(1971\)](#). A number of recent papers have revisited that idea, either analytically or quantitatively, using variations and extensions of models with firm-level heterogeneity and monopolistic competition, as in [Epifani and Gancia \(2011\)](#), [Dhingra and Morrow \(2012\)](#), and [Mrazova and Neary \(2013a\)](#), Bertrand competition, as in [de Blas and Russ \(2015\)](#) and [Holmes, Hsu, and Lee \(2015\)](#), and Cournot competition, as in [Edmond, Midrigan, and Xu \(2015\)](#). Our approach differs from existing work in three important ways. First, we focus on trade models with variable markups that satisfy the same macro-level restrictions as trade models with constant markups. Besides the empirical appeal of focusing on gravity models, this provides an ideal theoretical benchmark to study how departures from CES utility may affect the welfare gains from trade liberalization. Since the macro-level behavior of new trade models considered in this paper is exactly controlled for, new gains may only reflect new micro-level considerations. Second, we provide a theoretical framework in which the welfare implications of variable markups can be signed and quantified using only one new statistic,  $\eta$ . Hence counterfactual analysis can still be conducted in a parsimonious manner. Third, we develop a new empirical strategy to estimate  $\eta$  and to compute the welfare gains from trade liberalization using micro-level trade data.

The rest of the paper is organized as follows. Section 2 describes our theoretical framework. Section 3 characterizes the trade equilibrium. Section 4 derives our new welfare formula. Section 5 presents our empirical estimates. Section 6 explores the robustness of our results. Section 7 offers some concluding remarks.

## 2 Theoretical Framework

Consider a world economy comprising  $i = 1, \dots, n$  countries, one factor of production, labor, and a continuum of differentiated goods  $\omega \in \Omega$ . All individuals are endowed with one unit of labor, are perfectly mobile across the production of different goods, and are immobile across countries.  $L_i$  denotes the population and  $w_i$  denotes the wage in country  $i$ . Under free entry,  $w_i$  will also be equal to income per capita in country  $i$ .

### 2.1 Consumers

The goal of our paper is to study the implications of trade models with monopolistic competition for the magnitude of the gains from trade in economies in which markups are variable. This requires departing from the assumption of CES utility. Three prominent alternatives in the international trade and international macro literature are: (i) additively separable, but non-CES utility functions, as in the pioneering work of [Krugman \(1979\)](#) and the more recent work of [Behrens and Murata \(2009\)](#), [Behrens, Mion, Murata, and Sudekum \(2009\)](#), [Saure \(2009\)](#), [Simonovska \(2009\)](#), [Dhingra and Morrow \(2012\)](#) and [Zhelobodko, Kokovin, Parenti, and Thisse \(2011\)](#); (ii) a symmetric translog expenditure function, as in [Feenstra \(2003\)](#), [Bergin and Feenstra \(2009\)](#), [Feenstra and Weinstein \(2010\)](#), [Novy \(2010\)](#), and [Rodriguez-Lopez \(2010\)](#), as well as its strict generalization to quadratic mean of order  $r$  (QMOR) expenditure functions, as in [Feenstra \(2014\)](#); (iii) Kimball preferences, as in [Kimball \(1995\)](#) and [Klenow and Willis \(2006\)](#). In our baseline analysis, we study a general demand system for differentiated goods that encompasses all of them.<sup>1</sup>

All consumers have the same preferences. If a consumer with income  $w$  faces a schedule of prices  $\mathbf{p} \equiv \{p_\omega\}_{\omega \in \Omega}$ , her Marshallian demand for any differentiated good  $\omega$  is

$$q_\omega(\mathbf{p}, w) = Q(\mathbf{p}, w) D(p_\omega / P(\mathbf{p}, w)), \quad (1)$$

where  $Q(\mathbf{p}, w)$  and  $P(\mathbf{p}, w)$  are two aggregate demand shifters, which firms will take as given in subsequent sections. Note that whereas  $Q(\mathbf{p}, w)$  only affects the level of demand,  $P(\mathbf{p}, w)$  affects both the level and elasticity of demand, which will have implications for firm-level markups. As discussed in [Burstein and Gopinath \(2013\)](#), equation (1) is a common feature of many models in the macroeconomic literature on international pricing.

---

<sup>1</sup>A trivial generalization of this demand system also nests the case of quadratic, but non-separable utility function, as in [Ottaviano, Tabuchi, and Thisse \(2002\)](#) and [Melitz and Ottaviano \(2008\)](#), when a homogenous “outside good” is introduced. We have discussed the additional considerations associated with the existence of an outside good in the earlier version of this paper. Details are available upon request.

To complete the description of our demand system, we assume that  $Q(\mathbf{p}, w)$  and  $P(\mathbf{p}, w)$  are jointly determined as the solution of the following system of two equations,

$$\int_{\omega \in \Omega} [H(p_\omega/P)]^\beta [p_\omega QD(p_\omega/P)]^{1-\beta} d\omega = w^{1-\beta}, \quad (2)$$

$$Q^{1-\beta} \left[ \int_{\omega \in \Omega} p_\omega QD(p_\omega/P) d\omega \right]^\beta = w^\beta, \quad (3)$$

with  $\beta \in \{0, 1\}$  and  $H(\cdot)$  strictly increasing and strictly concave. As shown in Appendix A.1, our demand system nests the case of additively separable utility functions when  $\beta = 0$  and the case of QMOR expenditure functions and Kimball preferences when  $\beta = 1$ . In the former case, equation (2) reduces to the consumer's budget constraint with  $P(\mathbf{p}, w)$  equal to the inverse of the Lagrange multiplier associated with that constraint, whereas equation (3) merely implies that  $Q(\mathbf{p}, w) = 1$ . In the latter case,  $P(\mathbf{p}, w)$  remains determined by equation (2), which becomes  $\int_{\omega \in \Omega} H(p_\omega/P) d\omega = 1$ , but the consumer's budget constraint is now captured by equation (3) with  $Q(\mathbf{p}, w)$  set such that budget balance holds.

Three properties of the general demand system introduced above are worth emphasizing. First, the own-price elasticity  $\partial \ln D(p_\omega/P(\mathbf{p}, w)) / \partial \ln p_\omega$  is allowed to vary with prices, which will generate variable markups under monopolistic competition. Second, other prices only affect the demand for good  $\omega$  through their effect on the aggregate demand shifters,  $Q(\mathbf{p}, w)$  and  $P(\mathbf{p}, w)$ .<sup>2</sup> Third, the demand parameter  $\beta$  controls whether preferences are homothetic or not. If  $\beta = 1$ , equations (2) and (3) imply that  $P(\mathbf{p}, w)$  is independent of  $w$  and that  $Q(\mathbf{p}, w)$  is proportional to  $w$ . Thus preferences are homothetic. Conversely, if  $\beta = 0$ , one can show that preferences are non-homothetic unless  $D(\cdot)$  is iso-elastic, i.e. utility functions are CES.<sup>3</sup> The parameter  $\beta$  will influence the magnitude of general equilibrium effects and play a crucial role in our welfare analysis.

Compared to most papers in the existing trade literature, either theoretical or empirical, we do not impose any functional form restriction on  $D(\cdot)$ . The only restriction that we impose on  $D(\cdot)$  in our theoretical analysis is that it features a choke price.

**A1. [Choke Price]** *There exists  $a \in \mathbb{R}$  such that for all  $x \geq a$ ,  $D(x) = 0$ .*

<sup>2</sup>In this regard, our specification is more restrictive than the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980). Compared to AIDS, however, our specification does not impose any functional form restriction on  $Q(\mathbf{p}, w)$  and  $P(\mathbf{p}, w)$ .

<sup>3</sup>The formal argument can be found in the Appendix. Intuitively, CES utility functions correspond to the knife-edge case in which  $\beta$  admits multiple values. CES utility functions can be thought either as a special case of additively separable utility functions—and derived under the assumption  $\beta = 0$ —or as a special case of QMOR expenditure functions or Kimball preferences—and derived under the assumption  $\beta = 1$ .



Without loss of generality, we normalize  $a$  to one in the rest of our analysis so that the aggregate demand shifter  $P(\mathbf{p}, w)$  is also equal to the choke price. In the absence of fixed costs of accessing domestic and foreign markets—which is the situation that we will focus on—Assumption A1 implies that the creation and destruction of “cut-off” goods have no first-order effects on welfare at the margin. Indeed, if there was some benefit from consuming these goods, they would have been consumed in strictly positive amounts.

Assumption A1 is appealing on both theoretical and empirical grounds. From a theoretical standpoint, Assumption A1 provides an instructive polar case. In models with CES utility, such as those studied in [Arkolakis, Costinot, and Rodríguez-Clare \(2012\)](#), there are welfare gains from new goods, but markups are fixed. In our baseline analysis, firm-level markups can vary, but there are no welfare gains from new goods. From an empirical standpoint, Assumption A1 also appears more relevant in practice than the alternative assumption under which consumers never stop consuming any good, regardless of how expensive these goods may become. The two previous observations notwithstanding, we explore the sensitivity of our results to Assumption A1 in Section 6.

For future derivations, it is convenient to write the demand function in a way that makes explicit the symmetry across goods as well as the way in which the aggregate demand shifters,  $Q(\mathbf{p}, w)$  and  $P(\mathbf{p}, w)$ , affect the demand for all goods. Thus, we write  $q_\omega(\mathbf{p}, w) \equiv q(p_\omega, Q(\mathbf{p}, w), P(\mathbf{p}, w))$ , with

$$q(p_\omega, Q, P) = QD(p_\omega/P). \quad (4)$$

For future reference, it is also worth pointing out that under Assumption A1, CES utility functions are ruled out. Based on our earlier discussion, the preferences that we consider in this paper are therefore homothetic if and only if  $\beta = 1$ .

## 2.2 Firms

Firms compete under monopolistic competition with free entry. There is a large number of ex ante identical firms in each country  $i$  that have the option of hiring  $F_i > 0$  units of labor to enter the industry. We denote by  $N_i$  the measure of firms incurring this fixed entry cost in country  $i$ . After  $w_i F_i$  has been paid, production of any differentiated good is subject to constant returns to scale. For a firm with productivity  $z$  in country  $i$ , the constant cost of delivering one unit of the variety associated with that firm to country  $j$  is given by  $w_i \tau_{ij}/z$ , where  $\tau_{ij} \geq 1$  is an iceberg trade cost. We assume that only international trade is subject to frictions,  $\tau_{ii} = 1$ . As mentioned earlier, there are no fixed costs of accessing domestic and foreign markets. Thus, the selection of firms across markets is driven entirely by the

existence of a choke price, as in Melitz and Ottaviano (2008). Throughout our analysis, we assume that good markets are perfectly segmented across countries and that parallel trade is prohibited so that firms charge the optimal monopoly price in each market.

As in Melitz (2003), firm-level productivity  $z$  is the realization of a random variable  $Z_i$  drawn independently across firms from a distribution  $G_i$ . We assume that  $G_i$  is Pareto with the same shape parameter  $\theta > 0$  around the world.

**A2. [Pareto]** For all  $z \geq b_i$ ,  $G_i(z) \equiv \Pr(Z_i \leq z) = 1 - (b_i/z)^\theta$ , with  $\theta > 0$ .

While by far the most common distributional assumption in models of monopolistic competition with firm-level heterogeneity—even when utility functions are not CES, see e.g. Melitz and Ottaviano (2008), Behrens, Mion, Murata, and Sudekum (2009), Simonovska (2009), and Rodriguez-Lopez (2010)—Assumption A2 is obviously a strong restriction on the supply-side of our economy. So it is worth pausing to discuss its main implications.

As we will demonstrate below, the main benefit of Assumption A2 is that trade flows will satisfy the same gravity equation as in models with CES utility. This will allow us to calibrate our model and conduct counterfactual analysis in the exact same way as in ACR. Accordingly, we will be able to ask and answer the following question: Conditional on being consistent with the same macro data, do models featuring variable markups predict different welfare gains from trade liberalization? In our view, this is a theoretically clean way to compare the welfare predictions of different trade models.

Given the generality of the demand system considered in Section 2.1, it should be clear that Assumption A2 is no less appealing on empirical grounds than under the assumption of CES utility. As documented by Axtell (2001) and Eaton, Kortum, and Kramarz (2011), among others, Pareto distributions provide a reasonable approximation for the right tail of the observed distribution of firm sales. Since Pareto distributions of firm sales can be generated from a model of monopolistic competition with CES utility and Pareto distributions of firm-level productivity, the previous facts are often given as evidence in favor of Assumption A2. Although demand functions derived from CES utility do not satisfy A1, it is easy to construct generalizations of CES demands that satisfy A1, behave like CES demands for the right tail of the distribution of firm sales, and provide a better fit for the left tail. We provide one such example in our empirical application.

Perhaps the main concern regarding Assumption A2 is that it may be too much of a straight jacket, i.e., that we may be assuming through functional form assumptions whether gains from trade liberalization predicted by models with variable markups will be larger, smaller, or the same. As Proposition 1 will formally demonstrate, this is not so. Although Assumption A2 has strong implications for the univariate distribution of firm-

level markups—as we will see, it is unaffected by changes in trade costs—this knife-edge feature does not preclude the existence of variable markups to increase or decrease—in theory—the welfare gains from trade liberalization. As we discuss in Section 4.2, what matters for welfare is not the univariate distribution of markups, but the bivariate distribution of markups and employment, which is free to vary in our model. In Section 6, we further discuss the sensitivity of our results to departures from Assumption A2.

### 3 Trade Equilibrium

In this section we characterize the trade equilibrium for arbitrary values of trade costs. We proceed in two steps. We first study how the demand system introduced in Section 2 shapes firm-level variables. We then describe how firm-level decisions aggregate up to determine bilateral trade flows and the measure of firms active in each market.

#### 3.1 Firm-level Variables

Consider the optimization problem of a firm producing good  $\omega$  in country  $i$  and selling it in a certain destination  $j$ . To simplify notation, and without risk of confusion, we drop indices for now and denote by  $c \equiv w_i \tau_{ij} / z$  the constant marginal cost of serving the market for a particular firm and by  $Q$  and  $P$  the two aggregate shifters of demand in the destination country, respectively. Under monopolistic competition with segmented good markets, the firm chooses its market-specific price  $p$  in order to maximize profits in each market,

$$\pi(c, Q, P) = \max_p \{(p - c) q(p, Q, P)\},$$

taking  $Q$  and  $P$  as given. The associated first-order condition is

$$(p - c) / p = -1 / (\partial \ln q(p, Q, P) / \partial \ln p),$$

which states that monopoly markups are inversely related to the elasticity of demand.

**Firm-level markups.** We use  $m \equiv p / c$  as our measure of firm-level markups. Combining the previous expression with equation (4), we can express  $m$  as the implicit solution of

$$m = \varepsilon_D(m/v) / (\varepsilon_D(m/v) - 1), \tag{5}$$

where  $\varepsilon_D(x) \equiv -\partial \ln D(x) / \partial \ln x$  measures the elasticity of demand and  $v \equiv P / c$  can be thought of as a market-specific measure of the efficiency of the firm relative to other

firms participating in that market, as summarized by  $P$ . Equation (5) implies that the aggregate demand shifter  $P$  is a sufficient statistic for all indirect effects that may lead a firm to change its price in a particular market.

We assume that for any  $v > 0$ , there exists a unique  $m \equiv \mu(v)$  that solves equation (5). Assuming that  $\varepsilon'_D < 0$  is a sufficient, but not necessary condition for existence and uniqueness. The properties of the markup function  $\mu(v)$  derive from the properties of  $D(\cdot)$ . Since  $\lim_{x \rightarrow 1} D(x) = 0$  by Assumption A1, we must also have  $\lim_{x \rightarrow 1} \varepsilon_D(x) = \infty$ , which implies  $\mu(1) = 1$ . Thus, the choke price in a market is equal to the marginal cost of the least efficient firm active in that market. Whether markups are monotonically increasing in productivity depends on the monotonicity of  $\varepsilon_D$ . As is well-known and demonstrated in Appendix A.2, if demand functions are log-concave in log-prices,  $\varepsilon'_D < 0$ , then  $\mu' > 0$  so that more efficient firms charge higher markups.<sup>4</sup>

**Firm-level sales and profits.** In any given market, the price charged by a firm with marginal cost  $c$  and relative efficiency  $v$  is given by  $p(c, v) = c\mu(v)$ . Given this pricing rule, the total sales faced by a firm with marginal cost  $c$  and relative efficiency  $v$  in a market with aggregate demand shifter  $Q$  and population  $L$ , are equal to

$$x(c, v, Q, L) \equiv LQc\mu(v)D(\mu(v)/v). \quad (6)$$

In turn, the profits of a firm with marginal cost  $c$  and relative efficiency  $v$  selling in a market with aggregate shifter  $Q$  and population  $L$  are given by

$$\pi(c, v, Q, L) \equiv ((\mu(v) - 1)/\mu(v)) x(c, v, Q, L). \quad (7)$$

The relationship between profits and sales is the same as in models of monopolistic competition with CES utility, except that markups are now allowed to vary across firms.

## 3.2 Aggregate Variables

**Aggregate sales and profits.** Let  $X_{ij}$  denote the total sales by firms from country  $i$  in country  $j$ . Only firms with marginal cost  $c \leq P_j$  sell in country  $j$ . Thus there exists a productivity cut-off  $z_{ij}^* \equiv w_i\tau_{ij}/P_j$  such that a firm from country  $i$  sells in country  $j$  if and only if its productivity  $z \geq z_{ij}^*$ . Accordingly, we can express the bilateral trade flows

---

<sup>4</sup>Mrazova and Neary (2013b) refer to this condition as “subconvexity.” Although Assumption A1 requires demand functions to be log-concave in log-prices locally around 1, it should be clear that it does not require them to be log-concave in log-prices globally.

between the two countries as

$$X_{ij} = N_i \int_{z_{ij}^*}^{\infty} x(w_i \tau_{ij} / z, z / z_{ij}^*, Q_j, L_j) dG_i(z).$$

Combining this expression with equation (6) and using our Pareto assumption A2, we get, after simplifications,

$$X_{ij} = \chi N_i b_i^\theta (w_i \tau_{ij})^{-\theta} L_j Q_j (P_j)^{1+\theta}. \quad (8)$$

where  $\chi \equiv \theta \int_1^\infty (\mu(v)/v) D(\mu(v)/v) v^{-\theta-1} dv > 0$  is a constant that affects overall sales.<sup>5</sup>

Let  $\Pi_{ij}$  denote aggregate profits by firms from country  $i$  in country  $j$  gross of fixed entry costs. This is given by

$$\Pi_{ij} = N_i \int_{z_{ij}^*}^{\infty} \pi(w_i \tau_{ij} / z, z / z_{ij}^*, Q_j, L_j) dG_i(z).$$

Using equations (6) and (7), and again invoking our Pareto assumption A2, we get

$$\Pi_{ij} = \pi N_i b_i^\theta (w_i \tau_{ij})^{-\theta} L_j Q_j (P_j)^{1+\theta}, \quad (9)$$

where  $\pi \equiv \theta \int_1^\infty (\mu(v) - 1) D(\mu(v)/v) v^{-\theta-2} dv > 0$  is a constant that affect overall profits. For future reference, note that Equations (8) and (9) imply that aggregate profits are a constant share of aggregate sales,

$$\Pi_{ij} = (\pi/\chi) X_{ij}. \quad (10)$$

**Measure of Entrants.** Free entry requires that the sum of expected profits across all markets be equal to the entry costs, which can be expressed as

$$\sum_j \Pi_{ij} = w_i F_i N_i. \quad (11)$$

Labor market clearing and free entry, in turn, require that total sales across all markets be

---

<sup>5</sup>Equation (8) implicitly assumes that the lower-bound of the Pareto distribution  $b_i$  is small enough so that the firm with minimum productivity  $b_i$  always prefers to stay out of the market,  $b_i < z_{ij}^*$ . This implies that the “extensive” margin of trade is active for all country pairs, which is the empirically relevant case. It also implicitly assumes that the behavior of the distribution of firm-level productivity and demand in the upper-tail is such that  $\chi$  is finite.

equal to the total wage bill,

$$\sum_j X_{ij} = w_i L_i. \quad (12)$$

Equations (10), (11) and (12) imply that the measure of entrants in each country is fully determined by country size  $L_i$  and the fixed cost of entry  $F_i$ ,

$$N_i = (\pi/\chi)(L_i/F_i). \quad (13)$$

This implies, in particular, that entry levels are invariant to changes in trade costs.<sup>6</sup>

**Summary.** A trade equilibrium corresponds to price schedules,  $(p_1, \dots, p_n)$ , measures of entrants,  $(N_1, \dots, N_n)$ , and wages,  $(w_1, \dots, w_n)$ , such that (i) prices set in country  $j$  by firms with productivity  $z$  located in country  $i$  maximize their profits:

$$p_{ij}(z) = (w_i \tau_{ij}/z) \mu(P_j z / (w_i \tau_{ij})) \quad (14)$$

if  $z \geq w_i \tau_{ij}/P_j$  and  $p_{ij}(z) \geq w_i \tau_{ij}/z$  otherwise; (ii) measures of entrants are consistent with free entry, equation (13); and (iii) wages are consistent with labor market clearing, equation (12), with aggregate sales  $X_{ij}$  determined by equation (8) and aggregate demand shifters,  $Q_j$  and  $P_j$ , determined by equations (2) and (3).<sup>7</sup>

### 3.3 Discussion

In spite of the fact that the pricing behavior of firms, as summarized by equation (14), is very different in the present environment than in trade models with CES utility, bilateral trade flows still satisfy a gravity equation. Indeed, by equation (8), we have

$$X_{ij} = \frac{N_i b_i^\theta (w_i \tau_{ij})^{-\theta} E_j}{\sum_k N_k b_k^\theta (w_k \tau_{kj})^{-\theta}}, \quad (15)$$

where  $E_j \equiv \sum_k X_{kj}$  denotes total expenditure in country  $j$ . Together with the fact that aggregate profits are a constant share of aggregate sales, the previous gravity equation implies that the macro-level predictions of models considered in this paper, i.e., counterfactual predictions about wages and bilateral trade flows, are exactly the same as in

---

<sup>6</sup>In our baseline analysis, the assumption of free entry is therefore innocuous. Since aggregate profits are a constant share of aggregate sales, our welfare and counterfactual analysis would be unchanged if we were to assume that the measure of entrants,  $N_i$ , is exogenously given. As shown in ACR, a similar property holds in models of monopolistic competition with CES utility and Pareto distributions of firm-level productivity.

<sup>7</sup>Budget balance in all countries implies that one of these  $n$  labor market conditions is redundant.

gravity models with CES utility, such as [Krugman \(1980\)](#), [Eaton and Kortum \(2002\)](#), [Anderson and Van Wincoop \(2003\)](#), and [Eaton, Kortum, and Kramarz \(2011\)](#). As shown in [Appendix A.3](#), once the models with variable markups considered in this paper are calibrated to match the trade elasticity  $\theta$  and the observed trade flows  $\{X_{ij}\}$ , they must predict the exact same changes in wages and trade flows for any change in variable trade costs.<sup>8</sup> Yet, as we will see, differences in the behavior of firms at the micro-level open up the possibility of new welfare implications.

Before we turn to our welfare analysis, it is worth emphasizing again that there will be no gains from new varieties associated with trade liberalization in the present environment. Because  $N_i$  is unaffected by changes in trade costs, changes in the set of varieties available in a given importing country  $j$  may only come from selection effects, i.e. changes in the productivity cut-offs  $z_{ij}^*$ . But as mentioned before, in the absence of fixed costs of accessing domestic and foreign markets, the creation or destruction of “cut-off” varieties has no welfare consequences either. Thus our focus in this paper is squarely on the welfare implications of variable markups at the firm-level.

## 4 Welfare Analysis

In this section we explore the pro-competitive effects of trade, or lack thereof, in the economic environment described in [Sections 2 and 3](#). We focus on a small change in trade costs from  $\tau \equiv \{\tau_{ij}\}$  to  $\tau' \equiv \{\tau_{ij} + d\tau_{ij}\}$ . ACR show that under monopolistic competition with Pareto distributions of firm-level productivity and CES utility, the equivalent variation associated with such a change—namely, the percentage change in income that would be equivalent to the change in trade costs in terms of its welfare impact—is given by

$$d \ln W_j = -d \ln \lambda_{jj} / \theta,$$

where, like in the present paper,  $\theta$  is the shape parameter of the Pareto distribution and  $d \ln \lambda_{jj}$  is the change in the share of domestic expenditure on domestic goods caused by the change from  $\tau$  to  $\tau'$ . Since  $\theta > 0$ , the equivalent variation  $d \ln W_j$  is positive if a change in trade costs leads to more trade,  $d \ln \lambda_{jj} < 0$ . We now investigate how going from CES utility to the demand system described in [equation \(1\)](#) affects the above formula.

---

<sup>8</sup>This derives from the fact that the three macro-level restrictions, R1, R2, and R3', imposed in ACR still hold in this environment.



## 4.1 A New Formula

Without loss of generality, we use labor in country  $j$  as our numeraire so that  $w_j = 1$  before and after the change in trade costs. Under free entry, income per capita in country  $j$  is equal to the wage  $w_j$ . Thus, the percentage change in income,  $d \ln W_j$ , equivalent to the change in trade costs from  $\tau$  to  $\tau'$  can be computed as the negative of the percentage change in the expenditure function,  $d \ln e_j$ , of a representative consumer in country  $j$ . This is what we focus on next.

By Shephard's lemma, we know that  $de_j/dp_{\omega,j} = q(p_{\omega,j}, Q_j, P_j) \equiv q_{\omega,j}$  for all  $\omega \in \Omega$ . Since all price changes associated with a change from  $\tau$  to  $\tau'$  are infinitesimal,<sup>9</sup> we can express the associated change in expenditure as

$$de_j = \sum_i \int_{\omega \in \Omega_{ij}} q_{\omega,j} dp_{\omega,j} d\omega,$$

where  $\Omega_{ij}$  is the set of goods produced in country  $i$  and exported to country  $j$  and  $dp_{\omega,j}$  is the change in the price of good  $\omega$  in country  $j$  caused by the change from  $\tau$  to  $\tau'$ . The previous expression can be rearranged in logs as

$$d \ln e_j = \sum_i \int_{\omega \in \Omega_{ij}} \lambda_{\omega,j} d \ln p_{\omega,j} d\omega, \quad (16)$$

where  $\lambda_{\omega,j} \equiv p_{\omega,j} q_{\omega,j} / e_j$  is the share of expenditure on good  $\omega$  in country  $j$  in the initial equilibrium. Using equation (14) and the fact that firms from country  $i$  only sell in country  $j$  if  $z \geq z_{ij}^*$ , we obtain

$$d \ln e_j = \sum_i \int_{z_{ij}^*}^{\infty} \lambda_{ij}(z) (d \ln c_{ij} + d \ln m_{ij}(z)) dG_i(z), \quad (17)$$

where

$$\lambda_{ij}(z) \equiv \frac{N_i x(w_i \tau_{ij} / z, z / z_{ij}^*, Q_j, L_j)}{\sum_k \int_{z_{kj}^*}^{\infty} N_k x(w_k \tau_{kj} / z, z / z_{kj}^*, Q_j, L_j) dG_k(z)}$$

denotes the share of expenditure in country  $j$  on goods produced by firms from country  $i$  with productivity  $z$ ,  $c_{ij} \equiv w_i \tau_{ij}$ , and  $m_{ij}(z) \equiv \mu(z / z_{ij}^*)$ . Equation (17) states that the

<sup>9</sup>In principle, price changes may not be infinitesimal because of the creation of "new" goods or the destruction of "old" ones. This may happen for two reasons: (i) a change in the number of entrants  $N$  or (ii) a change in the productivity cut-off  $z^*$ . Since the number of entrants is independent of trade costs, as argued above, (i) is never an issue. Since the price of goods at the productivity cut-off is equal to the choke price, (ii) is never an issue either. This would not be true under CES utility functions and fixed exporting costs. In this case, changes in productivity cut-offs are associated with non-infinitesimal changes in prices since goods at the margin go from a finite (selling) price to an (infinite) reservation price, or vice versa. We come back to this point in detail in Section 6.



percentage change in expenditure is equal to a weighted sum of the percentage change in prices, with the percentage changes in prices themselves being the sum of the percentage change in marginal costs,  $d \ln c_{ij}$ , and markups,  $d \ln m_{ij}(z)$ .

Let  $\lambda_{ij} \equiv X_{ij}/E_j$  denote the total share of expenditure on goods from country  $i$  in country  $j$  and let  $\rho_{ij} \equiv \int_{z_{ij}^*}^{\infty} \rho(z/z_{ij}^*) \frac{\lambda_{ij}(z)}{\lambda_{ij}} dG_i(z) dz$  denote the weighted average of the markup elasticities,  $\rho(v) \equiv d \ln \mu(v)/d \ln v$ . Using this notation, we can simplify equation (17) into

$$d \ln e_j = \sum_i \lambda_{ij} \left( d \ln c_{ij} - \rho_{ij} d \ln z_{ij}^* \right).$$

Using Assumption A2, as well as the definition of  $\lambda_{ij}(z)$ , one can show that the markup elasticity, like the trade elasticity, must be common across countries (i.e.,  $\rho_{ij} = \rho$  for all  $i, j$ ) and given by the constant

$$\rho \equiv \int_1^{\infty} \frac{d \ln \mu(v)}{d \ln v} \frac{(\mu(v)/v) D(\mu(v)/v) v^{-\theta-1}}{\int_1^{\infty} (\mu(v')/v') D(\mu(v')/v') (v')^{-\theta-1} dv'} dv. \quad (18)$$

Finally, using the fact that the productivity cut-off satisfies  $z_{ij}^* = c_{ij}/P_j$ , we can rearrange the expression above as

$$d \ln e_j = \underbrace{\sum_i \lambda_{ij} d \ln c_{ij}}_{\text{Change in marginal costs}} + \underbrace{(-\rho) \sum_i \lambda_{ij} d \ln c_{ij}}_{\text{Direct markup effect}} + \underbrace{\rho d \ln P_j}_{\text{Indirect markup effect}}. \quad (19)$$

To fix ideas, consider a “good” trade shock,  $\sum_i \lambda_{ij} d \ln c_{ij} < 0$ . If markups were constant,  $\rho = 0$ , the only effect of such a shock would be given by the first term on the RHS of (19). Here, the fact that firms adjust their markups in response to a trade shock leads to two additional terms. The second term on the RHS of (19) is a direct effect. *Ceteris paribus*, a decrease in trade costs makes exporting firms relatively more productive, which leads to changes in markups, by equation (5). If  $\rho > 0$ , which, as we will argue, is the empirically relevant case, we see that the direct effect of markups tends to *lower* gains from trade liberalization. The reason is simple. There is incomplete pass-through of changes in marginal costs from foreign exporters to domestic consumers. Firms that become more productive because of lower trade costs tend to raise their markups ( $\rho > 0$ ), leading to lower welfare gains ( $(-\rho) \sum_i \lambda_{ij} d \ln c_{ij} > 0$ ). The third term on the RHS of (19) is an indirect effect. It captures the change in markups caused by changes in the aggregate demand shifter,  $P_j$ . If trade liberalization leads to a decline in  $P_j$ , reflecting a more intense level of competition, then  $\rho > 0$  implies a decline in domestic and foreign markups and *higher* gains from trade liberalization.

Based on the previous discussion, whether or not there are pro-competitive effects of trade liberalization—in the sense of larger welfare gains than in models with constant markups—depends on a horse race between the direct and indirect markup effects. In order to compare these two effects, we need to compare the change in marginal costs,  $\sum_i \lambda_{ij} d \ln c_{ij}$ , to the change in the aggregate demand shifter,  $d \ln P_j$ . We can do so by using equations (2) and (3). Given our choice of numeraire and Assumption A2, these two equations imply

$$\kappa Q_j^{1-\beta} P_j^{\theta+1-\beta} \left( \sum_i N_i b_i^\theta c_{ij}^{-\theta} \right) = 1, \quad (20)$$

$$\chi^\beta Q_j P_j^{\beta(1+\theta)} \left( \sum_i N_i b_i^\theta c_{ij}^{-\theta} \right)^\beta = 1, \quad (21)$$

with  $\kappa \equiv \theta \int_1^\infty [H(\mu(v)/v)]^\beta [(\mu(v)/v) D(\mu(v)/v)]^{1-\beta} v^{-1-\theta} dv$ . For  $\beta \in \{0, 1\}$ , equations (20) and (21) imply  $P_j = \left( \kappa \sum_i N_i b_i^\theta c_{ij}^{-\theta} \right)^{1/(\theta+1-\beta)}$ . Taking logs and totally differentiating, we therefore have

$$d \ln P_j = (\theta / (\theta + 1 - \beta)) \sum_i \lambda_{ij} d \ln c_{ij}. \quad (22)$$

Since  $\theta > 0$  and  $\beta \leq 1$ , we see that a “good” trade shock,  $\sum_i \lambda_{ij} d \ln c_{ij} < 0$ , is necessarily accompanied by a decline in the aggregate demand shifter,  $d \ln P_j < 0$ , as hinted to in the previous paragraph. As we can also see from equation (22), the ranking of the direct and indirect markup effects is pinned down by the preference parameter  $\beta$ . Namely, the indirect markup effect is larger if preferences are homothetic ( $\beta = 1$ ) than if they are not ( $\beta = 0$ ).

Plugging equation (22) into equation (19), we finally get

$$d \ln e_j = (1 - \rho((1 - \beta) / (1 - \beta + \theta))) \sum_i \lambda_{ij} d \ln c_{ij}. \quad (23)$$

As in ACR, by differentiating the gravity equation (15), one can show that  $\sum_i \lambda_{ij} d \ln c_{ij}$  is equal to  $d \ln \lambda_{jj} / \theta$ . Combining this observation with equation (23), we obtain

$$d \ln e_j = (1 - \rho((1 - \beta) / (1 - \beta + \theta))) d \ln \lambda_{jj} / \theta. \quad (24)$$

Given free entry and our choice of numeraire, we have already argued that  $d \ln W_j = -d \ln e_j$ . Thus, the main theoretical result of our paper can be stated as follows.

**Proposition 1** *Suppose that Assumptions A1 and A2 hold. Then the equivalent variation asso-*

ciated with a small trade shock in country  $j$  is given by

$$d \ln W_j = - (1 - \eta) d \ln \lambda_{jj} / \theta, \text{ with } \eta \equiv \rho ((1 - \beta) / (1 - \beta + \theta)).$$

Although markups are allowed to vary at the firm-level, we see that welfare analysis can still be conducted using only a few sufficient statistics. In particular, like in ACR, the share of expenditure on domestic goods,  $\lambda_{jj}$ , is the only endogenous variable whose changes need to be observed in order to evaluate the welfare consequences of changes in trade costs.

Compared to ACR, however, Proposition 1 highlights the potential importance of micro-level data. In spite of the fact that the models analyzed in this paper satisfy the same macro-level restrictions as in ACR, different predictions at the micro-level—namely the variation in markups across firms—lead to different welfare conclusions. Since bilateral trade flows satisfy the gravity equation (15) and the measure of entrants is independent of trade costs, the value of  $d \ln \lambda_{jj} / \theta$  caused by a given trade shock is exactly the same as in ACR. Yet, welfare changes are no longer pinned down by  $d \ln \lambda_{jj} / \theta$ , but depend on an extra statistic,  $\eta$ . According to Proposition 1, if  $\eta < 0$ , then conditional on matching the same macro data, models with variable markups will predict larger welfare gains. Conversely, if  $\eta > 0$ , they will predict smaller welfare gains.

The sign of  $\eta$ , in turn, depends on two considerations. First, is the preference parameter  $\beta$  equal to zero or one? This determines the relative importance of the direct and indirect markup effects. Second, is the average markup elasticity  $\rho$  positive or negative? This determines which of the direct and indirect markup effects is welfare-enhancing. While the answer to these questions is ultimately an empirical matter, which we deal with in Section 5, a number of theoretical issues are worth clarifying at this point.

## 4.2 Discussion

In Section 2, we have mentioned three special cases of our general demand system: (i) additively separable utility functions, which imply  $\beta = 0$ ; (ii) QMOR expenditure functions, which imply  $\beta = 1$ ; and (iii) Kimball preferences, which also imply  $\beta = 1$ . In cases (ii) and (iii), Proposition 1 implies that gains from trade liberalization are exactly the same as those predicted by the models with constant markups considered in ACR. In case (i), whether  $\eta > 0$  or  $< 0$  depends on the sign of the (average) markup elasticity,  $\rho$ . Since the pioneering work of Krugman (1979), a common assumption in the literature is that the demand elasticity is decreasing with the level consumption,  $\varepsilon'_D < 0$ , which

implies  $\rho > 0$ .<sup>10</sup> Under this assumption,  $\eta > 0$ , and the gains from trade liberalization predicted by models with variable markups are *lower* than those predicted by models with constant markups. In other words, under standard alternatives to CES utility, the existence of variable markups at the firm-level (weakly) dampens rather than magnifies the gains from trade liberalization.

**What are the economic forces behind lower gains from trade liberalization?** As we formally establish in Appendix A.4, a strong implication of Assumption A2 is that if markups are an increasing function of firm-level productivity—as they would be under standard alternatives to CES utility—then the univariate distribution of markups is independent of the level of trade costs. This reflects the countervailing effects of a change in trade costs on markups. On the one hand, a decline in trade costs,  $\tau_{ij}$ , leads current exporters from country  $i$  to increase their markups in country  $j$ . On the other hand, it leads less efficient firms from country  $i$  to start exporting to  $j$ , and such firms charge lower markups. When firm-level productivity is distributed Pareto, the second effect exactly offsets the first one so that the markup distribution is not affected.<sup>11</sup> Yet the entry of the less efficient firms is irrelevant from a welfare standpoint, which explains why the invariance of the markup distribution does not preclude trade liberalization to have different welfare implications in models with constant and variable markups. Formally,  $\eta$  could be positive or negative in Proposition 1.

The economic forces behind our welfare results are deeper and echo the two quotes from Helpman and Krugman (1989) given at the beginning of our Introduction. First, the existence of variable markups affects a country’s terms-of-trade effects by changing how trade cost shocks get passed through from foreign firms to domestic consumers. This is reflected in  $(-\rho) \sum_{i \neq j} \lambda_{ij} (d \ln c_{ij} - d \ln P_j)$  in equation (19), which captures both the direct and indirect effects on foreign markups. Second, the existence of variable markups implies that changes in trade costs also affect the degree of misallocation in the economy. This is reflected in  $(-\rho) \lambda_{jj} (d \ln c_{jj} - d \ln P_j)$  in equation (19), which captures the indirect (and only the indirect) effect on domestic markups. While domestic markups per se are a transfer from consumers to producers, it is a matter of simple algebra to check that under Assumption A2, changes in domestic markups,  $\rho \lambda_{jj} d \ln P_j$ , are proportional to the negative of the covariance between firm-level markups on the domestic market and changes in firm-level employment shares for that market; see Appendix A.4. Thus whenever do-

<sup>10</sup>As Mrazova and Neary (2013b) note, this condition is sometimes called “Marshall’s Second Law of Demand,” as Marshall (1920) argued it was the normal case.

<sup>11</sup>A similar compositional effect is at play in Bernard, Eaton, Jensen, and Kortum (2003). In their model, Bertrand competition leads to variable markups at the firm-level, but distributional assumptions similar to ours make the univariate distribution of markups invariant to changes in trade costs.

mestic markups go down on average, workers get reallocated towards firms with higher markups. Since their goods are under-supplied in the initial equilibrium, this increases welfare above and beyond what a model with constant markups would have predicted.<sup>12</sup>

At this point, it should therefore be clear that our theoretical analysis is perfectly consistent with a scenario in which after trade liberalization: (i) the least efficient domestic firms exit; (ii) domestic firms that stay in the industry reduce their markups; and yet (iii) welfare gains from trade liberalization are lower than those predicted by a simple trade model with constant markups and no firm heterogeneity like [Krugman \(1980\)](#). The underlying economics are simple: the exit of the least efficient firms has no first-order welfare effects; the decrease in domestic markups raises welfare by reducing distortions on the domestic market; but the welfare consequences of trade liberalization also depend on changes in foreign markups, which tend to push welfare in the opposite direction.

**The Role of Non-Homotheticity in Preferences.** A corollary of Proposition 1 is that if preferences are homothetic, which corresponds to  $\beta = 1$ , then  $\eta = 0$ . In this case, the direct and indirect markup effects exactly compensate one another, implying that welfare changes are equal to those predicted by models with constant markups considered in ACR. Intuitively, a good trade shock in an open economy is like a positive income shock in a closed economy. If preferences are homothetic, such a shock does not affect how domestic consumers allocate their expenditures across goods and, in turn, has no additional welfare effects even if the economy is distorted. In contrast, if preferences are non-homothetic, a positive income shock may additionally lower welfare in a distorted economy if it triggers a reallocation towards goods that have lower markups. This is what happens if  $\rho > 0$  and  $\beta = 0$ .<sup>13</sup>

Under the assumption that preferences are homothetic, it is worth noting that the equivalence between models with variable and constant markups extends beyond small changes in trade costs. Homotheticity in preferences implies that consumers that are subject to an income shock equivalent to the trade shock still consume goods in the exact same proportions as consumers that are not. In order to compute the equivalent variation associated with an arbitrary change in trade costs from  $\tau$  to  $\tau'$ , we can therefore integrate the expression given in Proposition 1 between the initial and final equilibria. Formally, if Assumptions A1 and A2 hold and  $\beta = 1$ , then the equivalent variation associated with

---

<sup>12</sup>The fact that changes in the degree of misallocation should be picked up by the covariance between markups and changes in factor share is not specific to the particular model that we consider; see [Basu and Fernald \(2002\)](#) for a general discussion.

<sup>13</sup>In this case, one can show that  $d \ln P_j / d \ln w_j > 0$ . Thus, the covariance between firm-level markups and log-changes in firm-level employment shares caused by the positive income shock is negative.

any trade shock in country  $j$  is given by

$$\hat{W}_j = (\hat{\lambda}_{jj})^{-1/\theta},$$

where  $\hat{\lambda}_{jj} \equiv \lambda'_{jj}/\lambda_{jj}$  denotes the proportional change in the share of expenditure on domestic goods caused by the trade shock. This is the exact same expression for large welfare changes as in ACR.

Although the set of models with homothetic preferences considered in this paper is rich enough to rationalize any cross-sectional distribution of markups—by appropriately choosing the demand function  $D(\cdot)$  that enters equation (5)—any model within that set would predict the same welfare gains from trade liberalization as in ACR, regardless of whether trade shocks are small or not.

**Relationship to Krugman (1979).** To conclude this discussion, let us briefly come back to the existing literature and clarify the relationship between our theoretical results and the seminal work of Krugman (1979). While the demand system described in equation (1) nests the case of additively separable utility functions considered in Krugman (1979), our analysis differs from his in three dimensions. First, we impose the existence of a choke price. Second, we assume that firms are heterogeneous in their productivity. Third, we focus on changes in iceberg trade costs, whereas he focuses on changes in market size. The last two differences have strong implications for the nature of distortions in the class of models that we analyze compared to his.

In models of monopolistic competition with homogeneous firms and no trade costs, such as the one considered in Krugman (1979), the level of the markups may change with the size of the market, but they are always common across goods in a given equilibrium. Thus markups are not a source of inefficiency. The only distortion in the economy is that there may be too many or too few goods produced in equilibrium, which changes in country size may exacerbate or not. In contrast, because of Assumption A2, the mass of entrants  $N$  is invariant to changes in trade costs in the models that we consider. The only distortion here is that markups vary across goods from the same country.<sup>14</sup> Thus, our theoretical results have little to say about whether gains from changes in trade costs in Krugman (1979) are bigger or smaller than those in Krugman (1980) in some well-defined sense that would remain to be specified. Our focus is on the existence of variable markups at the firm-level and whether, conditional on the same observed macro data, models that feature such markups should lead us to conclude that welfare gains from

---

<sup>14</sup>Under Assumption A2, the distribution of markups in a given destination is the same across all source countries. Thus all markup distortions are “within” rather than “between” distortions; see Appendix A.4 for details.

trade liberalization are larger than previously thought. We are now ready to answer this question empirically.

## 5 Empirical Estimates

We now describe a procedure to estimate  $\eta$  in Proposition 1. The basic idea is to estimate a demand system that satisfies equations (1)-(3) and then to use the full structure of the model to go from demand estimates to  $\eta$ .

### 5.1 From Theory to Data

Our choice of demand system is motivated by the two following considerations. First, we want to nest the case of CES demand, which is by far the most common in the field. Second, we want to allow the average elasticity of markups—and hence  $\eta$ —to be positive or negative, so that data can speak to whether the existence of variable markups increases or decreases the gains from trade liberalization. In order to achieve these two goals in a parsimonious manner, we focus on the additively-separable case,  $\beta = 0$ , and impose the following parametric restrictions on  $D(\cdot)$ :

$$D(p_\omega/P) = (p_\omega/P)^{1/\gamma} - \alpha,$$

where  $\alpha$  and  $\gamma$  are the two structural parameters to be estimated.<sup>15</sup> When  $\alpha = 0$ , the previous demand system reduces to the CES case, with elasticity of substitution given by  $-1/\gamma$ . So trade liberalization has no effects on markups and  $\eta = 0$ . In contrast, when  $\alpha > 0$ , the demand elasticity is decreasing with the level consumption,  $\varepsilon'_D < 0$ , which implies  $\rho > 0$  and  $\eta = \rho/(1 + \theta) > 0$ . In this case, variable markups lower the welfare gains from trade liberalization, as described in Proposition 1. The opposite happens if  $\alpha < 0$ .

To estimate this demand system, we follow a large literature that uses detailed data on bilateral U.S. imports within narrowly defined product codes to estimate the representative U.S. consumer's demand parameters; see e.g. Broda and Weinstein (2006) and Feenstra and Weinstein (2010). The best available data is at the 10-digit HS level, annually from 1989-2005.<sup>16</sup> In mapping these data to our model we assume that a variety  $\omega$  in the

<sup>15</sup>Mrazova and Neary (2013a) refer to this demand system as the “Pollak family”. Simonovska (2009) uses the log-version to analyze the relationship between income and prices across countries.

<sup>16</sup>We download this data from Peter Schott's homepage and use the concordances provided in Pierce and Schott (2009) to adjust for changes in 10-digit HS codes over this time period.



model corresponds to a particular 10-digit HS product, indexed by  $g$ , from a particular exporting country, indexed by  $i$ ; that is, a “variety”  $\omega$  in the model is a “product-country” pair  $gi$  in the data. Because the demand system in equation (4) is intended to represent demand for varieties within a differentiated industry, we assume that an “industry” in the data is a level of product aggregation that is higher than the 10-digit level; our baseline estimates use 4-digit HS categories as industries, which we index by  $k$ . In what follows, we let the price aggregator  $P_t^k$  vary across industries and over time, but restrict the demand parameters  $\alpha$  and  $\gamma$  to be common across all industries.<sup>17</sup>

In practice, we therefore focus on the following demand equation:

$$q_{git}^k = \left( \varepsilon_{git}^k p_{git}^k / P_t^k \right)^{1/\gamma} - \alpha, \quad (25)$$

where  $p_{git}^k$  is the price paid by U.S. consumers when buying quantity  $q_{git}^k$  for a narrowly defined product  $g$  in industry  $k$  from an exporting country  $i$  in year  $t$ . The import data contain measures of total expenditure, i.e., the empirical analogue of  $q_{git}^k \times p_{git}^k$ , and measures of total quantities purchased, which we take as our measure of  $q_{git}^k$ . To construct a measure of prices  $p_{git}^k$  we therefore simply take the ratio of expenditure to quantity. The variety-specific demand shifter,  $\varepsilon_{git}^k$ , captures the fact that physical units in the data may differ from the choice of units in Section 2, under which all varieties are implicitly assumed to enter utility in a symmetric fashion. Such differences in units of account can be interpreted as unobserved quality differences; see e.g. [Baldwin and Harrigan \(2007\)](#).

## 5.2 Estimation Procedure

There are two key challenges involved in estimating equation (25): (i) the price aggregator  $P_t^k$  is unobserved and endogenous; and (ii) the demand shifter  $\varepsilon_{git}^k$  is unobserved and endogenous. We describe below, in turn, a procedure to estimate the demand parameters,  $\alpha$  and  $\gamma$ , that overcomes these challenges.

First, consider the problem that the price aggregator  $P_t^k$  is unobserved and endogenous. The key restriction imposed in equation (25), however, is that the demand for all varieties depends symmetrically on this aggregator; that is, the price aggregator does not vary across products  $g$  and exporters  $i$  within an industry  $k$ . This suggests that iden-

---

<sup>17</sup>In the one-sector case, we have already pointed out that if  $\beta = 0$ ,  $P$  corresponds to the inverse of the lagrange multiplier on the budget constraint. In the multiple-sector case, if one assumes two-stage budgeting,  $P^k$  can be interpreted in a similar manner as the inverse of the lagrange multiplier associated with the constraint,  $\int_{\omega \in \Omega^k} \left[ p_{\omega t}^k D \left( p_{\omega t}^k / P \right) \right] d\omega = w_t^k$ , where  $w_t^k$  denotes total expenditure on varieties from industry  $k$  at date  $t$ .



tification of the demand parameters,  $\alpha$  and  $\gamma$ , can be achieved through a differencing procedure designed to eliminate the unobserved and endogenous  $P_t^k$  term in equation (25). Specifically, inverting our demand function and taking logs, we get

$$\ln p_{git}^k = \gamma \ln(q_{git}^k + \alpha) - \ln P_t^k + \ln \varepsilon_{git}^k.$$

Taking differences with respect to one reference good-country within the same industry  $k$ , we then obtain

$$\Delta_{gi} \ln p_{git}^k = \gamma \Delta_{gi} \ln(q_{git}^k + \alpha) + \Delta_{gi} \ln \varepsilon_{git}^k, \quad (26)$$

where  $\Delta_{gi}$  denotes the corresponding difference operator. While in principle the difference  $\Delta_{gi}$  could be taken across any two product-country  $gi$  observations within an industry-year  $kt$ , we use the convention of mean differencing such that, for any variable  $Y$ ,  $\Delta_{gi} Y_{git}^k = Y_{git}^k - \frac{1}{N_{kt}} \sum_{gi \in \mathcal{I}_{kt}} Y_{git}^k$  where  $\mathcal{I}_{kt}$  is the set of products  $g$  and countries  $i$  in industry  $k$  and year  $t$  and  $N_{kt}$  is the number of observations in this set.

Second, consider the problem posed by the endogeneity of the unobserved demand-shifter,  $\varepsilon_{git}^k$ . We first follow the literature on demand system estimation using international trade data—e.g. [Broda and Weinstein \(2006\)](#) and [Feenstra and Weinstein \(2010\)](#)—and decompose this demand-shifter into two terms:

$$\ln \varepsilon_{git}^k = \ln \delta_{gi}^k + \ln \epsilon_{git}^k.$$

In this decomposition, the first term,  $\ln \delta_{gi}^k$ , reflects systematic differences in quality across products from different countries within an industry, whereas the second term,  $\ln \epsilon_{git}^k$ , reflects idiosyncratic determinants of demand that are free to vary over time. To eliminate systematic unobserved differences in quality, we take a second difference of equation (26), now across time periods, to obtain

$$\Delta_t \Delta_{gi} \ln p_{git}^k = \gamma \Delta_t \Delta_{gi} \ln(q_{git}^k + \alpha) + \Delta_t \Delta_{gi} \ln \epsilon_{git}^k, \quad (27)$$

where  $\Delta_t$  denotes the corresponding difference operator. Again, while the difference  $\Delta_t$  could be taken across any two time periods we use mean differencing, as in  $\Delta_{gi}$  defined above. While this double-differencing procedure may mitigate cross-sectional sources of bias due to unobserved quality shifters, standard simultaneity bias concerns remain. As in virtually any demand estimation context, available data on price and quantity are obtained from an observed equilibrium between the supply and demand sides of a market. This codetermination of prices and quantities means that OLS estimates of equation (27) would be biased. A natural solution is to use an instrumental variable (IV) ap-

proach, where here the instrument must be exogenous with respect to demand shifters, i.e.  $\Delta_t \Delta_{gc} \ln \epsilon_{git}^k$ , and must be correlated with the endogenous variable, i.e. the double-demeaned quantity  $\Delta_t \Delta_{gi} \ln(q_{git}^k + \alpha)$ , for any value of  $\alpha$ . In our model a natural candidate for such an instrument is trade costs. For this purpose we use the (log of one plus the) value of tariff duties charged, expressed as a percentage of import value, as a measure of trade costs; this variable is reported in the US 10-digit HS imports data. This procedure of using trade costs as instruments to estimate demand-side parameters in an international trade setting is commonly employed in the empirical gravity literature; see e.g. [Head and Mayer \(2013\)](#).

Since the estimating equation (27) is linear in  $\gamma$ , but non-linear in  $\alpha$ , we separate our estimation procedure into an inner-loop and an outer-loop. In the inner-loop, we take the value of  $\alpha$  as given and compute  $\hat{\gamma}(\alpha)$  as the IV estimator of  $\gamma$  with  $\Delta_t \Delta_{gi} \ln(t_{git}^k + \alpha)$  the instrumental variable for  $\Delta_t \Delta_{gi} \ln(q_{git}^k + \alpha)$ , where  $t_{git}^k$  denotes the tariff rate charged by the United States on imports of product  $g$  in industry  $k$  from country  $i$  in year  $t$ . In the outer-loop, we then search for  $\hat{\alpha}$  that minimizes the sum of the squared residuals across all linear IV regressions. Our estimator of  $\gamma$  is finally given by  $\hat{\gamma} = \hat{\gamma}(\hat{\alpha})$ .

### 5.3 Demand Estimates

We begin by estimating the demand system in equation (27) under the restriction that  $\alpha = 0$ . This reduces equation (27) to the CES case, in which the estimating equation is linear. Our results are reported in Panel A of Table 1. Our IV estimate is  $\hat{\gamma} = -0.221$  with a standard error—clustered at the exporting country level to account for serial correlation over time and across products within exporters—that suggests that  $\gamma$  is statistically significantly different from zero at the 95% confidence level. This finding implies an elasticity of substitution equal to  $1/\hat{\gamma} = -4.53$ , which is in line with typical estimates of the CES demand parameter in international trade settings. This suggests that our particular instrumental variable, based on the reported value of tariff duties charged, is generating the same exogenous variation in trade costs that is typically exploited by other researchers.

We then estimate equation (27) without any restriction on  $\alpha$  beyond the fact that  $q_{git}^k + \alpha$  must be strictly positive for  $\ln(q_{git}^k + \alpha)$  to be well-defined. Namely, we require  $\alpha$  to be greater than minus the lowest value of  $q_{git}^k$  in our dataset, which is equal to 1 in all years and industries. These results are reported in Panel B of Table 1. Our non-linear IV estimate of equation (25) results in—with 95% confidence intervals, block-bootstrapped at the exporting country level, with 200 bootstrap replications, shown in parentheses—estimates of  $\hat{\gamma} = -0.253$  and  $\hat{\alpha} = 1.5$ . Notably, this estimate of  $\alpha$  has a 95% confidence

	$\gamma$	$\alpha$
Panel A: CES demand	-0.221*** (0.051)	
Panel B: Generalized CES demand	-0.253*** [0.095, 0.408]	1.5*** [0.9, 2.1]

Table 1: Demand Estimates. Panel A reports IV estimates of equation (27) with  $\alpha = 0$  and standard errors clustered at the exporter level. Panel B reports IV estimates of equation (27) without restrictions and with 95 percent confidence intervals from a block-bootstrap procedure, with blocks at the exporter level. \*\*\* indicates  $p < 0.05$ .

interval that excludes zero, suggesting that the departure from CES that is modeled in equation (25) is a real feature of these data. Furthermore,  $\hat{\alpha}$  is positive. As argued above, this implies that  $\eta$  must be positive as well. So, regardless of the value of other structural parameters, Proposition 1 establishes that there cannot be any pro-competitive effect of trade in the sense that welfare gains from trade liberalization must be lower than those predicted by a model with constant markups. Next, we determine how much lower they must be.

## 5.4 Welfare Implications

By definition,  $\eta = \rho ((1 - \beta)/(1 - \beta + \theta))$  with  $\beta \in \{0, 1\}$ . The demand system introduced in equation (25) imposes  $\beta = 0$ . So, in order to compute  $\eta$ , we only need estimates of  $\theta$  and  $\rho$ . In our model,  $\theta$  is equal to the elasticity of trade flows with respect to trade costs. We therefore use  $\theta = 5$ , which is in line with recent estimates of this parameter—e.g. Eaton, Kortum, and Kramarz (2011), Simonovska and Waugh (2011), and Costinot, Donaldson, and Komunjer (2012)—and is equal to the median estimate in the meta-analysis of gravity-based estimates of trade elasticities in Head and Mayer (2013). Using our estimates of  $\hat{\alpha}$ ,  $\hat{\gamma}$ , and  $\theta$ , we can then use equations (5) and (18) to compute the average markup elasticity,  $\hat{\rho} = 0.24$  and in turn  $\hat{\eta} = \hat{\rho}/(1 + \theta) = 0.04$ . Thus, micro-level trade data lead us to conclude that gains from trade liberalization are 4% lower than what we would have predicted by assuming (wrongly) that markups are constant across firms.

One potential concern regarding the previous number is that we have let the data speak, but not enough. For instance, we have assumed  $\beta = 0$ . We have not tried to estimate whether  $\beta = 0$  or 1 (or another value for some broader class of demand functions). We note, however, that even in the case  $\beta = 1$ , there are no pro-competitive effects of

trade. Under homotheticity, gains from trade for all demand systems considered in this paper are given by the welfare formula in ACR. More generally, one may want to explore how departures from our baseline assumptions, like the absence of fixed costs or Pareto distributions of productivity, affect the previous welfare estimates. This is what we will turn to in the next section.

Another potential concern is that the source of variation used to estimate  $\rho$  and hence  $\eta$  relies too much on the particular structure of the model. Economically speaking,  $\rho$  measures how, on average, changes in marginal costs map into changes in markups. Under monopolistic competition,  $\rho$  can be inferred by using information about the shape of demand and the distribution of firm-level sales. But one may imagine instead measuring this elasticity directly. Given the static nature of our model, this elasticity would have to be long-run. One possibility would be to analyze how markups vary with productivity in a cross-section of firms. The empirical work of [Goldberg, Loecker, Khandelwal, and Pavcnik \(2012\)](#) in India provides a useful benchmark. When running a cross-sectional regression of (log) prices on (log) marginal cost, they find a “pass-through” coefficient of 0.35. For a given firm in our model, the pass-through coefficient is equal to one minus the markup elasticity. Ignoring heterogeneity in markup elasticities across firms, this alternative estimation strategy would lead to  $\hat{\rho} = 0.65$ . Since  $\hat{\rho} > 0$ , gains from trade liberalization are again strictly lower than in the CES benchmark. Using the same trade elasticity ( $\theta = 5$ ), we now get  $\hat{\eta} = 0.11$ .<sup>18</sup> Pro-competitive effects, in the sense of Proposition 1, remain elusive.

## 6 Sensitivity Analysis

We have designed our baseline analysis with two objectives in mind: (i) generate the same aggregate predictions across models with and without variable markups; and (ii) abstract from welfare gains from new varieties. While this provides a clear benchmark to study the welfare implications of variable markups, conditions (i) and (ii) rely on strong assumptions. The goal of this final section is to relax these assumptions and explore the robustness of our earlier conclusions. Namely, we allow for changes in trade costs that are not infinitesimal, for distributions of productivity that are not Pareto, and for fixed marketing costs that are not zero.

---

<sup>18</sup>[Goldberg, Loecker, Khandelwal, and Pavcnik \(2012\)](#) also run a regression of (log) price on (log) marginal cost with firm fixed effects, so that the identification comes from the variation over time in their panel. This yields a lower pass-through coefficient of 0.2, which would imply  $\hat{\rho} = 0.8$  and  $\hat{\eta} = 0.13$ .

## 6.1 Calibrated Economy

To analyze welfare changes in these more general environments, we rely on numerical simulations. We focus on a world economy comprising two symmetric countries. We set country size to  $L = 1$  and fixed entry costs to  $F = 1$ . This affects welfare levels in the initial equilibrium—by affecting the number of firms—but not the welfare changes that we are interested in. In all simulations, we use the demand system estimated in Section 5 with  $\alpha = 1.5$  and  $\gamma = -0.253$ . Finally, we set trade costs and parameters of the firm-level productivity distributions to match the U.S. imports to expenditure ratio, the trade elasticity and, in the case of lognormal and bounded Pareto distributions, the share of U.S. firms exporting. The values of all calibrated parameters can be found in Table 2. For the baseline calibration with Pareto distribution this calibration implies a choice of the Pareto elasticity of  $\theta = 5$ .

Before turning to our counterfactual exercises, we briefly discuss the positive implications of our calibrated model. In the previous literature, a number of models with CES demand have been constructed to match salient features of firm-level data, including the distribution of exporting sales and the difference in measured productivity between exporters and non-exporters. Since our demand estimates have led us to depart from CES, it is natural to ask how well our calibrated model performs along these two dimensions.

Figure 1 depicts the sales distribution of French firms—normalized by mean sales—across markets for the 5th, 25th, 50th, 75th, and 95th percentiles. As first documented by Eaton, Kortum, and Kramarz (2004), the data reveal a striking heterogeneity: compared to the mean exporter, the exporters at the 5th, 50th and 95th percentiles sell less than 1/100th, around 1/10th, and 3 times as much, respectively.

The predictions of our calibrated model are plotted on the same figure (black line). For comparison, we also plot the predictions our model if we had assumed a CES demand instead (red line).<sup>19</sup> In both cases, we use the same Pareto elasticity ( $\theta = 5$ ). The two models predict well the distribution of sales for the the largest exporters. Intuitively, our estimated demand function asymptotically resembles a CES function. So, given Pareto distributions of productivity, both models predict a Pareto distribution of sales in the right tail. Interestingly, our calibrated model is also able to capture the heterogeneity between the largest and smallest exporters, whereas the model with CES demand does not. One can improve the fit of the CES model by introducing demand shocks and fixed marketing costs, as in Eaton, Kortum, and Kramarz (2011) and Arkolakis (2010). The fit of our model for the firm-level distribution of sales is as good as the fit of these richer

---

<sup>19</sup>In the CES case, we use the estimates of demand in Panel A of Table 1,  $\alpha = 0$  and  $\gamma = -0.221$ . This implies an elasticity of substitution equal to 4.5.

Parameter	Value	Target/Choice Calibration
Panel A: Demand (Sections 6.2-6.4)		
$\alpha$	1.5	Baseline estimate (Table 1, Panel B)
$\gamma$	-0.253	Baseline estimate (Table 1, Panel B)
Panel B: Pareto productivity distribution (Sections 6.2 and 6.4)		
$\theta$	5	Trade elasticity (Head and Mayer (2013))
$\tau$	1.68	Imports/expenditure = 7% (OECD Input-Output Database, 2000)
Panel C: Lognormal productivity distribution (Section 6.3)		
$\tau$	1.82	Targets for all the three parameters:
$\mu_l$	-1.52	(i) trade elasticity = 5, (ii) imports/expenditure = 7%,
$\sigma_l$	0.475	and (iii) share of firms exporting = 18% (BJRS, 2007)
Panel D: Bounded Pareto productivity distribution (Section 6.3)		
$\tau$	1.845	Targets for all the three parameters:
$\theta$	2.475	(i) trade elasticity = 5, (ii) imports/expenditure = 7%,
$\bar{z}_i^\mu$	0.54	and (iii) share of firms exporting = 18% (BJRS, 2007)

Table 2: Calibration procedure. Procedure for model parameter calibration throughout Section 6. BJRS (2007) refers to Bernard, Jensen, Redding, and Schott (2007).

models.

Another striking feature of firm-level data is the difference between the measured productivity of exporters and non-exporters. Bernard, Eaton, Jensen, and Kortum (2003) report that the relative advantage of US exporters to non-exporters in log-productivity is 33% overall and 15% within the same industry. Measured productivity in Bernard, Eaton, Jensen, and Kortum (2003) corresponds to the sum of revenues divided by the sum of labor payments,  $\sum_j r_{ij}(z) / (\sum_j w_i \tau_{ij} q_{ij}(z) / z)$ . For domestic firms in our model, this ratio is equal to their markups while for exports it is a weighted average of the domestic and the foreign markups. At the calibrated parameters, we find that that the exporter's advantage is 16%, very close to the 15% observed within industries in the data. Absent any fixed cost of production, of course, the same model with CES demand would predict no variation in markups and hence no variation in measured productivity across firms.

Having established the ability of our model to match key moments of firm-level data—moments that we did not target in our baseline calibration—we now turn to the welfare implications of our calibrated model.

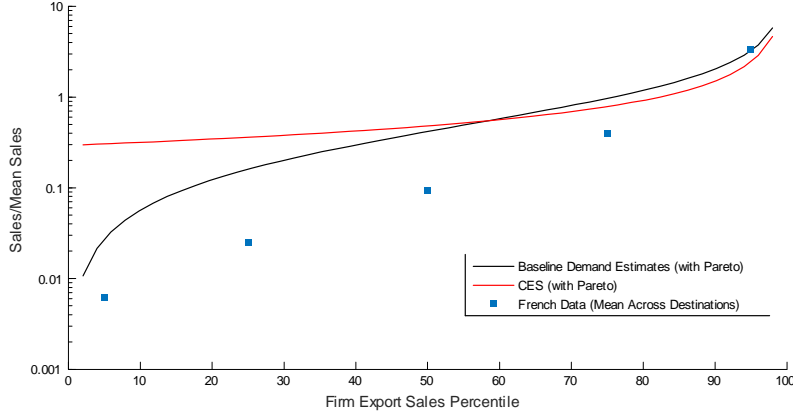


Figure 1: Distribution of Firm Export Sales. Source: [Eaton, Kortum, and Kramarz \(2011\)](#).

## 6.2 Large Changes in Trade Costs

For our first series of numerical exercises, we maintain the exact same assumptions as in our baseline analysis, but consider large changes in trade costs. Namely, we let symmetric iceberg trade costs,  $\tau$ , vary from twenty percent below to twenty percent above the calibrated value,  $\tau = 1.68$ .

To understand why large changes may affect our earlier conclusions, let us return to the expenditure minimization problem in country  $j$ . Under the restrictions imposed on demand in Section 5, one can check that the expenditure function is given by

$$e_j = \min_{\{q_{ij}(z)\}} \sum_i N_i \int_{z_{ij}^*} p_{ij}(z) q_{ij}(z) dG_i(z) \quad (28)$$

$$\sum_i N_i \int_{z_{ij}^*} u_{ij}(q_{ij}(z)) dG_i(z) dz \geq \bar{u},$$

with  $u_{ij}(q) = (q + \alpha)^{1+\gamma}$ . The Envelope Theorem then implies that

$$d \ln e_j = \sum_i \frac{N_i \int_{z_{ij}^*} [p_{ij}(z) q_{ij}(z, u) - \zeta u_{ij}(q_{ij}(z, u))] \lambda dG_i(z) dz}{e_j} d \ln N_i \quad (29)$$

$$- \sum_i \frac{N_i [p_{ij}(z_{ij}^*) q_{ij}(z_{ij}^*, u) - \zeta u_{ij}(q_{ij}(z_{ij}^*, u))] g_i(q_{ij}(z_{ij}^*))}{e_j} dz_{ij}^*$$

$$+ \sum_i \frac{N_i \int_{z_{ij}^*} [p_{ij}(z) q_{ij}(z) d \ln p_{ij}(z)] dG_i(z)}{e_j},$$



where  $\zeta$  is the Lagrange multiplier associated with the utility constraint and  $q_{ij}(z, \bar{u})$  is the compensated (Hicksian) demand. The first term in equation (29) corresponds to the total surplus associated with a change in the measure of varieties from country  $i$ ; the second term corresponds to the surplus associated with cut-off varieties; and the third term measures the effects of changes in the prices of existing varieties, either through changes in marginal costs or markups. This last term is the only one that is non-zero in our baseline analysis.

When productivity distributions are Pareto, the number of entrants is fixed by country size. So, the first term must always be equal to zero, regardless of whether changes in trade costs are large or small. Away from the initial equilibrium, however, the second term may not be. Although the consumer in the decentralized equilibrium would never consume the cut-off variety, the consumer whose utility has been held at some constant level  $\bar{u}$  may very well choose to do so. Put differently, non-homotheticities imply that gains and losses from cut-off varieties, which the formula in Proposition 1 ignores, may no longer be zero as one goes from small to large changes in trade costs.

To assess the importance of these considerations, we compute the equivalent variation associated with an arbitrary change in trade costs given by the expenditure function in (28). We refer to this number, expressed as a fraction of country's initial income, as the exact welfare change.<sup>20</sup> We then compare this number to the welfare change that one would obtain by integrating the welfare formula in Proposition 1, i.e.  $(\lambda'_{jj}/\lambda_{jj})^{-\frac{1-\eta}{\theta}} - 1$ , with  $\lambda'_{jj}$  the share of expenditure on domestic goods in the equilibrium with the new trade costs, as computed in Appendix A.3. Figure 2 plots the exact welfare changes (bold line) and the welfare changes obtained using our new formula with  $\eta = 0.04$  (dotted line) as a function of iceberg trade costs,  $\tau$ . The two curves almost coincide. For completeness, we also report the welfare changes one would obtain by using ACR's welfare formula, i.e.  $\eta = 0$  (dashed line). We see that the formula in Proposition 1, which holds exactly for small changes in trade costs, also provides an accurate approximation to the case of large changes. In this numerical example, the impact of cut-off varieties on the welfare implications of trade liberalization is minor.

---

<sup>20</sup>Formally, the exact welfare changes in country  $j$  are computed as  $e(\mathbf{p}_j, u'_j)/w_j - 1$ , with  $\mathbf{p}_j$  and  $w_j$  the schedule of good prices and the wage in the initial equilibrium, respectively, and  $u'_j$  the utility level in the counterfactual equilibrium.



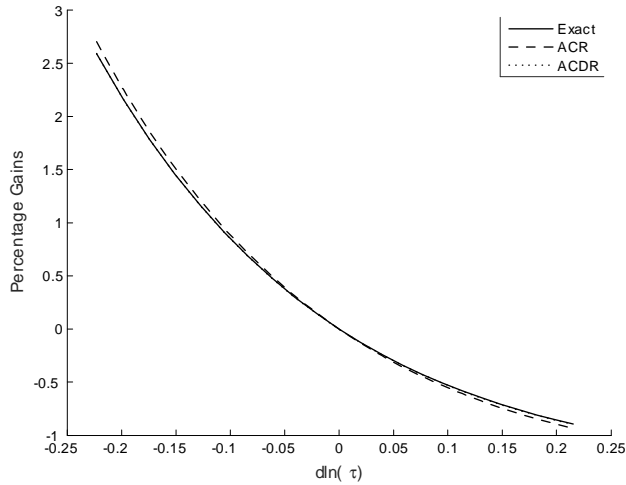


Figure 2: Welfare Gains Relative to Baseline ( $\tau = 1.68$ ), Pareto

### 6.3 Alternative Productivity Distributions

In our baseline analysis, we have assumed that the distribution of firm-level productivity was Pareto. This implies a gravity equation, which facilitates comparisons with earlier work, but it also implies that the univariate distribution of markups is invariant to changes in trade costs. The previous observation notwithstanding, it is not a priori obvious how departing from Pareto should affect our welfare results. First, as already discussed earlier, the fact that the univariate distribution of markups is fixed does not imply that the level of distortions is. The latter depends on the joint distribution of markups and expenditure shares. Second, since the number of entrants depends on the profitability of firms, any increase in average markups—which would lower the third term in equation (29)—should be accompanied by an increase in the number of entrants—which would increase the first term in the same equation.

The CES case nicely illustrates the potential importance of offsetting effects when studying aggregate welfare changes. Away from Pareto, we know that changes in trade costs not only affect the share of expenditure on domestic goods, but also the number of entrants in a given country. Yet, because the allocation is efficient under CES, we know from the work of [Atkeson and Burstein \(2010\)](#) that

$$d \ln e_j = (1 - \lambda_{jj}) d \ln \tau. \quad (30)$$

In a two-country symmetric economy, the formal definition of the trade elasticity in ACR reduces to  $\varepsilon = d \ln((1 - \lambda_{jj}) / \lambda_{jj}) / d \ln \tau$ . Using this definition and changing variable in

the previous equation, one therefore gets

$$d \ln e_j = d \ln \lambda_{jj} / \varepsilon.$$

In this CES example, the local version of ACR formula always holds, regardless of distributional assumptions and regardless of whether the number of entrants varies.

Without CES, and hence without efficiency, the situation is more subtle. To explore how our welfare results are affected by departures from Pareto under our estimated demand system, we focus on the two alternatives that have recently received attention in the literature: (i) log-normal distributions with mean  $\mu_l$  and standard deviation  $\sigma_l$ , as in [Head, Mayer, and Thoenig \(2014\)](#); and (ii) bounded Pareto distributions with shape parameter  $\theta$  and upper-bound  $\bar{z}_i^u$ , as in [Feenstra \(2014\)](#). The calibrated values of these parameters are reported in [Table 2](#). As discussed earlier, we set these parameters, together with the baseline iceberg trade cost, to target the U.S. imports to expenditure ratio, the trade elasticity and the share of U.S. firms exporting. Since the trade elasticity is no longer constant, we target its value for a 1% change in trade costs around the calibration point using the formal definition in ACR, applied to the case of two symmetric countries:  $\varepsilon = d \ln((1 - \lambda_{jj}) / \lambda_{jj}) / d \ln \tau$ .<sup>21</sup>

We then follow the same procedure as in [Section 6.2](#). We compute the exact welfare changes using the expenditure function in [\(28\)](#)—with the distribution  $G_i$  being either log-normal or bounded Pareto—and we compare those to the welfare changes that one would obtain by integrating our new welfare formula or the ACR formula. These results are reported in [Figures 3 and 4](#).<sup>22</sup> In both cases, we see that our formulae over-estimate both the gains from trade liberalization and the losses from trade protection.

The interpretation of these numerical results is less straightforward than before. As we go from Pareto distributions to other distributions, we not only change the extent of firm-level distortions, but also the aggregate predictions of the model. Although we still target the same trade elasticity in the initial equilibrium, it now varies with the level of the trade of costs, a point emphasized by [Head, Mayer, and Thoenig \(2014\)](#) and [Melitz and Redding \(2015\)](#) in the CES case. More precisely, the trade elasticity increases in absolute value with the level of trade costs, as documented in [Appendix A.5](#). The new welfare numbers therefore reflect different behavior both at the macro and micro levels. For our purposes,

---

<sup>21</sup>For these alternative productivity distributions, we obtain predictions for the distribution of exporting sales and for the productivity advantage of exporters that are similar to those in the Pareto case. Results are available upon request.

<sup>22</sup>When integrating our new formula and the ACR formula, we let the trade elasticity and the average markup elasticity vary as variable trade costs change from their initial to their counterfactual values.

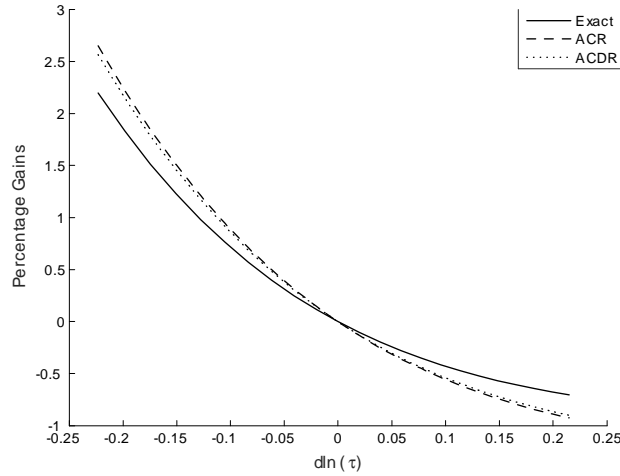


Figure 3: Welfare Gains Relative to Baseline ( $\tau = 1.82$ ), Log-normal

the important take-away from Figures 3 and 4 is that they provide very little support to the idea that welfare gains the Pareto case are special and unusually low, perhaps because the univariate distribution of markups is fixed. Under these two alternative distributional assumptions, gains from trade are lower, not larger.

## 6.4 Fixed Marketing Costs

For our last series of simulations, we introduce fixed marketing costs in our model. Such costs are potentially interesting from a welfare standpoint since they imply that creation and destruction of cut-off varieties may have first-order welfare effects, i.e. the second term in equation (29) is no longer zero, even for small changes in trade costs.

The economic environment is the same as in Section 2, except for the fact that after receiving their random productivity draws, firms must incur a fixed marketing cost,  $w_j f_j$ , in order to sell in market  $j$ . Fixed costs do not affect firm-level markups, which remain a function of relative efficiency alone, but they do affect firm-level profits. Without risk of confusion, let us drop the country indices as we did in Section 3.1. For a firm with marginal cost  $c$  and efficiency  $v$ , profits are now given by

$$\pi(c, v, Q, L) \equiv ((\mu(v) - 1) / \mu(v)) x(c, v, Q, L) - wf,$$

with firm-level sales,  $x(c, v, Q, L)$ , still given by (6). Accordingly, a firm will enter a given

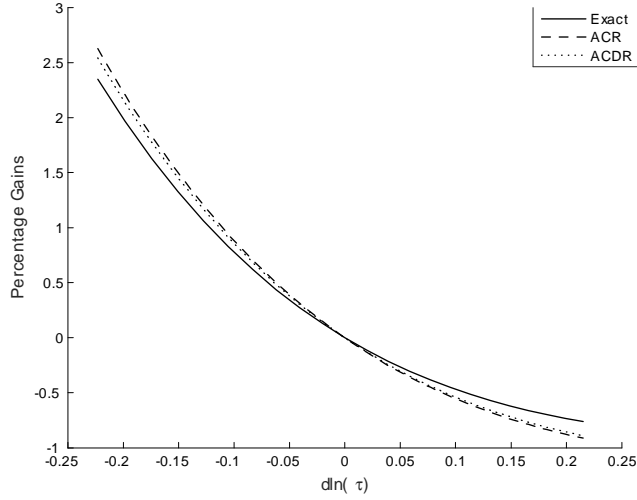


Figure 4: Welfare Gains Relative to Baseline ( $\tau = 1.845$ ), Truncated Pareto

market if and only if  $v \geq v^*$ , with  $v^*$  implicitly defined by

$$(\mu(v^*) - 1)D(\mu(v^*)/v^*) = (wf v^*)/(QLP). \quad (31)$$

When  $f = 0$ , equation (31) implies  $v^* = 1$ . So, firms only enter a market if their marginal cost,  $c$ , is below the reservation price,  $P$ . When  $f > 0$ , marginal costs must be strictly below  $P$  for firms to break even. The gravity equation (15) and the free entry condition (11) are the same as before. Only the labor market clearing condition (12) must be modified in order to take into account the resources associated with the fixed marketing costs.

To quantify the importance of fixed marketing costs, we focus on a 10% decrease in trade costs from the calibrated value,  $\tau = 1.68$ , to a counterfactual value,  $\tau = 1.51$ . We then vary the fixed marketing cost, without recalibrating other parameters, from  $f = 0$  to  $f = 3$ . Figure 5 reports the exact welfare changes together with the predictions that one would obtain by integrating our new welfare formula ( $\eta = 0.04$ ) or the ACR formula ( $\eta = 0$ ). Exact welfare changes are always bounded from above by our two formulas. As fixed costs increase, we see that both the exact welfare changes and our new formula converge towards ACR formula. This is intuitive. As fixed marketing costs increase, only the most productive firms select into a market. These firms operate in parts of the demand curve that are very close to CES. Hence, markups are close to constant across firms and welfare changes are well-approximated by the ACR formula.<sup>23</sup>

<sup>23</sup>Our estimated demand system imposes  $\beta = 0$ . In the homothetic case,  $\beta = 1$ , one can check that although the efficiency cut-off,  $v^*$ , is no longer equal to one, it remains unaffected by trade costs. Accordingly,

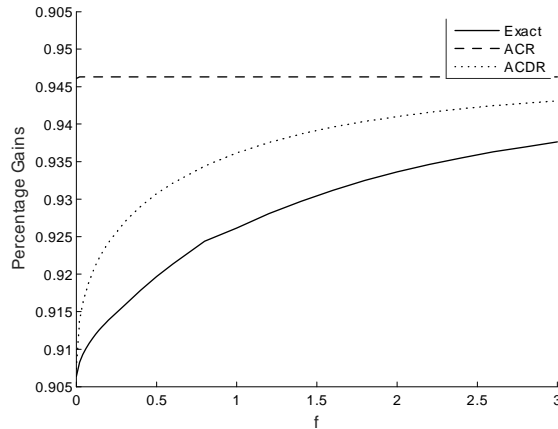


Figure 5: Welfare Gains and Fixed Costs ( $\tau = 1.68$  to  $1.51$ )

## 7 Concluding Remarks

We have studied the gains from trade liberalization in models with monopolistic competition, firm-level heterogeneity, and variable markups. Under standard restrictions on consumers' demand and the distribution of firms' productivity, we have developed a generalized version of ACR formula that highlights how micro- and macro-level considerations jointly shape the welfare gains from trade. We have then used micro-level trade data to quantify their importance. Our main finding is that (rightly) taking into account variable markups leads to gains from trade liberalization that are 4% lower than those that one would have predicted by (wrongly) assuming constant markups.

Our theoretical and empirical results, of course, only apply to a particular class of models. Monopolistic competition plays a central role in the field of international trade, but it is not the only market structure under which variable markups may arise. In a recent paper, [Edmond, Midrigan, and Xu \(2015\)](#) have studied the gains from trade liberalization in a model with Cournot competition. When calibrating their model to Taiwanese firm-level data, they find that aggregate welfare gains remain well-approximated by ACR formula. Taken together, these results suggest that if pro-competitive effects are to be found, they will require significant departures from standard assumptions about market structure and preferences. For now, pro-competitive effects of trade are elusive.

---

the distribution of markups and the number of entrants remain constant. Thus, whether fixed marketing costs are zero or not, gains from trade liberalization are given by the ACR formula.

## A Proofs

### A.1 Section 2.1

**Additively Separable Utility.** We first establish that our demand system under  $\beta = 0$  encompasses the case of additively separable utility functions considered in [Krugman \(1979\)](#). Using our notation, his model corresponds to a situation in which preferences are represented by a utility function,  $U = \int_{\omega \in \Omega} u(q_\omega) d\omega$ . The first-order conditions associated with utility maximization imply  $u'(q_\omega) = \lambda p_\omega$ , where  $\lambda$  is the Lagrangian multiplier associated with the budget constraint. Inverting the first-order conditions implies

$$q_\omega = u'^{-1}(\lambda p_\omega), \quad (32)$$

together with the budget constraint,

$$\int_{\omega \in \Omega} p_\omega q_\omega d\omega = w. \quad (33)$$

Under  $\beta = 0$ , equations (2) and (3) are equivalent to equation (33) and  $Q = 1$ , respectively. In turn, equations (1) and  $Q = 1$  imply  $q_\omega = D(p_\omega/P)$ . Thus, setting  $P \equiv 1/\lambda$  and  $D(\cdot) \equiv u'^{-1}(\cdot)$ , our demand system under  $\beta = 0$  replicates the demand system under additively separable utility.

**Kimball Preferences.** We now show that our demand system under  $\beta = 1$  encompasses the case of Kimball preferences. Under Kimball preferences, utility  $Q$  from consuming  $\{q_\omega\}_{\omega \in \Omega}$  is implicitly given by

$$\int Y\left(\frac{q_\omega}{Q}\right) d\omega = 1, \quad (34)$$

for some function  $Y$  that satisfies  $Y' > 0$  and  $Y'' < 0$ . The utility maximization program of the consumer is to  $\max_{Q, \{q_\omega\}} Q$  subject to equations (34) and (33). Let  $\gamma$  and  $\lambda$  denote the Lagrange multipliers associated with these two constraints. Manipulating the first-order conditions of this problem we get

$$q_\omega = QY'^{-1}\left(\frac{\lambda \int q_\omega Y'\left(\frac{q_\omega}{Q}\right) d\omega}{Q} p_\omega\right) \text{ for all } \omega. \quad (35)$$

The demand system under Kimball preferences is characterized by equations (33)-(35). Under  $\beta = 1$ , equations (2) and (3) are equivalent to  $\int_{\omega \in \Omega} H(p_\omega/P) d\omega = 1$  and equation (33), respectively. Thus, setting  $P \equiv Q / \left(\lambda \int q_\omega Y'\left(\frac{q_\omega}{Q}\right) d\omega\right)$ ,  $D(\cdot) \equiv Y'^{-1}(\cdot)$ , and  $H(\cdot) \equiv Y(D(\cdot))$ , our demand system with  $\beta = 1$  replicates the demand system under Kimball preferences.

**QMOR Expenditure.** Finally, we show that our demand system under  $\beta = 1$  also encompasses the demand system corresponding to QMOR expenditure functions in Feenstra (2014). The QMOR demand system entails  $q_\omega = QD(p_\omega/P)$  with

$$D(x) \equiv \begin{cases} \alpha x^{r-1} [1 - x^{-r/2}] & \text{if } x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}, \quad (36)$$

where  $P$  acts as a choke price defined implicitly by

$$P = \left( \left( \frac{N}{N - (\tilde{N} - \alpha/\gamma)} \right)^{r/2} \int_{p_\omega \leq P} \frac{1}{N} p_\omega^{r/2} d\omega \right)^{2/r}, \quad (37)$$

and where  $Q$  is determined such that the budget constraint (33) is satisfied.<sup>24</sup> In the previous expressions,  $\alpha$  and  $\gamma$  are parameters,  $\tilde{N} \equiv \int_\Omega d\omega$  is the measure of all possible goods,  $N \equiv \int_{p_\omega \leq P} d\omega$  is the measure of the set of goods with prices equal or below the choke price  $P$ . To proceed, note that equation (37) can be rearranged as

$$1 = \frac{1}{N - (\tilde{N} - \alpha/\gamma)} \int_{p_\omega \leq P} \left( \frac{p_\omega}{P} \right)^{r/2} d\omega. \quad (38)$$

To conclude, let us show that this is equivalent to equation (2) under  $\beta = 1$  if one sets

$$H\left(\frac{p_\omega}{P}\right) \equiv \frac{1}{\alpha(\alpha/\gamma - \tilde{N})} \left(\frac{p_\omega}{P}\right)^{1-r/2} D\left(\frac{p_\omega}{P}\right).$$

Together with the definition of  $D(\cdot)$  in equation 36, the previous definition implies

$$\int_\Omega H\left(\frac{p_\omega}{P}\right) d\omega = \frac{1}{\alpha/\gamma - \tilde{N}} \int_{p_\omega \leq P} \left[ \left(\frac{p_\omega}{P}\right)^{r/2} - 1 \right] d\omega.$$

Thus, as argued above,  $\int_\Omega H\left(\frac{p_\omega}{P}\right) d\omega = 1$  is equivalent to equation (38).<sup>25</sup>

**Homothetic Preferences.** In Section 2.1 we have also argued that if  $D(\cdot)$  satisfies As-

<sup>24</sup>Equations (36) and (37) are the counterparts of equations (7) and (2) in Feenstra (2014), respectively.

<sup>25</sup>Since the translog expenditure system is a special case of QMOR expenditure functions, as shown in Feenstra (2014), this establishes that our demand system encompasses the translog case. But it is useful to show directly that our demand system leads to translog demand if we set  $D(x) \equiv \zeta x^{-1} \ln x^{-1}$  for  $x \leq 1$  and  $D(x) = 0$  otherwise, with  $\zeta$  some positive constant, and  $H(x) \equiv xD(x)$ . Equation (2) with  $\beta = 1$  then implies  $\int_{p_\omega \leq P} \zeta \ln(p_\omega/P)^{-1} d\omega = 1$ , which is equivalent to

$$\ln P = \frac{1}{\zeta N} + \frac{1}{N} \int_{p_\omega \leq P} \ln p_\omega d\omega,$$

sumption A1, then consumers have homothetic preferences if and only if  $\beta = 1$ . We now establish this result formally. Throughout this proof we will repeatedly use the fact that preferences are homothetic if and only if the income elasticity,  $\partial \ln q_\omega(\mathbf{p}, w) / \partial \ln w$ , is equal to one for all goods  $\omega \in \Omega$ .

Suppose first that  $\beta = 1$ . Then equation (2) implies  $\int_{\omega \in \Omega} H(p_\omega/P) d\omega = 1$ , so  $P(\mathbf{p}, w)$  is independent of  $w$ . Differentiating equation (1), we therefore get:

$$\frac{\partial \ln q_\omega(\mathbf{p}, w)}{\partial \ln w} = \frac{\partial \ln Q(\mathbf{p}, w)}{\partial \ln w}.$$

But Equation (3) implies  $\frac{\partial \ln Q(\mathbf{p}, w)}{\partial \ln w} = 1$ , hence the income elasticity is equal to one for all goods  $\omega \in \Omega$ , so preferences are homothetic.

Now suppose that  $\beta = 0$ . In this case, equation (3) implies  $Q = 1$ . By equation (1), we therefore have

$$\frac{\partial \ln q_\omega(\mathbf{p}, w)}{\partial \ln w} = \varepsilon_D(p_\omega/P(\mathbf{p}, w)) \times \frac{\partial \ln P(\mathbf{p}, w)}{\partial \ln w}, \quad (39)$$

with

$$\varepsilon_D(p_\omega/P(\mathbf{p}, w)) \equiv \frac{p_\omega}{P(\mathbf{p}, w)} \frac{D'(p_\omega/P(\mathbf{p}, w))}{D(p_\omega/P(\mathbf{p}, w))}.$$

Now take two goods,  $\omega_1$  and  $\omega_2$ , and a price schedule and wage,  $(\mathbf{p}, w)$ , such that

$$\varepsilon_D(p_{\omega_1}/P(\mathbf{p}, w)) \neq \varepsilon_D(p_{\omega_2}/P(\mathbf{p}, w)). \quad (40)$$

Since Assumption A1 rules out CES preferences, we know that such  $\omega_1$ ,  $\omega_2$ , and  $(\mathbf{p}, w)$  must exist. By conditions (39) and (40), we cannot have

$$\frac{\partial \ln q_{\omega_1}(\mathbf{p}, w)}{\partial \ln w} = \frac{\partial \ln q_{\omega_2}(\mathbf{p}, w)}{\partial \ln w} = 1.$$

Thus preferences cannot be homothetic.

## A.2 Section 3.1

In Section 3.1 we have argued that if demand functions are log-concave in log-prices,  $\partial^2 \ln D / \partial \ln p^2 < 0$ , then  $\varepsilon_D'' > 0$  and hence  $\mu' > 0$  so that more efficient firms charge higher markups. To see this, let  $f(m, v) \equiv m - \frac{\varepsilon_D(m/v)}{\varepsilon_D(m/v)-1}$ . Equation (5) then entails

which is the condition that determines  $P$  in the translog demand; see equation (8) in Feenstra (2014). Equation (3) with  $\beta = 1$  is just the budget constraint, which given equation (2) immediately implies  $Q = w/P$ .



$f(m, v) = 0$ . Differentiating with respect to  $m$  and  $v$ , we obtain

$$\begin{aligned}\frac{\partial f(m, v)}{\partial m} &= 1 + \frac{\varepsilon'_D(m/v)}{(\varepsilon_D(m/v) - 1)^2} \frac{1}{v} > 0, \\ \frac{\partial f(m, v)}{\partial v} &= -\frac{\varepsilon'_D(m/v)}{(\varepsilon_D(m/v) - 1)^2} \frac{m}{v^2} < 0,\end{aligned}$$

where the two inequalities derive from  $\varepsilon''_D > 0$ , which follows immediately from  $\partial^2 \ln D / \partial \ln p^2 < 0$  and  $\varepsilon_D(x) \equiv -\partial \ln D(x) / \partial \ln x$ . By the Implicit Function Theorem, equation (5) therefore implies  $\mu'(v) = -(\partial f(m, v) / \partial v) / (\partial f(m, v) / \partial m) > 0$ .

### A.3 Section 3.3

In Section 3.3, we have argued that once models with variable markups considered in this paper are calibrated to match the trade elasticity  $\theta$  and the observed trade flows  $\{X_{ij}\}$ , they must predict the exact same changes in wages and trade flows for any change in variable trade costs as gravity models with CES utility, such as [Krugman \(1980\)](#), [Eaton and Kortum \(2002\)](#), [Anderson and Van Wincoop \(2003\)](#), and [Eaton, Kortum, and Kra-marz \(2011\)](#). We now establish this result formally. Let  $\lambda_{ij} \equiv X_{ij}/E_j$  denote the share of expenditure on goods from country  $i$  in country  $j$ . By equations (15) and (12), we know that

$$\begin{aligned}\lambda_{ij} &= \frac{N_i b_i^\theta (w_i \tau_{ij})^{-\theta}}{\sum_k N_k b_k^\theta (w_k \tau_{kj})^{-\theta}}, \\ w_i L_i &= \sum_j \lambda_{ij} w_j L_j.\end{aligned}$$

Since these are the same equilibrium conditions as in gravity models with CES utility (macro-level restrictions R1-R3' in ACR), we only need to show that counterfactual changes in wages and trade flows only depend on trade flows and expenditures in the initial equilibrium as well as the value of the trade elasticity.

We can use the same argument as in the proof of Proposition 2 in ACR. Consider a counterfactual change in variable trade costs from  $\tau \equiv \{\tau_{ij}\}$  to  $\tau' \equiv \{\tau'_{ij}\}$ . Let  $\hat{x} \equiv x'/x$  denote the change in any variable  $x$  between the initial and the counterfactual equilibrium. Since  $N_i$  is fixed for all  $i$ , one can show that  $\{\hat{w}_i\}_{i \neq j}$  are implicitly given by the solution of

$$\hat{w}_i = \frac{\sum_{j'=1}^n \lambda_{ij'} \hat{w}_{j'} E_{j'} (\hat{w}_i \hat{\tau}_{ij'})^{-\theta}}{E_i \sum_{j'=1}^n \lambda_{ij'} (\hat{w}_{j'} \hat{\tau}_{ij'})^{-\theta}}. \quad (41)$$

where  $\hat{w}_j = 1$  by choice of numeraire. Given changes in wages,  $\{\hat{w}_i\}$ , changes in expen-

diture shares are then given by

$$\hat{\lambda}_{ij} = \frac{(\hat{w}_i \hat{\tau}_{ij})^{-\theta}}{\sum_{i'=1}^n \lambda_{i'j} (\hat{w}_{i'} \hat{\tau}_{i'j})^{-\theta}}. \quad (42)$$

Equations (41) and (42) imply  $\{\hat{w}_i\}$  and  $\{\hat{\lambda}_{ij}\}$  only depend on the value of trade flows and expenditures in the initial equilibrium as well as the trade elasticity. Once changes in expenditure shares,  $\{\hat{\lambda}_{ij}\}$ , are known, changes in bilateral trade flows can be computed using the identity,  $\hat{X}_{ij} = \hat{\lambda}_{ij} \hat{w}_j$ . Thus the same observation applies to changes in bilateral trade flows, which concludes the argument.

#### A.4 Section 4.2

**Invariance of Distribution of Markups.** In Section 4.2, we have argued that if markups are an increasing function of firm-level productivity, then the univariate distribution of markups is independent of the level of trade costs. We now establish this result formally. Let  $M_{ij}(m; \tau)$  denote the distribution of markups set by firms from country  $i$  in country  $j$  in a trade equilibrium if trade costs are equal to  $\tau \equiv \{\tau_{ij}\}$ . Since firm-level markups only depend on the relative efficiency of firms, we can express

$$M_{ij}(m; \tau) = \Pr \{ \mu(v) \leq m | v \geq 1 \},$$

where the distribution of  $v$  depends, in principle, on the identity of both the exporting and the importing country. Recall that  $v \equiv P/c$  and  $c = c_{ij}/z$ . Thus for a firm with productivity  $z$  located in  $i$  and selling in  $j$ , we have  $v = P_j z / c_{ij} = z/z_{ij}^*$ . Combining this observation with Bayes' rule, we can rearrange the expression above as

$$M_{ij}(m; \tau) = \frac{\Pr \{ \mu(z/z_{ij}^*) \leq m, z_{ij}^* \leq z \}}{\Pr \{ z_{ij}^* \leq z \}}.$$

Using Assumption A2 and the fact that  $\mu(\cdot)$  is monotone, we can rearrange the previous expression as

$$M_{ij}(m; \tau) = \frac{\int_{z_{ij}^*}^{z_{ij}^* \mu^{-1}(m)} dG_i(z)}{\int_{z_{ij}^*}^{\infty} dG_i(z)} = 1 - \left( \mu^{-1}(m) \right)^{-\theta}.$$

Since the function  $\mu(\cdot)$  is identical across countries and independent of  $\tau$ , by equation (5), this establishes that for any exporter  $i$  and any importer  $j$ , the distribution of markups  $M_{ij}(\cdot; \tau)$  is independent of the identity of the exporter  $i$ , the identity of the importer  $j$ , and the level of trade costs  $\tau$ . As a result, the overall distribution of markups in any country  $j$  is also invariant to changes in trade costs.

**Domestic markups and Misallocation.** In Section 4.2, we have argued that changes in domestic markups,  $\rho \lambda_{jj} d \ln P_j$ , are proportional to the opposite of the covariance between firm-level markups on the domestic market and changes in firm-level employment shares for that market. We now establish this result formally.

Let us denote by  $L_{jj}(z)$  the number of workers allocated by a firm with productivity  $z$  in country  $j$  to production of goods for market  $j$ . We must have

$$L_{jj}(z) = \tau_{jj} q_{jj}(z) / z,$$

where  $q_{jj}(z)$  is such that

$$q_{jj}(z) = Q_j D \left( z_{jj}^* \mu(z/z_{jj}^*) / z \right).$$

Similarly, let us denote by  $\sigma_{jj}(z) \equiv L_{jj}(z) / L_{jj}$  denote the employment share that goes to a firm with productivity  $z$ . We have

$$\sigma_{jj}(z) = \frac{D \left( z_{jj}^* \mu(z/z_{jj}^*) / z \right) / z}{\int_{z_{jj}^*}^{\infty} N_j D \left( z_{jj}^* \mu(z'/z_{jj}^*) / z' \right) / z' dG_j(z')}.$$

Let us now compute the average of markups,  $\bar{\mu}_{jj} \equiv \int_{z_{jj}^*}^{\infty} \mu \left( z/z_{jj}^* \right) \sigma_{jj}(z) N_j dG_j(z)$ , for firms from country  $j$  selling in country  $j$  weighted by employment. We have:

$$\bar{\mu}_{jj} = \int_{z_{jj}^*}^{\infty} \mu \left( z/z_{jj}^* \right) \frac{D \left( z_{jj}^* \mu(z/z_{jj}^*) / z \right) / z}{\int_{z_{jj}^*}^{\infty} D \left( z_{jj}^* \mu(z'/z_{jj}^*) / z' \right) / z' dG_j(z')} dG_j(z).$$

Under Assumption A2, we can rearrange the previous expression as

$$\bar{\mu}_{jj} = \int_1^{\infty} \mu(v) \frac{D(\mu(v)/v) v^{-\theta-2} dv}{\int_1^{\infty} D(\mu(v')/v') (v')^{-\theta-2} dv'}.$$

This implies

$$\frac{d\bar{\mu}_{jj}}{dz_{jj}^*} = \int_{z_{jj}^*}^{\infty} \frac{d\mu \left( z/z_{jj}^* \right)}{dz_{jj}^*} \sigma_{jj}(z) N_j dG_j(z) + \int_{z_{jj}^*}^{\infty} \mu \left( z/z_{jj}^* \right) \frac{d\sigma_{jj}(z)}{dz_{jj}^*} N_j dG_j(z) = 0,$$

where we have used the fact that  $\sigma_{jj} \left( z_{jj}^* \right) = 0$ . The first term can be rearranged as

$$\int_{z_{jj}^*}^{\infty} \frac{d\mu \left( z/z_{jj}^* \right)}{dz_{jj}^*} \sigma_{jj}(z) N_j dG_j(z) = -\frac{\rho \bar{\mu}_{jj}}{z_{jj}^*}.$$

By construction,  $\int_{z_{jj}^*}^{\infty} \sigma_{jj}(z) N_j dG_j(z) = 1$ . Using again  $\sigma_{jj}(z_{jj}^*) = 0$ , we therefore have  $\int_{z_{jj}^*}^{\infty} \frac{d\sigma_{jj}(z)}{dz_{jj}^*} N_j dG_j(z) = 0$ . Thus the second term can be rearranged as

$$\int_{z_{jj}^*}^{\infty} \mu(z/z_{jj}^*) \frac{d\sigma_{jj}(z)}{dz_{jj}^*} N_j dG_j(z) = \int_{z_{jj}^*}^{\infty} \left( \mu(z/z_{jj}^*) - \bar{\mu}_{jj} \right) \left( \frac{d\sigma_{jj}(z)}{dz_{jj}^*} - 0 \right) N_j dG_j(z),$$

Combining the three previous expressions we therefore get

$$\frac{\rho \bar{\mu}_{jj}}{z_{jj}^*} = \int_{z_{jj}^*}^{\infty} \left( \mu(z/z_{jj}^*) - \bar{\mu}_{jj} \right) \left( \frac{d\sigma_{jj}(z)}{dz_{jj}^*} - 0 \right) N_j dG_j(z).$$

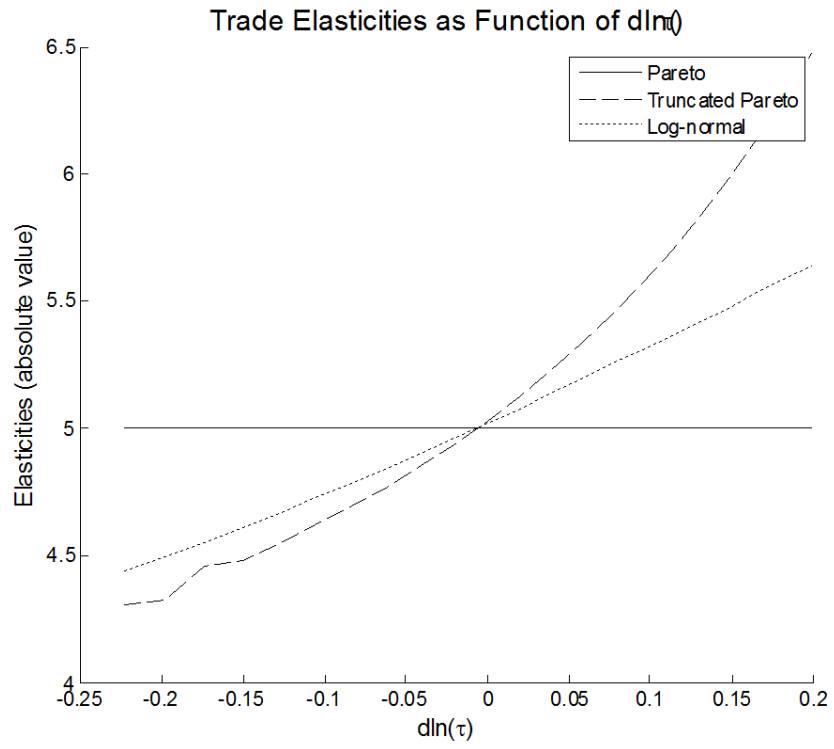
To conclude note that  $z_{jj}^* = 1/P_j$ , by our choice of numeraire. Thus the previous expression implies

$$\rho \lambda_{jj} d \ln P_j = - \left( \frac{\lambda_{jj}}{\bar{\mu}_{jj}} \right) \left( \int_{z_{jj}^*}^{\infty} \left( \mu(z/z_{jj}^*) - \bar{\mu}_{jj} \right) (d\sigma_{jj}(z) - 0) N_j dG_j(z) \right),$$

where the integral on the right-hand side is equal to the covariance between firm-level markups on the domestic market and changes in firm-level employment shares for that market.

## A.5 Section 6

All models that we consider are calibrated so that the trade elasticity for a 1% change in trade costs is equal to 5 in the initial equilibrium. Except when the distribution of productivity is Pareto, however, this elasticity will vary with the level of trade costs. Figure A.5 plots the trade elasticity as a function of trade costs in the case of Pareto, log-normal and bounded Pareto distributions. In both the log-normal and bounded Pareto cases, we see that the trade elasticity increases, in absolute value, with the level of trade costs.



## References

- ANDERSON, J. E., AND E. VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *The American Economic Review*, 93(1), 170–192.
- ARKOLAKIS, C. (2010): "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 118(6), 1151–1199.
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94–130.
- ATKESON, A., AND A. BURSTEIN (2010): "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118(3), 433–489.
- AXTELL, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293(5536), 1818–1820.
- BALDWIN, R., AND J. HARRIGAN (2007): "Zeros, Quality and Space: Trade Theory and Trade Evidence," *NBER Working paper 13214*.
- BASU, S., AND J. G. FERNALD (2002): "Aggregate Productivity and Aggregate Technology," *European Economic Review*, 46, 963–991.
- BEHRENS, K., G. MION, Y. MURATA, AND J. SUDEKUM (2009): "Trade, Wages, and Productivity," *CEP Discussion Papers*, 0942.
- BEHRENS, K., AND Y. MURATA (2009): "Globalization and Individual Gains from Trade," *CEPR discussion paper*, 7438.
- BERGIN, P. R., AND R. FEENSTRA (2009): "Pass-Through of Exchange Rates and Competition Between Floaters and Fixers," *Journal of Money Credit and Bankin*, 41(1), 35–70.
- BERNARD, A., B. JENSEN, S. REDDING, AND P. J. SCHOTT (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105–130.
- BERNARD, A. B., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268–1290.
- BHAGWATI, J. N. (1971): *The Generalized Theory of Distortions and Welfare*, vol. Trade, Balance of Payments, and Growth: Papers in International Economics in Honor of Charles P. Kindleberger. Amsterdam: North-Holland.
- BRODA, C., AND D. WEINSTEIN (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121(2), 541–585.
- BURSTEIN, A., AND G. GOPINATH (2013): "International prices and exchange rates," Discussion paper, National Bureau of Economic Research.
- CHEN, N., J. IMBS, AND A. SCOTT (2009): "The Dynamics of Trade and Competition," *Journal of International Economics*, 77(1), 50–62.

- COSTINOT, A., D. DONALDSON, AND I. KOMUNJER (2012): "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 79(2), 581–608.
- COSTINOT, A., AND A. RODRIGUEZ-CLARE (2014): "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, vol. 4, chap. 4.
- DE BLAS, B., AND K. N. RUSS (2015): "Understanding Markups in the Open Economy," *American Economic Journal: Macroeconomics*, 7(2), 157–180.
- DEATON, A., AND J. MUELLBAUER (1980): "An Almost Ideal Demand System," *American Economic Review*, 70(3), 312–326.
- DHINGRA, S., AND J. MORROW (2012): "The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition," *mimeo, LSE*.
- EATON, J., AND S. KORTUM (2002): "Technology, Geography and Trade," *Econometrica*, 70(5), 1741–1779.
- EATON, J., S. KORTUM, AND F. KRAMARZ (2004): "Dissecting Trade: Firms, Industries, and Export Destinations," *The American Economic Review, Papers and Proceedings*, 94(2), 150–154.
- (2011): "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 79(5), 1453–1498.
- EDMOND, C., V. MIDRIGAN, AND D. XU (2015): "Competition, Markups and the Gains from Trade," *American Economic Review*, forthcoming.
- EPIFANI, P., AND G. GANCIA (2011): "Trade, Markup Heterogeneity and Misallocations," *Journal of International Economics*, 83(1), 1–13.
- FEENSTRA, R. (2014): "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," *NBER working paper*, 19833.
- FEENSTRA, R. C. (2003): "A Homothetic Utility Function for Monopolistic Competition Models, Without Constant Price Elasticity," *Economics Letters*, 78(1), 79–86.
- FEENSTRA, R. C., AND D. WEINSTEIN (2010): "Globalization, Markups, and the U.S. Price Level," *NBER working paper*, 15749.
- GOLDBERG, P., J. D. LOECKER, A. KHANDELWAL, AND N. PAVCNIK (2012): "Prices, Markups and Trade Reform," *mimeo*.
- HARRISON, A. E. (1994): "Productivity, Imperfect Competition and Trade Reform. Theory and Evidence," *Journal of International Economics*, 36(1-2), 53–73.
- HEAD, K., AND T. MAYER (2013): *Gravity Equations: Toolkit, Cookbook, Workhorse*, vol. 4 of *Handbook of International Economics*. Elsevier.



- HEAD, K., T. MAYER, AND M. THOENIG (2014): "Welfare and Trade Without Pareto," *mimeo*.
- HELPMAN, E., AND P. R. KRUGMAN (1989): *Trade Policy and Market Structure*. MIT Press, Cambridge, Massachusetts.
- HOLMES, T. J., W.-T. HSU, AND S. LEE (2015): "Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade," *Journal of International Economics*, forthcoming.
- KIMBALL, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, pp. 1241–1277.
- KLENOW, P. J., AND J. L. WILLIS (2006): "Real rigidities and nominal price changes," *Available at SSRN 896325*.
- KONINGS, J., P. VAN CAYSEELE, AND F. WARZYNSKI (2001): "The Dynamics of Industrial Mark-Ups in Two Small Open Economies: Does National Competition Policy Matter?," *International Journal of Industrial Organization*, 19(5), 841–859.
- KRISHNA, P., AND D. MITRA (1998): "Trade Liberalization, Market Discipline and Productivity Growth: New Evidence from India," *Journal of Development Economics*, 56(2), 447–462.
- KRUGMAN, P. (1979): "Increasing Returns Monopolistic Competition and International Trade," *Journal of International Economics*, 9(4), 469–479.
- KRUGMAN, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *The American Economic Review*, 70(5), 950–959.
- LEVINSOHN, J. (1993): "Testing the Imports-as-Market-Discipline Hypothesis," *Journal of International Economics*, 35(1), 1–22.
- LOECKER, J. D., AND F. WARZYNSKI (2012): "Markups and Firm-Level Export Status," *forthcoming American Economic Review*.
- MARSHALL, A. (1920): *Principles of Economics, An Introductory Volume*. London: Mcmillan, eighth edition.
- MELITZ, M., AND S. REDDING (2015): "New Trade models, New Welfare Implications," *American Economic Review*, 105(3), 1105–1146.
- MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.
- MELITZ, M. J., AND G. I. P. OTTAVIANO (2008): "Market Size, Trade, and Productivity," *The Review of Economic Studies*, 75(1), 295–316.
- MRAZOVA, M., AND P. NEARY (2013a): "Not so demanding: preference structure, firm behavior, and welfare," *mimeo Oxford University*.
- (2013b): "Selection Effects with Heterogeneous Firms," *mimeo Oxford University*.

- NOVY, D. (2010): "International Trade Without CES: Estimating Translog Gravity," *CEP Discussion paper*, 1031.
- OTTAVIANO, G. I., T. TABUCHI, AND J.-F. THISSE (2002): "Agglomeration and Trade Revisited," *International Economic Review*, 43(2), 409–436.
- PIERCE, J. R., AND P. K. SCHOTT (2009): "Concording US harmonized system categories over time," Discussion paper, National Bureau of Economic Research.
- RODRIGUEZ-LOPEZ, J.-A. (2010): "Prices and Exchange Rates: A Theory of Disconnect," Manuscript, University of California Irvine.
- SAURE, P. (2009): "Bounded Love of Variety and Patterns of Trade," *mimeo*, Swiss National Bank.
- SIMONOVSKA, I. (2009): "Income Differences and Prices of Tradables," Manuscript, University of California, Davis.
- SIMONOVSKA, I., AND M. WAUGH (2011): "The Elasticity of Trade: Estimates and Evidence," *mimeo*.
- ZHELOBODKO, E., S. KOKOVIN, M. PARENTI, AND J.-F. THISSE (2011): "Monopolistic Competition in General Equilibrium: Beyond the CES," *Econometrica*, 80(6), 2765–2784.