

NBER WORKING PAPER SERIES

UNDERSTANDING MARKUPS IN THE OPEN ECONOMY UNDER BERTRAND
COMPETITION

Beatriz de Blas
Katheryn Russ

Working Paper 16587
<http://www.nber.org/papers/w16587>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2010

The authors thank seminar and conference participants at the 2010 American Economics Association winter meetings, the Banco do Portugal, the 2011 European Economics Association/Econometrics Society Meetings, the 2010 European Trade Study Group, DEGIT XV at Goethe University of Frankfurt, the Federal Reserve Bank of St. Louis, the 2011 NBER International Trade and Investment Group spring meeting, Stanford University, and the 2010 Western Economics Association International meetings. They are particularly indebted to Andrew Bernard, Ariel Burstein, Jeffrey Campbell, Robert Feenstra, Stefania Garetto, Yingyao Hu, Marc Melitz, Virgiliu Midrigan, Dave Richardson, Barbara Spencer, Jonathan Vogel, and Nick Zolas for detailed comments. Beatriz de Blas acknowledges financial support from the ECO2008-04073 project of the Spanish MEC, and the CCG08-UAM/HUM-4438. This paper circulated in an earlier form under the title "Teams of Rivals: Endogenous Markups in a Ricardian World." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by Beatriz de Blas and Katheryn Russ. 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.

Understanding Markups in the Open Economy under Bertrand Competition
Beatriz de Blas and Kathryn Russ
NBER Working Paper No. 16587
December 2010, Revised October 2011
JEL No. F0,F1,F4

ABSTRACT

This paper introduces a new model to capture the key stylized facts of firm price setting behavior in the open economy. At the firm level, producers adjust their domestic prices more frequently and fully in response to marginal cost shocks than for the same goods sold abroad. At the macro level, the terms of trade are less volatile than the real exchange rate and trade liberalization reduces markups. We present tractable analytical distributions of markups that are sensitive to market structure when firms are heterogeneous, bringing trade theory up to date with existing numerical and empirical studies.

Beatriz de Blas
Dpto. A.E.: Teoría e Historia Económica
Universidad Autónoma de Madrid
C/ Francisco Tomás y Valiente, 5
28049 Madrid, Spain
beatriz.deblas@uam.es

Kathryn Russ
Department of Economics
University of California, Davis
One Shields Avenue
Davis, CA 95616
and NBER
kkruss@ucdavis.edu

1 Introduction

At the firm level, producers adjust their domestic prices more frequently and fully in response to shocks to marginal cost than they do for the same goods sold abroad. At the macro level, the terms of trade are less volatile than the real exchange rate. This paper introduces a new model to capture these stylized facts of open economy pricing behavior in a simple way. To do so, we make a nontrivial extension to a canonical model of trade with heterogeneous firms, bringing theory up to date with existing empirical and numerical studies. The main engine driving the results is a form of price rigidity that arises endogenously due to cutthroat competition, even though prices are otherwise perfectly flexible.

Our model involves a finite number of firms competing within each industry. The most efficient firm in the industry ultimately becomes the sole supplier of that particular good, but only because it beats back its competitors by underselling them: it cannot charge a price higher than the marginal cost of its next best rival. We argue that trade costs make firms' prices more likely to be bound by their next best rival when selling overseas compared to their home market. We show analytically that this cutthroat competition can generate reduced markups under trade, pricing to market, and imperfect pass-through by causing price rigidity even when prices are not set in advance due to menu costs or other constraints. Bernard, Eaton, Jensen, and Kortum (2003, hereafter BEJK) pioneered the use of this form of price competition with heterogeneous firms and trade almost a decade ago. They integrated out the number of competitors to simplify their model, which makes the distribution of markups invariant to entry or trade. In contrast, we endogenize and explicitly focus on the number of entrants that compete in each industry so that markups are sensitive to market size and structure, in line with findings in the closed-economy literature such as those by Campbell and Hopenhayn (2005).

This micro feature of a trade model has macroeconomic implications, as the degree of rivalry affects the variation in markups that firms charge across industries and across countries. A fortunate fraction of firms are so far superior to their next best rival that they charge the maximum markup, which is the

Dixit-Stiglitz (1977) markup determined by the elasticity of substitution across industries. Only these firms can adjust prices upward without being undersold by their competitors. Having more competitors reduces the likelihood of being so fortunate, reducing the degree and frequency of price adjustment. The intensity of the Bertrand rivalry thus governs price setting behavior both within and across countries in a brand new way.

Empirical studies in international macroeconomics and trade present a rich and seemingly disparate array of stylized facts regarding firm pricing behavior when selling for domestic and export sale. In particular, papers such as Fitzgerald and Haller (2010), Schoenle (2010), Gopinath, Itshoki, and Rigobon (2010), and Gopinath and Itshoki (2010) show that firms often price to market and do not fully pass on changes in marginal costs and exchange rates to foreign buyers. In addition, several of these recent studies using firm-level data report more frequent price changes on domestic sales than on goods sold abroad and higher rates of pass-through among firms that change their prices more frequently.¹ An impressive list of empirical studies in international trade also demonstrates that trade liberalization is associated with firms charging lower markups over marginal costs when setting prices. Among these are Levinsohn (1993), Harrison (1994), Roberts and Supina (1996), Bottasso and Sembenelli (2001), Novy (2010), and Feenstra and Weinstein (2010). We believe that the two sets of findings— lower markups under trade and relatively rigid export prices— are related through trade costs and Bertrand competition.

The basic insight stems from the numerical simulations of Atkeson and Burstein (2007 and 2008), Garetto (2009), and de Blas and Russ (2011). All of these numerical studies start with a Fréchet or lognormal distribution of firm efficiency levels, then build on BEJK by computing markups under Bertrand competition (also Cournot in the case of Atkeson and Burstein (2008)). Collectively, they note that the size of the markup shrinks under trade and that trade costs make firms less able to pass on shocks to marginal costs by raising export prices. They also note another key feature— that the number of competi-

¹See Schoenle (2010) and Gopinath and Itshoki (2010) for U.S. data and Fitzgerald and Haller (2010) for Irish data.

tors within each industry, either the number of domestic competitors or foreign trading partners, affects both the size of the average markup and the degree or frequency of pass-through. Atkeson and Burstein (2007 and 2008) show that the special behavior of the endogenous markups combined with costly trade results in a lower volatility of the terms of trade relative to the real exchange rate.

We use free entry to introduce a cutthroat group of rivals for each industry within the BEJK framework that generates the markup and pricing behavior observed both empirically and numerically in this raft of previous studies. We assume an independent Fréchet distribution of efficiency draws for each country, following Eaton and Kortum (2002). Our model achieves tractable analytical solutions for the distribution of markups under autarky and trade when the finite number of firms competing to supply the market is not filtered out. We show that the distribution of markups in BEJK, which is invariant to market structure, can arise either by assuming an infinite number of rivals or an underlying Pareto distribution of efficiency draws. In short, we use free entry to build a standard Ricardian model of trade with a fully specified distribution of markups that is consistent with observed domestic and export pricing behavior. The assumption in our model that rivals to the best firm in each industry are latent is not necessary to achieve these distributions. We maintain this assumption to nest within existing literature and keep the model as simple as possible while we illustrate their implications for firm-level and aggregate price adjustment.

In contrast with Melitz (2003), entry in this Ricardian model does not affect the number of goods produced, but rather the number of firms competing to be the low-cost supplier of a particular good. “Competing” in this sense means drawing an efficiency parameter from an identical distribution and being ready to jump into production if a chance arises to undersell an active firm. The most efficient firm will have the lowest cost—the first order statistic for costs in the industry²—and become the only active supplier. An increase in the

²The first order statistic is the first (lowest) cost in a random sample arranged in ascending order of magnitude (see David and Nagaraja, 2003).

number of firms that compete to be the low-cost supplier of a good changes the shape of the entire distribution of marginal costs and markups. In our order-statistic framework, increasing entry reduces the expected marginal cost of the best suppliers, moving the mass toward the lower end of the cost distribution. It also shifts the mass of the distribution of markups toward the lower end. Therefore, increased entry reduces the aggregate price level under autarky.

Openness to trade has a first-order effect that is similar to increasing domestic entry under autarky. Higher geographic frictions impede trade as in BEJK, but also increase suppliers' market power, allowing them to charge higher markups conditional on the trade cost. Trade has the potential to push out the technological frontier directly by increasing the total number of rivals worldwide for any market, as each additional competitor for a market represents a chance of having a lower cost supplier. Under plausible conditions, trade can even increase domestic entry. Both because the distribution of markups is itself a function of trade costs and because increasing the number of rivals lowers both the average marginal cost and the average markup across industries, the model includes gains from trade above and beyond the productivity gains from reallocation across firms discussed in Arkolakis, Costinot, and Rodriguez-Clare (2010). It also allows for a very different process governing markups than in Melitz and Ottaviano (2008). In Melitz and Ottaviano's (2008) seminal work, the limit price is determined principally by a quadratic term in the utility function, so that the markup defined as price divided by marginal cost (rather than price minus marginal cost) is invariant to both trade and entry.³ In our model, the markup is determined by the marginal cost of the second best competitor, which creeps closer to the marginal cost of the best firm in an industry as either domestic entry or openness toward foreign competitors increases.

We also show the importance of the pre-existing level of domestic competition when evaluating the impact of trade on markups and prices. Higher domestic entry results in fewer firms charging the maximum markup, leaving less room for foreign competitors to challenge high profit margins in the domestic market. Thus, trade has a bigger effect on markups and prices in countries

³This is proven by Arkolakis (2011), online theoretical appendix.

or industries with small pools of rivals before liberalization due to a small domestic market size or high entry cost. Further, trade openness can increase or decrease domestic entry, according to whether the fixed cost of exporting is smaller or larger than the fixed cost of domestic entry. In the case where the fixed cost of exporting is smaller, trade can introduce exponentially large increases in domestic entry, providing an additional source of gains from trade.

The paper is organized as follows. Section 2 presents a simple closed economy model with analytical solutions for the distribution of markups and prices which include the number of rivals. We show the relationship between entry and the aggregate domestic price level. Section 3 considers the implications of trade in goods for these distributions given symmetric or asymmetric trading partners. Our distribution of markups for the open economy reveals pricing to market in the case of costly trade and that trade reduces markups in most cases. We also discuss the implications of trade for entry and aggregate output. In Section 4, we briefly review recent evidence on the quantitative importance of idiosyncratic shocks to marginal costs, then show that our distribution of markups implies a higher degree and frequency of price adjustment (higher price volatility) in response to these shocks for domestic versus export sales. Section 5 concludes.

2 Autarky

The heart of the model lies in the production of intermediate goods by heterogeneous firms. For simplicity, we assume that producers of the final good are perfectly competitive and assemble the intermediate goods, with no additional capital or labor necessary. The continuum of intermediate goods j spans the fixed interval $[0,1]$. The assembly process uses a technology involving a constant elasticity of substitution across inputs, with aggregate output given by

$$Y = \left[\int_0^1 Y(j)^{\frac{\sigma-1}{\sigma}} dj \right]^{\frac{\sigma}{\sigma-1}}.$$

We consider each intermediate input j as representing a different industry and assume that the price elasticity of substitution between output from different industries σ is greater than one. The demand for an individual input is downward sloping in its price, $Y(j) = \left(\frac{P(j)}{P}\right)^{-\sigma} Y$, and the aggregate price level P is given by

$$P = \left[\int_0^1 P(j)^{1-\sigma} dj \right]^{\frac{1}{1-\sigma}}. \quad (1)$$

Each producer of an intermediate good draws an efficiency parameter z from a cumulative distribution $F(z)$ with positive support over the interval $(0, \infty]$. Eaton and Kortum (2009, Chapter 4) describe a process whereby over time, $F(z)$ can emerge as a frontier distribution representing the efficiency levels associated with the best surviving ideas available to produce a particular good j . Being the distribution of the best surviving ideas, $F(z)$ naturally takes on an extreme value form and under mild assumptions, it can be characterized by a Fréchet distribution.⁴ Thus, we assume that an endogenous number of firms r each draw an efficiency parameter from a distribution given by

$$F(z) = e^{-Tz^{-\theta}}.$$

We assume that $T > 0$ and also that the shape parameter, θ , is positive. Only the most efficient firm with efficiency level $Z_1(j)$ in any industry supplies the market. This efficiency parameter increases the level of output a firm produces from one unit of a composite input Q :

$$Y(j) = Z_1(j)Q(j).$$

Marginal cost for this most efficient firm, $C_1(j)$, is inversely related to the

⁴In particular, EK suppose that each period a group of new ideas emerges with the quality of these ideas distributed as Pareto. Over time, the distribution of the best (most efficient) idea surviving from each period then becomes Fréchet, also known as an inverse Weibull (Pawlas and Szynal, 2000). Costs are inversely related to efficiency levels, so costs in this case are Weibull distributed, as in Eaton and Kortum (2002).

efficiency parameter,

$$C_1(j) = \frac{wd}{Z_1(j)},$$

which accounts for both the cost of the composite input, w , and any frictions involved in sending intermediate goods to the assemblers of the final good, $d \geq 1$. We assume that both labor and intermediate goods are used in the production of intermediate goods with constant cost shares: $w = \omega^\beta p^{1-\beta}$, with ω being the labor wage rate and p the cost of a bundle of intermediate goods. The cost parameter drawn by any firm hoping to produce good j is distributed

$$G(c) = 1 - e^{-T(wd)^{-\theta} c^\theta}.$$

Given that some number of rivals r draw an efficiency parameter hoping to be the low-cost supplier of industry j , the distribution of the lowest cost $C_1(j)$ is⁵

$$G_1(c_1) = 1 - e^{-rT(wd)^{-\theta} c_1^\theta}. \quad (2)$$

We assume that $d = 1$ under autarky in this section and for domestic sales in the open economy in Section 3.

2.1 The distribution of markups

Let $C_2(j)$ represent the unit cost of the second-best competitor in industry j , who sits inactive but ready to begin production instantly should the opportunity arise. Given the CES assembly technology for the final good, the lowest-cost firm producing good j would like to set a price that provides the maximum markup possible subject to demand—the CES markup, $\bar{m} \equiv \frac{\sigma}{\sigma-1} > 1$. However, if charging the CES markup results in a price that exceeds the marginal cost of the second-best competitor waiting in the wings, the lowest-cost supplier may find itself undersold. In short, no firm can charge a price that exceeds the

⁵See Rinne (2009), p.237 for derivation. As shown in Eaton and Kortum (2009), EK and BEJK simplify their frameworks by using the underlying assumption that the number of firms competing to be the low-cost supplier in any industry is a random variable with a Poisson distribution. It elegantly drops from the analysis. In contrast, we preserve the number of rivals in the following analysis.

unit cost of its next best rival. The low-cost supplier in each industry j takes the prices of the low-cost supplier in every other industry as given. The markup for industry j is then

$$M(j) = \min \left\{ \frac{C_2(j)}{C_1(j)}, \bar{m} \right\}.$$

With this formula for the markup, we compute the expected output-weighted price for any good j in several steps. First, note that the price for good j , $P(j)$, is given by

$$P(j) = \begin{cases} \frac{C_2(j)}{C_1(j)} C_1(j) = C_2(j) & \text{for } \frac{C_2(j)}{C_1(j)} \leq \bar{m} \\ \bar{m} C_1(j) & \text{for } \frac{C_2(j)}{C_1(j)} \geq \bar{m} \end{cases}$$

Thus, the pricing rule depends not only upon the distribution of the first and second order statistic of the marginal costs, but also upon the distribution of the ratio of the two order statistics. In Appendix A we use a straightforward Jacobian transformation on a result from Malik and Trudel (1982) to obtain the distribution of $\frac{C_2(j)}{C_1(j)}$, which is the distribution of the markup before imposing the maximum markup from the CES preferences. Assuming that the frontier distribution of efficiency parameters is identical for every industry j , the probability density of the ratio $\frac{C_2(j)}{C_1(j)}$ is given by

$$h(m) = \begin{cases} \frac{r(r-1)\theta m^{-(\theta+1)}}{[(r-1)+m^{-\theta}]^2} & \text{for } 1 \leq m < \bar{m} \\ \int_{\bar{m}}^{\infty} \frac{r(r-1)\theta m^{-(\theta+1)}}{[(r-1)+m^{-\theta}]^2} dm & \text{for } m = \bar{m} \\ 0 & \text{for } m > \bar{m} \end{cases}, \quad (3)$$

At the maximum markup, there is a mass point, which we show in Figure 1.

Like the distribution of markups given in BEJK, this distribution is statistically independent of $C_1(j)$ and $C_2(j)$. In fact, for very large r , we have $\lim_{r \rightarrow \infty} h(m) = \theta m^{-\theta-1}$ for $1 \leq m \leq \bar{m}$, which is a Pareto density for markups identical to the one in BEJK. However, because we explicitly include the number of rivals r — rather than integrating it out to focus on the role of gravity in a Ricardian setting as they do— we see that the distribution of markups is

directly affected by the number of firms competing to be the low-cost supplier.⁶ One can conceptualize r as an exogenous policy parameter, as in the numerical analysis by Atkeson and Burstein (2007 and 2008) and de Blas and Russ (2011), or endogenize it using a free entry condition as in Melitz (2003). The key is that unlike models using a Pareto distribution of firm efficiency parameters, the degree of entry embodied in r changes the shape of the entire distribution of markups, costs, and firm size.

Proposition 1 *The average markup is decreasing in the number of rivals r under autarky.*

Proof. For any given value $1 \leq m' \leq \bar{m}$, the probability that $M(j) \geq \frac{C_2(j)}{C_1(j)}$ is greater than or equal to m' is decreasing in r :

$$\frac{\partial \Pr[M(j) \geq m']}{\partial r} = \frac{\partial \left(\int_{m'}^{\infty} \frac{r(r-1)\theta m^{-(\theta+1)}}{[(r-1)+m^{-\theta}]^2} dm \right)}{\partial r} = \frac{-[(m')^\theta - 1]}{[1 + (r-1)(m')^\theta]^2} < 0.$$

Equivalently, we can say that the distribution of markups when r is low first-order stochastically dominates the distribution of markups with a higher r . First-order stochastic dominance implies a higher expected value; therefore $E[M(j)]$ must be decreasing in r . ■

With the CES bundling technology, firms will never set a markup greater than \bar{m} , creating a mass point in the density at \bar{m} , since all cases where $\frac{C_2(j)}{C_1(j)}$ exceeds \bar{m} are assigned a value of \bar{m} . The probability of charging the maximum markup is simply

$$\Pr \left[\frac{C_2(j)}{C_1(j)} \geq \bar{m} \right] = \int_{\bar{m}}^{\infty} h(m) dm = \frac{r}{1 + (r-1)\bar{m}^\theta}. \quad (4)$$

Note that as \bar{m} goes from its own upperbound of ∞ (for $\sigma = 1$) to its lowerbound of 1 (for $\sigma \rightarrow \infty$), this probability moves monotonically from 0 to 1, so

⁶Claessens and Laeven (2004) and de Blas and Russ (2011) refer to this phenomenon as “contestability.”

it is a well behaved cumulative distribution function over the range of possible markups.

Corollary 1 *In expectation, the fraction of firms charging the maximum markup is decreasing in the number of rivals r under autarky.*

Proof. The proof of Proposition 1, combined with equation (4) shows that the probability of $\frac{C_2(j)}{C_1(j)}$ being at least as large as \bar{m} is decreasing in the number of rivals. Markups are set equal to \bar{m} whenever $M(j)$ would be greater than \bar{m} without the restriction of the CES upperbound. Thus, in expectation, the fraction of firms charging the maximum markup is decreasing in the number of rivals. ■

As the number of rivals in an industry j increases, both the average markup and the probability that any firm charges the maximum markup falls—increased rivalry squeezes markups. Intuitively, the result emerges because, on average, increasing the number of rivals in our order-statistic framework diminishes the difference between the costs of the two best potential suppliers. This is not the case for a Pareto distribution of firm efficiency levels, as shown in Buch, Russ, and Schnitzer (2011). When firms draw from a Pareto distribution of efficiency levels, markups are again Pareto distributed as in BEJK (and in our special case above with many competitors), with no impact from the number of rivals. To reinterpret BEJK’s sports analogy in our setup: with the distribution of costs in equation (2), a competitor running second in a race will run even faster relative to the winner when there are more competitors behind him. However, with Pareto efficiency draws, no matter how many additional competitors trail behind in the race, each runner maintains both his speed and spacing *relative* to the person in front of him.⁷

To illustrate our new distribution of markups, Figure 1 shows the restricted distribution of markups when r equals its minimum value of 2, versus 20, the

⁷We believe the key difference is that the value of any outcome z enters the inverse of the hazard function linearly, which is not the case for the Fréchet used here or the lognormal used in Atkeson and Burstein (2008). Or more simply, the Pareto mean is linear in its minimum, which is also the case for the uniform distribution. Their density functions are flat or convex, rather than being strictly concave around the mode as in our model.

number of rivals chosen by Atkeson and Burstein (2007) calibrated to match U.S. industry concentration. We use $\theta=3.6$ and $\sigma=3.79$, as estimated by BEJK. The fraction of firms charging the maximum markup falls drastically, from one-half to just over one-third. We will discuss the implications of this statistic for price rigidity but first, we use the distribution of markups to compute the aggregate price level.

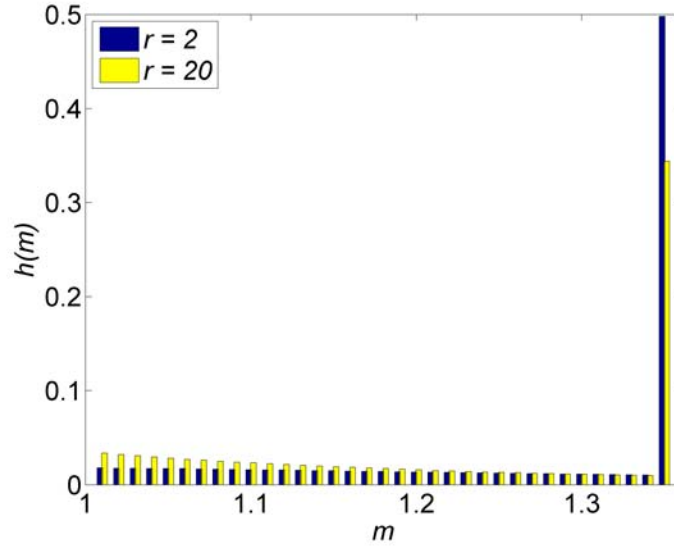


Figure 1: Increasing the number of rivals reduces markups

2.2 The distribution of prices

As shown in de Blas and Russ (2011), the joint distribution for the first and second order statistic also contains the number of rivals r :

$$g_{1,2}(c_1, c_2) = r(r-1) [\theta T w^{-\theta}]^2 c_1^{\theta-1} c_2^{\theta-1} e^{-T w^{-\theta} c_1^{\theta}} e^{-(r-1) T w^{-\theta} c_2^{\theta}}.$$

To find the marginal distribution for $C_1(j)$ ($C_2(j)$), one can integrate the joint distribution over values of c_2 (c_1).⁸ We find that increasing the number of rivals

⁸Integrating the joint distribution over c_2 from c_1 to ∞ , for instance, one obtains the marginal distribution $g_1(c_1)$ and sees immediately that it is equal to the first derivative of

leads, on average, to lower costs in the industry. We compute the moment $1 - \sigma$, which appears in the formula for the aggregate price level (1), for the first and second order statistics of marginal costs, so that we can use them below to construct the aggregate price level:

$$\begin{aligned} E[C_1(j)^{1-\sigma}] &= (rTw^{-\theta})^{\frac{\sigma-1}{\theta}} \Gamma\left(\frac{1-\sigma}{\theta} + 1\right), \\ E[C_2(j)^{1-\sigma}] &= (Tw^{-\theta})^{\frac{\sigma-1}{\theta}} \Gamma\left(\frac{1-\sigma}{\theta} + 1\right) \left[r(r-1)^{\frac{\sigma-1}{\theta}} - (r-1)r^{\frac{\sigma-1}{\theta}}\right]. \end{aligned}$$

We know from Proposition 1 that $E[C_2(j)]$ falls faster in r than $E[C_1(j)]$, since the expected ratio, $E\left[\frac{C_2(j)}{C_1(j)}\right]$ is falling in r .

Because the distribution of the markup is independent of outcomes for the individual order statistics $C_1(j)$ and $C_2(j)$, we can compute the expected price $P(j)^{1-\sigma}$ as

$$E[P(j)^{1-\sigma}] = \Pr[M(j) > \bar{m}] \bar{m}^{1-\sigma} E[C_1(j)^{1-\sigma}] + \Pr[M(j) \leq \bar{m}] E[C_2(j)^{1-\sigma}],$$

which we explain below is also increasing in r under very feasible conditions. Since firms in all industries draw from the same underlying distribution, using the law of large numbers one can calculate the aggregate price level,

$$P^{1-\sigma} = E\left[\int_0^1 P(j)^{1-\sigma} dj\right] = \int_0^1 E[P(j)^{1-\sigma}] dj = E[P(j)^{1-\sigma}].$$

Proposition 2 *The aggregate price level P is decreasing in the level of the number of rivals r under autarky as long as efficiency is sufficiently disperse.*

Proof. Intuitively, Proposition 2 is true because an increase in r shifts the distribution of markups to the left at the same time it reduces the first- and second-lowest unit costs on average. More rigorously, taking the derivative of

$G_1(c_1)$. To obtain the marginal for $C_2(j)$, one instead integrates over c_1 from zero to c_2 , as we do later for the open economy in Appendix D.

$P^{1-\sigma} = E[P(j)^{1-\sigma}]$ with respect to r yields

$$\begin{aligned} \frac{\partial[P^{1-\sigma}]}{\partial r} = & Pr[M(j) \geq \bar{m}] \frac{\partial E[(\bar{m}C_1(j))^{1-\sigma}]}{\partial r} + (1 - Pr[M(j) \geq \bar{m}]) \frac{\partial[C_2(j)^{1-\sigma}]}{\partial r} \\ & - \frac{\partial Pr[M(j) \geq \bar{m}]}{\partial r} (E[C_2(j)^{1-\sigma}] - E[(\bar{m}C_1(j))^{1-\sigma}]) . \end{aligned}$$

The first two terms on the right-hand side are positive, while it has been shown above that the probability of charging the maximum markup is falling in r , making its partial derivative negative in the third term negative. Thus, a sufficient condition for $P^{1-\sigma}$ to increasing in r is $E[(\bar{m}C_1)^{1-\sigma}] \leq E[C_2^{1-\sigma}]$. Using the expressions for $E[C_2(j)^{1-\sigma}]$ and $E[C_1(j)^{1-\sigma}]$ derived above, it is straightforward to check that this is possible whenever θ is not too large relative to σ . BEJK estimate a σ of 3.79, with θ equal to 3.6.⁹ For $\sigma = 3.6$, we have $E[(\bar{m}C_1)^{1-\sigma}] \leq E[C_2^{1-\sigma}]$ for all r given θ as high as 9 and as low as $\sigma - 1$. Intuitively, this means that increased contestability reduces the aggregate price level as long as the dispersion in firm efficiency levels is large enough to balance the consumer's love of variety. The sufficient condition coincides with structural estimates by BEJK (and also EK, for θ by itself). ■

2.3 The number of rivals

The key variable of interest in our model is the number of rivals r , which we pin down using a free entry condition.¹⁰ Following Melitz (2003), we assume that there is a uniform probability of death, $0 < \delta < 1$, in every period. Entrepreneurs must pay a fixed cost f . This fixed cost is denominated in units of revenue here, but we can also specify f in units of labor as in Melitz (2003) without affecting our qualitative results below at all. In equilibrium, the number of rivals must be such that the expected present discounted value

⁹That is, Recall that the elasticity σ here refers to the elasticity of substitution between industries, not the elasticity of substitution between goods within an industry, which is infinity since goods within an industry are all perfect substitutes. Thus σ can be finite and within the usual bounds.

¹⁰Since the circulation of this paper, new working papers by Holmes, Hsu, and Li (2011) and Zolas (2011) have also begun to do so, with different applications relating to agglomeration and patenting.

of output for an active producer equals the sunk cost of entry,

$$E_t \left[\sum_{s=0}^{\infty} (1 - \delta)^{t+s} (P_{t+s}(j)Y_{t+s}(j) - C_1(j)Y_{t+s}(j)) \right] \equiv f. \quad (5)$$

We also use the labor market clearing condition to define market size Y . In steady state, it is

$$\omega L = \beta \lambda P Y, \quad (6)$$

where L is the number of workers, β is labor's cost share in the input bundle used to produce intermediate goods and λ is the share of variable costs as a fraction of total profits,

$$\lambda = \frac{E[C_1(j)Y(j)]}{E[P(j)Y(j)]} = \frac{E[M^{-\sigma}(j)]}{E[M^{1-\sigma}(j)]}. \quad (7)$$

Isolating Y in equation (6), normalizing the wage $w \equiv 1$, and then substituting for Y and λ in the free entry condition, equation (5), yields the steady-state expression

$$\frac{E[M^{1-\sigma}(j)]}{E[M^{-\sigma}(j)]} = 1 + \frac{\beta \delta f}{L}. \quad (8)$$

Recall that the probability of forced exit, δ , is independent of firm efficiency, and that the distribution of the markup is independent of the distribution of costs,¹¹ so in Appendix B we show that the free entry condition reduces to

$$E[\ln M(j)] \geq \ln \left(1 + \frac{\beta \delta f}{L} \right). \quad (9)$$

The distribution of the markup derived above does not yield a closed-form solution for the expected markup $E[M(j)]$ or for the expected log markup, $E[\ln M(j)]$. However, we can determine an upper- and lowerbound for r . Noting from Jensen's inequality that $E[M(j)] \geq \ln E[M(j)]$ and that $\ln E[M(j)] \geq$

¹¹To see this, recall that the cost parameters C_k do not enter into the expression for $h(m)$ for $k \in N$.

$E[\ln M(j)]$, the form of the free entry condition in equation (9) implies

$$E[M(j)] \geq \ln E[M(j)] \geq E[\ln M(j)] \geq \ln \left(1 + \frac{\beta \delta f}{L} \right). \quad (10)$$

Proposition 1 tells us that the mean markup, $E[M(j)]$, is decreasing in the number of rivals, r . In combination with this insight from Proposition 1, $E[M(j)] \geq \ln(1 + \frac{\beta \delta f}{L})$ tells us that rivals will keep “entering” the industry (i.e., draw a productivity parameter) as long as the markup a rival expects to charge, if it is the low-cost supplier, generates expected profits at least as large as the amortized fixed cost of entry. We explore this graphically in the next section, contrasting domestic entry under autarky versus trade.

Appendix B.2 uses expression (9) to derive the upper- and lower- bounds for r given by

$$\frac{\ln \left(1 + \frac{\beta \delta f}{L} \right) (e^{\theta \bar{m}} - 1)}{\ln \left(1 + \frac{\beta \delta f}{L} \right) e^{\theta \bar{m}} - \bar{m}} \geq r \geq \frac{E[\ln M(j)] (e^{\theta \bar{m}} - 1)}{E[\ln M(j)] e^{\theta \bar{m}} - \bar{m}}.$$

Notice that the number of rivals in each industry grows as the fixed cost f , the share of labor in the input bundle β , and the exit rate δ fall, as well as when market size L is bigger. Note also that entry is not proportional to changes in market size L , but can grow much faster than L , a departure from the Melitz model that will impact gains from trade below. This yields many interesting implications, for example, from equation (2) it is clear that increasing the number of rivals influences the distribution of costs exactly like an increase in the technology parameter T . Thus, reducing barriers to entry pushes out the technological frontier, in addition to lowering the average markup. Since λ and P are both falling in r , we can see from equation (6) that reducing either the fixed cost or the exit rate increases aggregate output Y by boosting the number of rivals.

3 Trade in goods

Trade in our model not only shifts production toward lower-cost producers in the classic Ricardian sense, but also reduces markups in countries with low contestability, lowering the aggregate price level for all trading partners. The reason is simple: all else equal, openness increases the number of firms competing to serve the domestic market. In addition, trade costs increase the marginal cost for exporters situated far from a destination market relative to their rivals, making it more likely that their price will be bounded by a geographically closer rival. As trade costs eat into markups due to the direct competition of firms from closer locations, this relative gravity effect prohibits a larger fraction of exporters from being able to adjust prices in response to idiosyncratic shocks and limits the degree to which those that can actually *do* adjust them.

The squeeze on markups from openness generates a gain from trade that is new to the BEJK framework. Trade also invites increased domestic entry (a higher r), which reduces markups, generating a second gain from trade to the BEJK framework, though not to Bergin and Feenstra (2008), Devereux and Lee (2001), Melitz and Ottaviano (2008), or Rodriguez (2010). Furthermore, an increase in entry due to market scale effects by itself can shift the distribution of efficiency levels among active firms to the right, an effect not captured by either BEJK, Bergin and Feenstra (2008), Melitz and Ottaviano (2008), or Rodriguez (2010). The increase in entry acts both as a technological advance and an increase in the intensity of competition. Thus, trade always reduces the prices of imported goods relative to autarky *and* can also reduce the prices of domestically produced goods, as increasing the number of domestic rivals increases average efficiency. Geography, in the form of trade frictions, interferes with all three of these sources of welfare gains.

3.1 The distribution of costs in the open economy

We follow BEJK's notation, adding the subscript n to the terms $C_k(j)$, $g_k(c_k)$, and $G_k(c_k)$ from the autarkic case to refer to the costs and distribution of costs for goods supplied to country n in the open economy. When the potential

supplier is from country i we add the subscript i , so that the unit cost of the k^{th} most efficient firm from country i when supplying any good (j) to country n becomes $C_{kni}(j)$, drawn from the underlying cumulative distribution function $G_{kni}(c_k)$, with the corresponding probability density $g_{kni}(c_k)$. We assume that Eaton and Kortum's (2002) no arbitrage condition for trade costs holds: $d_{ni} < d_{ui}d_{nu}$.

Let $G_{1n}(c_1)$ be the probability that the low-cost supplier of a good j to the home country n has a marginal cost less than or equal to some level c_1 under trade. The probability is equal to one minus the probability that any other potential supplier— domestic or foreign— has a marginal cost greater than c_1 . The cumulative distribution for low-cost suppliers under trade is thus

$$\begin{aligned} G_{1n}(c_1) &= \Pr[C_{1n}(j) \leq c_1] = 1 - \prod_{i=1}^N [1 - G_{1ni}(c_1)] \\ &= 1 - e^{-\Phi_n c_1^\theta}, \end{aligned} \quad (11)$$

where $G_{1ni}(c_1)$ is the distribution of low-cost suppliers to n from country i , $\Phi_n = \sum_{i=1}^N T_i(w_i d_{ni})^{-\theta} r_i$, and $d_{ni} \geq 1$ is an iceberg trade cost involved in shipping goods from country i to country n for $i \neq n$. It is straightforward to show that the probability that a country exports to n is the same as in Eaton and Kortum (2002) and BEJK, but allowing for the number of rivals:

$$\pi_{ni} = \Pr[EXPORT_{ni}] = \frac{r_i T_i(w_i d_{ni})^{-\theta}}{\Phi_n}. \quad (12)$$

3.2 Geography and markups

In three steps, we can compute the full distribution of markups under costly trade with asymmetric countries. First, we consider the case that the best two rivals for a destination market originate in the same country. Let ψ_{ni} be the probability that the two best rivals to supply country n both originate in country i . Then, it must be that the two best rivals in a particular industry in country i are more efficient (have lower marginal costs) than any other potential

suppliers of the good to country n . Let c_{2i} be the second-best cost draw for an industry in country i . Then the probability that it is lower than the best draw for the same industry in any country $u \neq i$ is

$$\begin{aligned}\psi_{ni} &= \int_0^\infty \int_{c_1}^\infty g_{2ni}(c_{2i}) \prod_{u \neq i}^N [1 - G_{1nu}(c_{2i})] dc_{2i} \\ &= \pi_{ni} \psi'_{ni},\end{aligned}\tag{13}$$

where we define $\psi'_{ni} = \frac{(r_i-1)T_i(w_i d_{ni})^{-\theta}}{\Phi_n - T_i(w_i d_{ni})^{-\theta}}$, the probability that the second best producer in i will be the second best supplier to country n in the world market as a whole, given that the best producer of a good in i is also the best supplier to n worldwide.¹² The distribution of markups in this case is a simple application of our autarkic distribution, renaming r in equation 3 as r_i .

The second step is to compute the probability that the best supplier to n is from country i and the best rival supplier to supply country n is in country $u \neq i$, denoted ψ_{niu} . The unconditional probability that this occurs is the probability that the best supplier native to country u has some marginal cost c_{1u} , which lies between the first- and second-best draws in country i , c_{1i} and $c_{2i} > c_{1i}$, while the best rivals from all third countries ($v \neq u, i$) have a marginal cost that is larger than c_{1u} . See Appendix D for the full derivation of this probability, given by

$$\begin{aligned}\psi_{niu} &= \int_0^\infty \int_{c_{1i}}^{c_{2i}} \left(g_{1nu}(c_{1u}) \prod_{v \neq u, i}^N [1 - G_{1nv}(c_{1u})] dc_{1u} \right) g_{1ni, 2ni}(c_{1i}, c_{2i}) dc_{1i} dc_{2i} \\ &= \psi'_{niu} \pi_{ni} (1 - \psi'_{ni}),\end{aligned}\tag{14}$$

where $\psi'_{niu} = \frac{r_u T_u(w_u d_{nu})^{-\theta}}{\Phi_n - r_i T_i(w_i d_{ni})^{-\theta}}$, the probability that the second best supplier to n is in country u conditional on the best supplier being from i while the second best is not. Note that $\sum_{u \neq i} \psi_{niu} = 1$.

Finally, we compute the distribution of markups charged in country n given that the best rival to supply a good is in i and the second-best is in coun-

¹²Under symmetry, the probability collapses to the very intuitive expression $\psi = \frac{1}{N} * \frac{r-1}{Nr-1}$.

try u , which we call $h_{niu}(m)$. We use the formula for the distribution of the ratio of two independent random variables, $C_{1ni}(j)$ and $C_{1nu}(j)$, described by Mood, Graybill and Boes (1974, pp.187-88), given that, $C_{1nu}(j)$ is greater than $C_{1ni}(j)$:¹³

$$\begin{aligned} h_{niu}(m) &= \int_0^\infty \theta c_{1i} g_{1ni}(c_{1i}) \frac{g_{1nu}(mc_{1i})}{1 - G_{1nu}(c_{1i})} dc_{1i} \\ &= \frac{\theta r_i T_i(w_i d_{ni})^{-\theta} r_u T_u(w_u d_{nu})^{-\theta} (m)^{\theta-1}}{[r_i T_i(w_i d_{ni})^{-\theta} + r_u T_u(w_u d_{nu})^{-\theta} (m^\theta - 1)]^2} \end{aligned}$$

Then, the full distribution of markups in country n under trade, $\tilde{h}_n(m)$, is given by

$$\tilde{h}_n(m) = \sum_{i=1}^N \psi_{ni} h_i(m) + \sum_{i=1}^N \sum_{u \neq i}^N \psi_{niu} h_{niu}.$$

It is easy to verify that the relevant weights sum to one: $\sum_{i=1}^N \psi_{ni} + \sum_{i=1}^N \sum_{u \neq i}^N \psi_{niu} = 1$. The important result for our purposes is the probability that the supplier charges the maximum markup when its next-best rival is an exporter in a different country,¹⁴

$$\Pr[M_{niu} \geq \bar{m}] = \frac{r_i T_i(w_i d_{ni})^{-\theta}}{r_i T_i(w_i d_{ni})^{-\theta} + r_u T_u(w_u d_{nu})^{-\theta} (\bar{m}^\theta - 1)}. \quad (15)$$

One can see immediately that the supplier to country n exporting from country i will be more likely to charge the maximum markup when its next-best rival (a) resides in a country far from the destination country n (high d_{nu}), or (b) resides in a country with low contestability, low technology, or a high wage relative to country i . The country- i supplier's own distance from the destination country lowers the probability that it can charge the maximum markup. If all countries

¹³We can integrate the density over the domain $[1, \bar{m}]$, noting the mass point at \bar{m} and see that it forms a well behaved cumulative distribution function that integrates to one.

¹⁴More generally, the cumulative probability $\Pr[M(j) \leq m'] = 1 - \Pr[M_{niu} \geq m']$ ranges from 0 to 1 as m' increases from 1 to ∞ , so it is a well behaved cumulative distribution function for markups.

are identical, this probability that a firm in i supplying country n charges the maximum markup when its next best rival is in another country $u \neq i$ reduces to $\bar{m}^{-\theta}$. This expression $\bar{m}^{-\theta}$ is easily shown to be lower than the probability under autarky in equation (4) for finite r .¹⁵

The only way that the average markup can increase under trade is if the home country n opened its borders to trade with a world dominated by one country that was both much closer than other trading partners (low d_{ni}) and was far superior to all other countries by having much lower labor input costs (low ω_i), or very advanced technology (high T_i). What is more, equation (15) implies that reducing the trade cost d_{ni} for one particular country i increases the probability that a foreign supplier from i will be able to charge their full autarkic markup when selling to country n , yielding an important argument for multilateral trade negotiations.

Lemma 1 *Trade lowers the aggregate price level.*

Proof. A country will never import a good with a higher price than it pays under autarky and the second-best competitor will never be less efficient than the second-best competitor under autarky. To quantify the impact on the aggregate price level, we can compute

$$\begin{aligned} P_n^{1-\sigma} &= E[P_n(j)^{1-\sigma}] \\ &= \Pr[M_n(j) > \bar{m}] \bar{m}^{1-\sigma} E[C_{1n}(j)^{1-\sigma}] + \Pr[M_n(j) \leq \bar{m}] E[C_{2n}(j)^{1-\sigma}] \end{aligned}$$

and note that

$$E[C_{1n}(j)^{1-\sigma}] = \int_0^{\infty} c_1^{1-\sigma} g_{1n}(c_1) dc_1 = (\Phi_n)^{\frac{\sigma-1}{\theta}} \Gamma\left(\frac{\theta - (\sigma - 1)}{\theta}\right),$$

which is strictly greater than its counterpart under autarky. We also can compute the same $1 - \sigma^th$ moment for the marginal cost of the second-best rival

¹⁵It also demonstrates another special case where the Pareto distribution nests within our model.

by using the probability that it is in the same source country i as the actual supplier, ψ_{ni} :

$$E[(C_{2n})^{1-\sigma}] = \sum_{i=1}^N \psi_{ni} E[(C_{2ni})^{1-\sigma}] + \sum_{i=1}^N \sum_{u \neq i}^N \psi_{niu} E[(C_{1nu})^{1-\sigma}],$$

which we know is at least as great as its counterpart under autarky because the second-best rival producer of a good j in the entire world (including the home country) by definition could not have a marginal cost any higher than the second-best rival under autarky. ■

Under costly trade, the markups that firms charge are different when they sell domestically compared to when they export. The formula for the distribution of markups, $\tilde{h}_n(m)$, reveals that they internalize a portion of the trade cost, unless they are so technologically superior or have such a huge unit input cost advantage that they can pass the entire cost on to the foreign consumer. We demonstrated that the probability of charging the maximum markup is lower when one's next best rival is from a different country. The effect of incremental reductions in the trade cost on the import penetration ratio is no longer a constant, which Arkolakis, Costinot, and Rodriguez-Clare (2010) report is the case for the BEJK model without entry. More formally, profits are no longer a constant share of revenues, independent of the variable trade cost. Instead, the share of profits in total revenues varies with the variable trade cost d , shrinking as d falls and firms are forced to charge lower markups due to competition from new foreign and possibly new domestic entrants. This violates the gravity restriction satisfied by many trade models, even though the probability of exporting to any country n , π_{ni} , appears very similar to the export equations in Eaton and Kortum (2002) and BEJK. Put more simply, the gains from trade liberalization can not be inferred from the value of aggregate flows alone because liberalization reduces markups, distorting the relationship between the trade cost and observed expenditures.

As in the variable-markup frameworks of Bergin and Feenstra (2008), Melitz and Ottaviano (2008) and Feenstra and Weinstein (2010), entry changes the effective elasticity of demand (the price-elasticity of marginal revenues), even

though the Dixit-Stiglitz elasticity of demand governing the upperbound for the markup is a constant. Thus, trade liberalization has the potential to create welfare gains not only through productivity-based comparative advantage, but also by reducing firms' market power. We close the model and show output growth under free trade versus autarky under symmetry and free trade below, but save detailed analysis of gains from costly trade with variable markups in this generalized Ricardian setting for future research and in order to focus our analysis on entry, pricing behavior, and the aggregate price level.

To the degree that trade induces new entry (increased r), it shifts the entire distribution of marginal costs to the left, similar to an innovation in available technology T . A particularly clean case occurs when countries are identical and that trade is costless.

Proposition 3 *In a world with symmetric countries, free trade (a) increases the total number of rivals competing to supply a destination market, (b) reduces the aggregate price level, and (c) reduces the expected markup, as well as the probability that firms will charge the maximum markup.*

Proof. To illustrate more intuitively how trade affects the full distribution of markups, it is useful to suppose for a moment that countries are identical and trade is costless, so that $T_i = T$, $\omega_i = \omega \equiv 1$, and $d_{ni} = 1$ for all i . Then we see that the distribution for the lowest unit cost among all potential suppliers to any country n in equation (11) reduces to the Weibull distribution

$$G_{1n}(c_1) = 1 - e^{-rN T c_1^\theta},$$

which is observationally equivalent to a world with $R = rN$ rivals who all draw from an underlying distribution that takes the same form as the distribution of cost parameters for any individual country, $G(c) = 1 - e^{-T c^\theta}$.¹⁶ The distribution

¹⁶The distribution of first order statistics for samples drawn from a Weibull distribution is also Weibull.

of markups in this special case takes the form

$$\tilde{h}(m) = \frac{R(R-1)\theta m^{-(\theta+1)}}{[(R-1) + m^{-\theta}]^2}.$$

The implication is clear: trade has the same effect on the distribution of markups as increasing contestability and therefore reduces the number of firms charging the maximum markup and, all else equal, the aggregate price level, which takes the same form as under autarky, but with the total number of rivals for each market, R . Defining r^a as the number of rivals under autarky, we will show that $R > r^a$. From here, all three pieces of Proposition 3 are straightforward.

Part a) To show that the number of rivals under trade equals a number $R > r^a$, we use the open economy version of the free entry condition and a labor market clearing condition that takes the same form as the closed economy version in equation (6) for each country. If all countries are identical and trade is costless, we have

$$E_t \left[\sum_{s=0}^{\infty} (1-\delta)^{t+s} \left(1 + \frac{N-1}{N} \right) (P_{t+s}(j)Y_{t+s}(j) - C_1(j)Y_{t+s}(j)) \right] \equiv f. \quad (16)$$

The condition simplifies to

$$\frac{E[M^{1-\sigma}]}{E[M^{-\sigma}]} = 1 + \frac{\beta\delta f}{(1 + \frac{(N-1)}{N})L}. \quad (17)$$

Since the left-hand side is decreasing in r and $\frac{\beta\delta f}{(1 + \frac{(N-1)}{N})L} < 1 + \frac{\beta\delta f}{L}$, it is clear that the possibility of exporting strictly increases entry. Thus, leaping from autarky to free trade increases the number of rivals competing to produce any good ($R = Nr > r^a$), in addition to reducing prices by reallocating production to more efficient producers.¹⁷

Part b) $E[C_2(j)^{1-\sigma}]$ under free trade and symmetry takes the exact form of its counterpart under autarky, only substituting $R > r^a$ for the number of

¹⁷We assume the fixed cost of exporting is zero for simplicity, but one can also derive a reasonable restriction on the size of a fixed cost of exporting that preserves this result.

rivals, making $E[C_2(j)^{1-\sigma}]$ greater than its counterpart under autarky. From the discussion in Lemma 1, we know that $E[C_{1n}(j)^{1-\sigma}]$ must also be greater than its counterpart under autarky. Therefore, $(P)^{1-\sigma}$ must be greater than its counterpart under autarky P^a , revealing that the aggregate price level falls under trade: $P < P^a$.

Part c) It follows directly from the derivative in Proposition 1 and the fact that $R > r^a$ that the average markup falls under trade. Similarly, the likelihood of charging the maximum markup falls when opening to trade. ■

The results from Proposition 3 echo those of Bergin and Feenstra (2008) and Melitz and Ottaviano (2008), but now within the *homothetic* preference structure of Bernard, Eaton, Jensen, and Kortum (2003). Atkeson and Burstein (2007 and 2008) show the results in Parts (b) and (c) numerically, while de Blas and Russ (2011) demonstrate that having a large number of rivals under autarky reduces the impact of trade liberalization on markups. Note also that increasing the number of trading partners has a similar effect to increasing the number of rivals in any trading partner, seen in numerical solutions calculated by Garetto (2009). Under costless trade, it does not matter how the rivals are distributed across countries. Markups respond as though all entrants worldwide compete on equal footing to be the low-cost supplier. As in classic studies of trade and endogenous market structure, geographic frictions here increase market power, dampening the effect of foreign industrial structure on domestic markups and prices.

What is more, trade has the potential to increase the number of entrants in any particular country. Figure 2 shows the number of rivals entering under autarky versus trade using equations (8) and (17). The curved lines represent the left-hand side of each equation, while the flat ones represent the amortized labor-weighted upfront fixed cost per customer. As shown in Section 2, the left-hand side of the equations represent a figure closely to the expected markup, which is lower under free trade, where there are $R = N * r^t$ rivals¹⁸ for any national market in each industry. We use BEJK's estimates for σ and θ , as

¹⁸We add the superscript for r under trade here for the sake of clarity if the figure is viewed independently from the text.

well as their calibration for β , 0.21. In addition, we assume that the firm exit rate δ is 0.025, according to default rates for firms in the U.S. reported in Russ and Valderrama (2010). To calibrate the fixed cost per customer, $\frac{f}{L}$ and $\frac{f^t}{1+\frac{N-1}{N}}$, we use gross fixed capital formation in the year 2000-2003 (before the housing boom) in terms of 1985 producer prices reported in the International Monetary Fund's *International Financial Statistics*, supposing that 0.025 percent of it is used for new firm entry to replace the exiting firms. The right hand side using the figures from 2000-2003 varies between 1.12 and 1.22. We choose figures that result in 20 domestic firms competing in each industry under trade to match Atkeson and Burstein's (2008) calibration for the U.S., and a lower number, 16, under autarky. This requires some fixed (not necessarily sunk) cost of exporting such that $f^t > f$ in our graph. If this amortized fixed cost per customer falls only a very small amount, domestic entry increases under trade. However, we show that this need not be the case: if it remains constant due to extremely high fixed costs of exporting, for instance, entry actually falls from 16 to 8, generating a dramatic selection effect.

In Figure 3, we show a conservative estimate of the gains for a country opening to free trade with one partner, assuming a dramatic selection effect where half of the rivals in both countries exit (minus one). That is, we suppose that the total (global) number of rivals in each industry is equal to its autarkic level plus only one more. The ratio $\frac{Y^t}{Y^a}$ is computed as in Appendix E. We also show a case with no selection, analogous to Krugman (1980), where there is no selection effect. In either case, when the autarkic number of rivals r^a increases, the gains from trade fall. For the case with selection, when the autarkic number of rivals grows very large—approaching 200, for instance, the gains disappear. Equation (8) implies that small countries, ones with smaller market size L , will have fewer rivals in each industry under autarky. Thus, gains from trade are greatest for small countries opening to trade, even if a pair of small countries establishes a free trade agreement, echoing a result in representative firm models such as Devereux and Lee (2001) under Cournot competition or Novy (2010) with a translog preferences. This is very intuitive, since we saw above that the distribution of markups is Pareto and thus invariant to trade and market

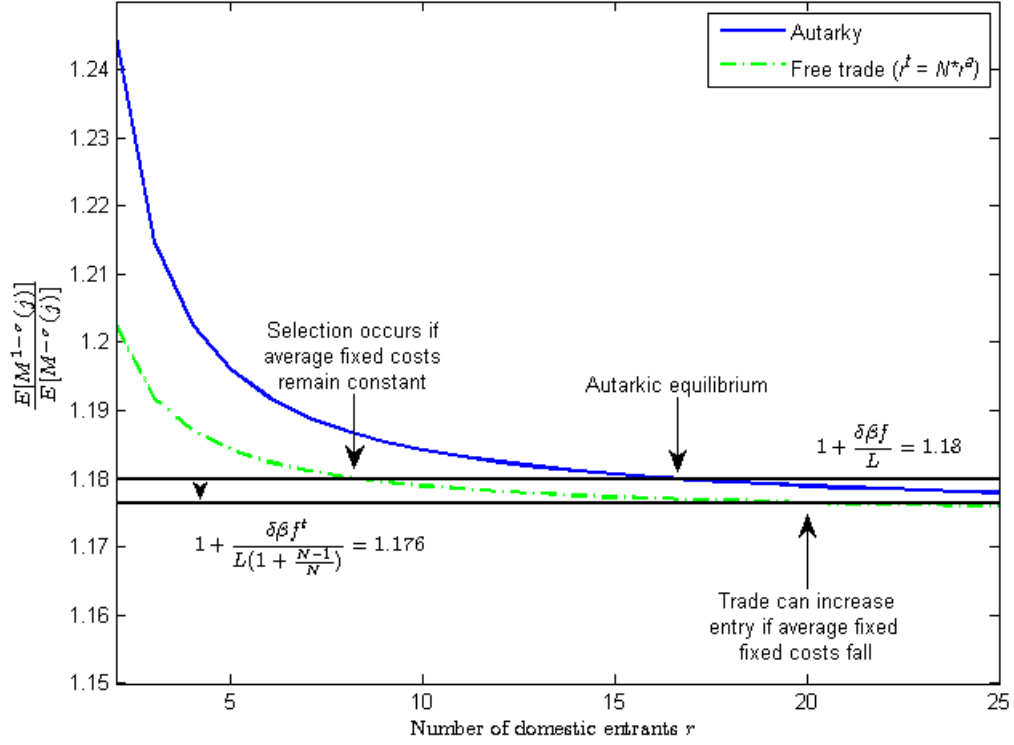


Figure 2: Free entry and the number of home-grown rivals under autarky versus trade

structure for large numbers of rivals. However, for a country opening to free trade with another that has 19 rivals in each industry under autarky, which under trade has 20 potential suppliers– 10 domestic rivals and 10 more rivals in the foreign country– gains from trade in terms of increased GDP are 23.3 percent.

3.2.1 A note on generality

Atkeson and Burstein (2008) point out that if the products of different firms within the same industry are less than perfect substitutes, the endogenous markup behavior of firms depicted here breaks down to some degree, making

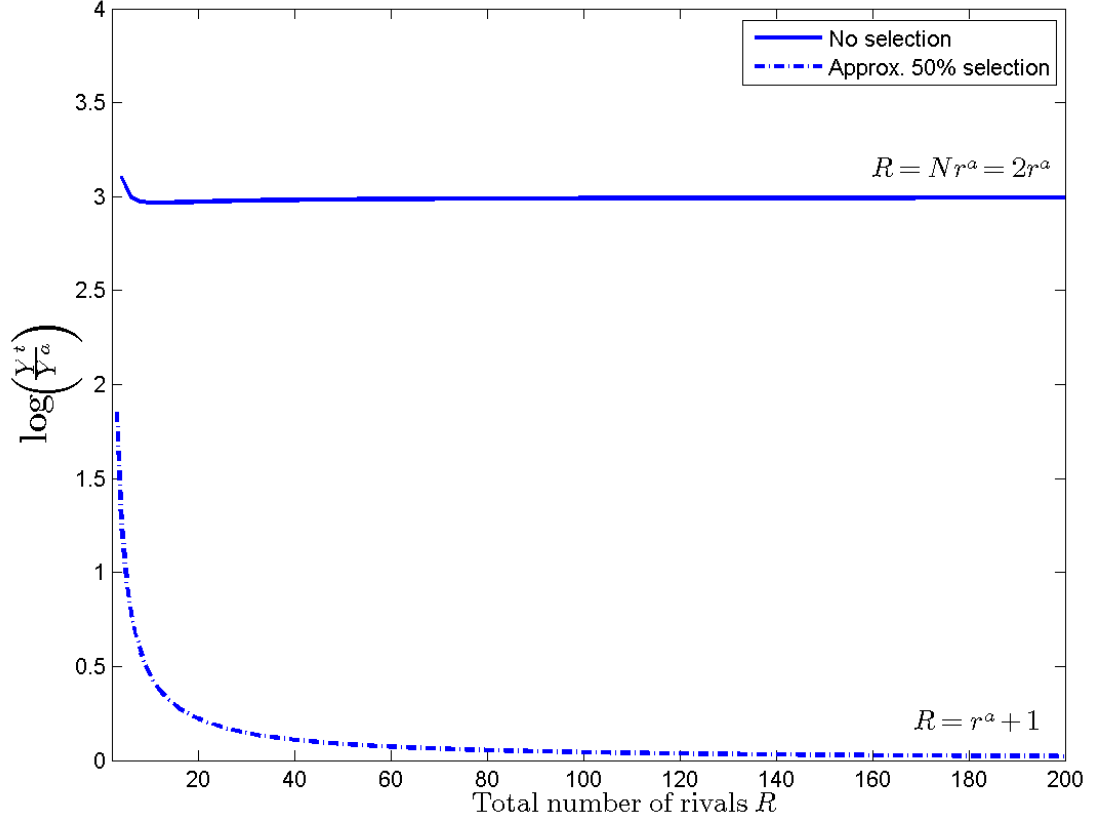


Figure 3: Gains from trade diminish when autarkic entry is high

our approach to oligopolistic competition less tractable. This is a reasonable criticism. However, Kucheryavyy (2010) shows that the number of firms influencing the markup converges to two continuously as the elasticity of substitution increases. We argue that at some level of disaggregation, products will be extremely close substitutes. Even if all upper layers of the CES nesting involve finite elasticities of substitution, the impact of endogenous markups from an underlying level of perfect substitutes would have similar implications for the frequency of observed price changes and the behavior of the aggregate price level, as long as the goods with prices characterized by endogenous markups were assembled in the country where they are consumed as part of the final

good.¹⁹ The location where intermediate goods are assembled is a critical factor in determining how the endogenous markup behavior affects relative prices and pass-through across countries. Our model assumes, like BEJK and Atkeson and Burstein (2008), that intermediate goods are assembled in the same country where the final good is consumed.

How disaggregate do product categories have to be for varieties within an industry to be close substitutes? In a structural estimation of the Atkeson and Burstein (2008) model, which uses quantity-based competition but the same nested CES framework as in BEJK and Atkeson and Burstein (2007), Edmond, Midrigan, and Xu (2009) estimate that the elasticity of substitution for the intermediate goods is 8.7 for goods *within* the 4-digit SIC level (goods are identified at the 7-digit SIC level) among firms in the Taiwan Annual Manufacturing Survey.²⁰ While the number 8.7 is obviously much lower than infinity, in terms of the implied markup it is not so far away: the markup shoots toward infinity when the elasticity of substitution between the goods of rival firms approaches 1, but it tends toward zero as the elasticity approaches infinity and already falls to 15% when the elasticity is 8. Their estimate of the elasticity is just under 8 for simulations with price-based competition of the form in BEJK. So even at an intermediate level of disaggregation, goods are already very close substitutes when analyzed in the context of a Ricardian model with endogenous markups.²¹

¹⁹The idea of a constant markup being imposed at an upper retail level in the destination country is consistent with findings by Berger, Faust, Rogers, and Steverson (2009), who find that the markup added onto imported goods by distributors *after* the goods arrive at the dock is stable over time.

²⁰There is a rich literature estimating the elasticity of substitution in the Dixit-Stiglitz (1977) framework with a constant elasticity of substitution, but we omit discussion of it here, as we are focusing on estimates implied by the type of model in this paper, a Ricardian model with endogenous markups.

²¹The probability density of markups resulting from the quantity-based competition in Atkeson and Burstein (2008) for their preferred calibration also looks strikingly Pareto-like in shape. Edmond, Midrigan, and Xu (2006) show that their (quantity-based) Cournot competition produces a greater dispersion in markups than the price-based Bertrand from BEJK, closer to the actual dispersion observed in their sample. However, it is not clear whether this would be the case once one takes into account the Taiwanese producers shipping to export markets, or the import competition within Taiwan's own domestic market, or calibrates the number of rivals to match observed industry concentration as in Atkeson and

One other potential question is whether the fact that rivals within each industry are latent— they do not produce and thus can not be observed for empirical analysis— is important in interpreting our results. The answer is a resounding “No.” De Blas and Russ (2011) generalize the type of Bertrand competition employed here to use a search framework, which allows for a continuum of firms to produce the same good as long as search costs restrict the number of prices that buyers check before making a purchase, in the spirit of Burdett and Judd (1983). They use numerics rather than the analytics here, but all of the same intuition is reflected in the numerical results, suggesting fertile ground for future research applying our analytical distributions.

4 Price adjustment, volatility, and passthrough

Empirical studies indicate that idiosyncratic shocks are likely to be prevalent and economically important: Gabaix (2010) finds that a substantial portion of observed aggregate fluctuations in U.S. output can be explained by idiosyncratic shocks falling across a distribution of heterogeneous firms, while Foster, Haltiwanger, and Syverson (2008) determine that idiosyncratic shocks affecting plant-level output have a standard deviation five times as large as that of industry-level productivity shocks. Thus, recent literature indicates that idiosyncratic shocks are likely to be important from a macroeconomic perspective. In this context, our setup can shed some light on the pricing behavior of individual firms subject to idiosyncratic shocks in domestic versus foreign markets. We start by looking at price adjustment in the closed economy with idiosyncratic shocks to fix ideas. Afterward, we apply a country-specific shock in the open economy model to look at pass-through.

Burstein (2008) rather than by counting the number of suppliers of 7-digit micro-industries. We find the study illuminating as the first to apply both Cournot and Bertrand competition in the Ricardian setting, running a horserace to explain observed markups in firm-level data.

4.1 Price adjustment in the closed economy

In the simple, closed-economy framework, a lower number of rivals leads to more frequent price changes in response to idiosyncratic shocks to marginal costs. The reason is clear from Figure 1. When r is low, more firms charge the maximum CES markup— their prices are not tightly bounded by the marginal costs of their next-best rival, so they are better able to pass on idiosyncratic increases in marginal cost to their customers. The fraction of firms that set their price equal to the marginal cost of the next-best rival are unable to do this. Since firms will not change prices in response to an idiosyncratic shock unless they charge the maximum markup, Figure 1 suggests that at least half of firms will never be able to adjust their prices upward *ever*, unless they experience a shock common to all rivals and which affects all rivals at exactly the same time. We apply a lognormally distributed idiosyncratic shock with the log of the shock being distributed $N(0,10)$, so that the standard deviation of the shock is 10%.²² After 1000 simulations, using the same parameters as in Figure 1, we compute that, all else equal, 73.3% of firms adjust their price in response to a shock when $r = 2$, while the figure falls to 64.8% when $r = 20$. This is consistent with results from Nakamura and Steinsson (2010), who find that no price changes are observed for 40% of products over the period 1982-2007, as well as Gopinath and Rigobon (2008) and Gopinath, Itskhoki, and Rigobon (2010), who report static prices for approximately 30% of their sample. The following corollary to Proposition 1 formalizes this result.

Corollary 2 *The degree and frequency of pass-through for idiosyncratic shocks to marginal cost is falling in the number of rivals r , as is price volatility.*

Proof. For some random i.i.d. shock ε to firm-specific marginal cost with probability density $\eta(\varepsilon)$ over the domain $(0, \bar{\varepsilon})$, we can compute the fraction of firms that will raise prices in response to an idiosyncratic increase in marginal

²²This is in line with calibration by Feenstra, Obstfeld, and Russ (2011) for micro-level shocks, drawing on empirical estimates by Basu, Fernald, and Kimball (2006) and Foster, Haltiwanger, and Syverson (2008).

costs. Suppose a shock occurs such that $\varepsilon > 1$, increasing the marginal cost for a particular active firm but not its rivals in the industry.

First, we note that only firms charging the maximum CES markup would be able to increase their prices, since firms setting prices bounded by the marginal cost of their next-best rival can not. Then, the probability that a firm will pass an idiosyncratic increase in marginal cost fully to buyers by raising its price is equal to the probability that the current price (\bar{m} times marginal cost) times the shock does not exceed the marginal cost of the next best rival,

$$\Pr [\bar{m}\varepsilon C_1(j) \leq C_2(j)] = \Pr \left[\frac{C_2(j)}{C_1(j)} \geq \bar{m}\varepsilon \right] = \Pr [M(j) \geq \bar{m}\varepsilon].$$

Since the distribution of markups is independent of ε , we can compute this probability as

$$\begin{aligned} \Pr [M(j) \geq \bar{m}\varepsilon] &= \int_{-\infty}^{\infty} \int_{\bar{m}}^{\infty} h(\varepsilon m) \eta(\varepsilon) dm d\varepsilon \\ &= \int_{-\infty}^{\infty} \frac{r}{1 + (r-1)(\varepsilon \bar{m})^\theta} \eta(\varepsilon) d\varepsilon. \end{aligned} \quad (18)$$

It follows from Corollary 1 that regardless of the probability distribution for ε , as long as the marginal cost shock is independent of the markup, the probability of full pass-through under autarky is decreasing in the number of rivals.²³ Multiplying ε above by some positive constant less than one, we see that the result is general to any degree of pass-through, not just full pass-through.²⁴ The intuition also applies for a downward cost shock, which is omitted here

²³That is, given the calculus used to prove Proposition 1, equation (18) implies that the probability of the markup being high enough to permit adjustment to positive price shocks is decreasing in the number of rivals r .

²⁴Our assumption that firms pay a fixed cost when they become active prevents the lowest-cost producer from having to adjust prices in response to temporary idiosyncratic shocks hitting its next-best rivals. The rivals will not find it profitable to try to undercut an existing producer unless they experience a transitory shock large enough to cover the entire fixed cost. We assume that the variance of costs is small enough that the likelihood of such a large shock is negligible.

for the sake of brevity. In this case, all firms charging the maximum markup would have to lower their prices, otherwise their markup would rise above \bar{m} , implying marginal revenues less than marginal costs. Further, some portion of firms charging a price equal to $C_2(j)$ would also lower prices, namely those for whom leaving the price at $C_2(j)$ resulted in a markup greater than \bar{m} . Thus, downward adjustment is most likely when firms are more likely to have relatively inefficient rivals, which is the case when r is low. Note that having less complete and less frequent price adjustment in response to idiosyncratic shocks implies lower price volatility. ■

4.2 Trade and prices

The expressions for markup behavior in the Section 3.2 yield pricing-to-market, incomplete pass-through, and the closely related facts that firms change prices on exported goods less frequently and with less synchronization relative to prices in the domestic market. Atkeson and Burstein (2007 and 2008) describe in brilliant detail the manner in which numerical simulations of BEJK and an innovative new quantity-based competitive framework result in pricing-to-market and incomplete pass-through, matching them with data on pricing behavior. Here, we demonstrate similar results algebraically. First, pricing-to-market is evident in the formula for $\tilde{h}_n(m)$ and both of its components, $h_i(m)$ and $h_{niu}(m)$. Unless trade is costless, firms can charge higher markups in their home markets than abroad because trade costs increase their domestic market power, as discussed above. The formulas also depict how firms set markups depending on the proximity of other export competitors in a particular destination market, if their next best rival is another exporter. Second, as under autarky, firms will only fully pass an increase in marginal cost to buyers in export market n if (a) they are already charging the maximum markup and (b) the price increase would not surpass the marginal cost of the next-best rival to supply country n . Although the logic is quite general, we can show this mathematically if we again invoke symmetry, this time with costly trade.

Suppose again that there is a shock to marginal cost ε such that a shock $\varepsilon > 1$ reduces efficiency and increases the marginal cost of an industry's low-

cost supplier in country n . The probability that pass-through occurs under trade is now

$$Pr \left[M_n(j) = \frac{C_{2n}(j)}{\varepsilon C_{1n}(j)} \geq \bar{m} \right] = \int_{-\infty}^{\infty} \int_{\bar{m}}^{\infty} \tilde{h}_n(\varepsilon m) \eta(\varepsilon) dm d\varepsilon. \quad (19)$$

Under symmetry, we have $\pi_{ni} = \pi = \frac{1}{N}$ and $\psi'_{ni} = \psi' = \left(\frac{r-1}{Nr-1} \right)$, yielding $\psi_{ni} = \psi = \pi\psi'$. In addition, $\psi'_{niu} = \frac{1}{N-1}$, so that we also have $\psi_{niu} = \frac{\pi(1-\psi')}{N-1}$. With this in mind, equation (19) becomes

$$Pr \left[\frac{C_{2n}(j)}{C_{1n}(j)} \geq \bar{m}\varepsilon \right] = \psi \int_{-\infty}^{\infty} \frac{r}{1 + (r-1)(\varepsilon\bar{m})^\theta} \eta(\varepsilon) d\varepsilon + \frac{\pi(1-\psi')}{N-1} \int_{-\infty}^{\infty} \frac{1}{1 + (\varepsilon\bar{m})^\theta} \eta(\varepsilon) d\varepsilon,$$

where $\psi = \frac{r-1}{N}$. Since $\frac{r}{1+(r-1)(\varepsilon\bar{m})^\theta} > \frac{1}{1+(\varepsilon\bar{m})^\theta}$ for any $r \geq 2$ and $\bar{m} \geq 1$, and r is at least as large under trade as under autarky, $Pr \left[\frac{C_{2n}(j)}{C_{1n}(j)} \geq \bar{m}\varepsilon \right]$ must be less than its autarkic counterpart given by equation (18). Therefore, the probability of full pass-through of cost shocks under trade must be less than the probability of full pass-through under autarky. The same can be shown for any degree of partial pass-through, as well.²⁵

For the case where countries are not symmetric, our markup formulas demonstrate results described in the numerical simulations of Garetto (2009). As we noted above from equation (15), the probability that a firm charges the maximum markup (and as a result, the degree of passthrough) in an export market is greater when the exporting country has a higher level of technology T or a lower wage ω than its competitor's source country. Thus, we show the point Garetto (2009) argues—“firms should do less pricing-to-market when exporting to relatively more productive (richer) countries.”

²⁵Although several studies have shown that pass-through depends on the choice of currency invoicing, Goldberg and Tille (2009) demonstrate that this currency invoicing choice also depends on the degree of competition in the destination market, so we view our market structure approach as quite relevant.

4.3 Frequency and synchronization of export price changes

Our distributions in Section 3 imply that the frequency of price changes will be smaller in export markets than in domestic markets. Unless an exporting country has a huge advantage in the form of high T , high r , or low labor costs, it is harder for firms to charge the maximum markup in an export market compared to their native market. This is due to the trade cost, which effectively increases exporters' marginal cost relative to domestic firms in the destination country. Since firms must be charging the maximum markup in order to pass on idiosyncratic or country-specific shocks in the form of higher export prices, fewer firms will change prices in export markets (as compared to their native market) when marginal costs increase. As a consequence, the median and average frequency of price changes must be lower for exports, as shown by Schoenle (2010) and Fitzgerald and Haller (2010). More intuitively, trade costs eat away a portion of firms' markups, giving them less leeway to adjust prices in response to increases in marginal costs. Atkeson and Burstein (2007 and 2008) demonstrate this point numerically. Below we illustrate it analytically for the first time in the formula for the distribution of markups and thus to other fundamentals, including Ricardian differences in technology and domestic rates of entry across markets.

Schoenle (2010) reports that export price changes are less synchronized than domestic price changes. Trade costs can generate this effect in the same way they reduce the frequency of price changes.²⁶ In Table 1, we summarize data generated by applying a small, country-specific shock²⁷ to our model calibrated as described above with two symmetric countries. These are small deviations from steady state to focus on the workings of the pricing mechanism, apart

²⁶The dichotomy could also come about because when marginal costs are subject to country-specific shocks, all domestic rivals experience the same shock, making it more likely that a firm can adjust its domestic price even if it is not charging the maximum markup. We focus on the macro shock here.

²⁷We use the same calibration as in Figure 1, with lognormal shocks that enter like ε above, but applied to all firms within a country. The shock is lognormally distributed with log of these shock distributed as normal with mean zero and variance one, so that the standard deviation is rather small, less than half of the standard deviation of industry-level shocks estimated by Basu, Fernald, and Kimball (2006)

from general equilibrium effects. The relative cost (i.e., country-specific) shock results in a frequency of export adjustment similar to that found by Schoenle (2010). Interestingly, it also replicates the frequency of export price changes relative to the frequency of domestic price changes, but only for lower trade costs. High trade costs give domestic suppliers enormous leeway to adjust prices in response to the relative cost shocks.

Gopinath and Itskhoki (2010) find that the frequency of price adjustment is positively correlated with the degree of exchange rate passthrough. This is true in our model, as the most frequent price adjusters among exporters are those who are able to charge the maximum markup in the foreign market, giving them the greatest degree of pass-through. As discussed above, firms charging the maximum markup are most able to pass through shocks to their relative marginal costs, including those arising from exchange rate movements, for instance. In the discussion of the distribution $\tilde{h}_n(m)$ in Section 3.2, we also noted that lower trade costs make it more likely that foreign suppliers charge the maximum markup. In Table 1, we see these effects illustrated by the positive quantitative relationship between the degree of pass-through and the frequency of price adjustment as trade costs fall.

The macro-level manifestation of this restricted price adjustment in export markets is reduced volatility in the terms of trade relative to the real exchange rate for high levels of contestability (r) in the host market or in the presence of trade costs. To illustrate the relationship between the micro and macro effects of relative cost shocks across countries, such as a small movement in the nominal exchange rate (see Burstein, Eichenbaum, and Rebelo (2005) for a discussion contrasting the impact of large versus small shocks on the real exchange rate), Table 1 lists the volatility of the terms of trade relative to the real exchange rate in U.S. data alongside results from simulated data for small cost shocks in our model.²⁸ We see in the table that the model delivers a variance in the terms of trade that is approximately one-half the variance of the real exchange rate when $d = 1.74$, the value for trade costs estimated by Anderson and van

²⁸That is, to focus on the main mechanism of the Bertrand pricing behavior, these are small departures from a symmetric steady state without second-order effects on wages.

Table 1: Effect of relative cost shocks on simulated price adjustment

		U.S. data*	$d = 1.75$		$d = 1.5$		$d = 1.25$		
			r=2	r=20	r=2	r=20	r=2	r=20	
		$\frac{\sigma_{TOT}}{\sigma_{RER}}$	0.56	0.71	0.67	1.02	0.96	1.97	1.88
Export prices	frequency of change	0.14-0.21	0.11	0.11	0.18	0.18	0.29	0.30	
	average pass-through	0.25	0.14	0.14	0.20	0.20	0.29	0.29	
Relative frequency of price changes (for domestic relative to export sales)		2.1-2.7	7.9	8.0	4.44	4.44	2.31	2.31	

*U.S. figures for $\frac{\sigma_{TOT}}{\sigma_{RER}}$ are from Corsetti, Dedola, and Leduc (2008), frequencies from Schoenle (2010), and pass-through from Gopinath, Itshoki, and Rogobon (2010).

Wincoop (2004). This corresponds with the figure reported for the U.S. in Corsetti, Dedola, and Leduc (2008). The variance of the terms of trade relative to the real exchange rate decreases monotonically as trade costs fall. This relative variance is lower when the level of domestic rivalry is higher. High trade costs suppress exporters' ability to adjust prices more strongly than the level of domestic entry, which is quite similar to the results from the variable elasticity of substitution framework developed by Gust, Leduc, and Vigfusson (2010). The results in our stylized setup present a puzzle in that they capture the relative variance of the terms of trade when trade costs are high, but replicate the relative observed frequency of domestic versus export price changes only when trade costs are low.

5 Conclusions

The purpose of this paper is to understand the fundamentals of pricing behavior for heterogeneous firms under Bertrand competition. We provide new distributions of markups which are sensitive to market structure, demonstrating how market structure has important implications for gains from trade and both the

level and volatility of prices. The distributions allow us to characterize in an analytically clean way firm markup behavior under trade, as well as the percentage of firms who can change their price in response to an idiosyncratic shock in any market, or in response to a source-country-specific shock in an export market. As in previous the numerical studies using similar frameworks, key results include imperfect passthrough, pricing-to-market, and a lower volatility of the terms of trade relative to the real exchange rate. Our breakthrough is that we explicitly characterize in tractable formulas an endogenous degree of pricing rigidity that depends on market structure and varies across destination markets due to the degree of domestic entry and the level of trade costs.

References

- [1] Anderson, James E. and Eric van Wincoop. 2004. Trade Costs. *Journal of Economic Literature* 42:692-751.
- [2] Anderson, Simon P., Andre de Palma and Jacques-Francois Thisse. 1987. The CES is a Discrete Choice Model? *Economics Letters* 24: 139-40.
- [3] Arkolakis, Costas. 2011. Market Penetration Costs and the New Consumers Margin in International Trade. *Journal of Political Economy* 118(6): 1151-1199.
- [4] Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare. 2010. New Models, Same Old Gains from Trade? Forthcoming, *American Economic Review*.
- [5] Atkeson, Andrew and Ariel Burstein. 2007. Pricing to Market in a Ricardian Model of International Trade. *American Economic Review Papers and Proceedings* (May) 97(2): 362-367.
- [6] Atkeson, Andrew and Ariel Burstein. 2008. Pricing to Market, Trade Costs, and International Relative Prices. *American Economic Review* (December) 98(5): 1998-2031.
- [7] Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. Are Technology Improvements Contractionary? *American Economic Review* 96(5): 1418-1448.
- [8] Berger, David, Jon Faust, John H. Rogers, and Kai Stevenson. 2009. Border Prices and Retail Prices. Federal Reserve Board of Governors International Finance Discussion Paper No. 2009-972.

- [9] Bergin, Paul C. and Robert C. Feenstra. 2008. Pass-through of Exchange Rates and Competition between Floaters and Fixers. *Journal of Money Credit and Banking* 41(1):35-70.
- [10] Bernard, Andrew, Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade," *American Economic Review* 93(4): 1268-1290.
- [11] Bottasso, Anna, and Alessandro Sembenelli. 2001. Market Power, Productivity, and the EU Single Market Program: Evidence from a Panel of Italian Firms. *European Economic Review* 45(1): 167-186.
- [12] Buch, Claudia, Katheryn N. Russ and Monika Schnitzer. 2011. Systemic Risk in Financial Intermediation. Manuscript, University of California, Davis.
- [13] Burdett, Kanneth and Kenneth L. Judd. 1983. Equilibrium Price Dispersion. *Econometrica* 51(4): 955-969.
- [14] Burstein, Ariel, Martin Eichenbaum and Sergio Rebelo. 2005. Large Devaluations and the Real Exchange Rate. *Journal of Political Economy* 113(4): 742-784.
- [15] Campbell, Jeffrey R. and Hugo A. Hopenhayn. "Market Size Matters," *Journal of Industrial Economics*, 2005, v53(1,Mar), 1-25.
- [16] Claessens, Stijn and Luc Laeven. 2004. What drives bank competition? *Journal of Money, Credit, and Banking* 36(3): 563-583.
- [17] de Blas, Beatriz and Katheryn N. Russ. 2011. All Banks Great, Small and Global: Loan Pricing and Foreign Competition. Manuscript, University of California, Davis.
- [18] Corsetti, Giancarlo, Luca Dedola and Sylvain Leduc. 2008. High Exchange Rate Volatility and Low Pass-Through. *Journal of Monetary Economics* 55(6): 1113-1128.
- [19] David, Herbert A. and Haikady N. Nagaraja. *Order Statistics*. Hoboken, NJ: John Wiley & Sons, Inc.
- [20] Devereux, Michael B. and Khang Min Lee. 2001. Dynamic Gains from International Trade with Imperfect Competition and Market Power. *Review of Development Economics* 5(2): 239-255.
- [21] Dixit, Avinash K. and Joseph E. Stiglitz. 1977. Monopolistic Competition and Optimum Product Diversity. *American Economic Review* 67: 297-308.
- [22] Eaton, Jonathan and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70(5): 1741-1779.
- [23] Eaton, Jonathan and Samuel Kortum. 2009. Technology in the Global Economy: A Framework for Quantitative Analysis. Manuscript, University of Minnesota.

- [24] Edmond, Christopher, Virgiliu Midrigan, and Daniel Xu. Competition and markups: Micro evidence and aggregate implications. Presentation for the July 2006 Meetings of the Society for Economic Dynamics, Istanbul.
- [25] Erdélyi, Arthur, Wilhelm Magnus, Fritz Oberhettinger, and Francesco G. Tricomi. 1954. *Tables of Integrals and Transforms*, Vol.1. New York: McGraw-Hill Book Company, Inc.
- [26] Feenstra, Robert C. and David E. Weinstein. 2010. Globalization, Markups, and the U.S. Price Level. NBER Working Paper No. 15749.
- [27] Feenstra, Robert C., Maurice Obstfeld, and Katheryn Russ. 2011. In Search of the Armington Elasticity. Manuscript, University of California, Davis.
- [28] Fitzgerald, Doireanne and Stephanie Haller. 2010. Pricing-to-Market: Evidence from Plant-Level Prices. Manuscript, Stanford University.
- [29] Foster, Lucia, John Haltiwanger and Chad Syverson. 2008. Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* (March).
- [30] Gabaix, Xavier. The Granular Origins of Aggregate Fluctuations. 2010. Forthcoming, *Econometrica*.
- [31] Garetto, Stefania. 2009. Firm Heterogeneity and Incomplete Pass-Through. Manuscript, Boston University.
- [32] Ghironi, Fabio and Marc J. Melitz. 2005. International Trade and Macroeconomic Dynamics with Heterogeneous Firms. *Quarterly Journal of Economics* 120: 865-915 (August).
- [33] Goldberg, Linda and Cedric Tille. 2009. Micro, Macro, and Strategic Forces in International Trade Invoicing. CEPR Discussion Paper 7534.
- [34] Gopinath, Gita and Oleg Itshoki. 2010. Frequency of Price Adjustment and Pass-through. *Quarterly Journal of Economics* 125(2): 675-727.
- [35] Gopinath, Gita, Oleg Itshoki, and Roberto Rigobon. 2010. Currency Choice and Exchange-Rate Pass-through. *American Economic Review* 100(1): 306-336.
- [36] Gopinath, Gita and Roberto Rigobon. 2008. Sticky Borders. *The Quarterly Journal of Economics* 123(2):531-575.
- [37] Gust, Christopher, Sylvain Leduc, and Robert Vigfusson. 2010. Trade Integration, Competition, and the Decline in Exchange-Rate Pass-Through. *Journal of Monetary Economics* 57: 309-324.
- [38] Harrison, Ann E. 1994. Productivity, Imperfect Competition and Trade Reform. *Journal of International Economics* 36(1-2): 53-73 (February).
- [39] Holmes, Thomas J., Wen-Tai Hsu, and Sanghoon Lee. 2011. Plants, Productivity, and Market Size with Head-to-Head Competition. Manuscript University of Minnesota (23 June).

- [40] Kortum, S.S. 1997. Research, Patenting, and Technological Change. *Econometrica* 65(6): 1389-1419.
- [41] Krugman, Paul R. 1980. Scale Economies, Product Differentiation, and Patterns of Trade. *American Economics Review* 70: 950-959.
- [42] Kucheryavyi, Konstantin. 2010. Continuity of a Model with a Nested CES Utility Function and Bertrand Competition. Manuscript, Pennsylvania State University.
- [43] Levinsohn, J. 1993. Testing the imports-as-market-discipline hypothesis. *Journal of International Economics* 35, pp. 122.
- [44] Malik, Henrick J. and Roger Trudel. 1982. Probability Density Function of Quotient of Order Statistics from the Pareto, Power and Weibull Distributions. *Communications in Statistics - Theory and Methods* 11(7): 801-814.
- [45] Melitz, Marc J. 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* 71: 1695-1725 (November).
- [46] Melitz, Marc J. and Giancarlo I. P. Ottaviano. 2008. Market Size, Trade, and Productivity. *Review of Economic Studies* 75(1): 295-316.
- [47] Mood, Alexander McFarlane, Franklin A. Graybill, and Duane C. Boes. 1974. *Introduction to the Theory of Statistics*. McGraw Hill Companies.
- [48] Nakamura, Emi and Jon Steinsson. 2010. Lost in Transit: Product Replacement Bias and Pricing to Market. Manuscript, Columbia University.
- [49] Novy, Dennis. 2010. International Trade and Monopolistic Competition without CES: Estimating Translog Gravity. Manuscript, University of Warwick.
- [50] Pawlas, Piotr and Dominik Szynal. 2000. *Applicationes Mathematicae* 27(2): 197-202.
- [51] Rinne, Horst. 2009. *The Weibull Distribution: a Handbook*. New York: CRC Press.
- [52] Roberts, Mark and Dylan Supina. 1996. Output Price, Markups, and Producer Size. *European Economic Review* 40(3-5): 909-21.
- [53] Rodriguez, Jose Antonio Lopez. 2010. Prices and Exchange Rates: A Theory of Disconnect. Forthcoming, *Review of Economic Studies*.
- [54] Russ, Katheryn N. and Diego Valderrama. 2010. Financial Choice in a Ricardian Model of Trade. NBER Working Paper No.15528.
- [55] Schoenle, Raphael. 2010. International Menu Costs and Price Dynamics. Manuscript, Brandeis University.
- [56] Zolas, Nicholas J. 2011. Firms' Locational Patenting Decisions. Manuscript, University of California, Davis.

APPENDIX FOR ONLINE PUBLICATION

A Deriving the distribution of markups

Malik and Trudel (1982, equation (5.17)) use Mellin transforms to derive the following distribution for the ratio $z = \frac{c_1}{c_2}$, the ratio of the first order statistic to the second, given that the sample is from the Weibull distribution $G(c)$ in the main text. This distribution is also reported in Rinne (2009, p.244, equation (5.42c)):

$$\hat{h}(z) = \frac{r(r-1)\theta z^{\theta-1}}{[(r-1) + z^\theta]^2},$$

Since we specify the (unrestricted) markup as $m = \frac{c_2}{c_1}$, we note that m is a function of z , $m = \frac{1}{z}$, and apply a straightforward Jacobian transformation:

$$\begin{aligned} h(m) &= \left(\frac{r(r-1)\theta \left(\frac{1}{m}\right)^{\theta-1}}{[(r-1) + \left(\frac{1}{m}\right)^\theta]^2} \right) \left(\frac{1}{m^2} \right) \\ &= \frac{r(r-1)\theta m^{-(\theta+1)}}{[(r-1) + m^{-\theta}]^2}. \end{aligned}$$

Similarly, Malik and Trudel (1982) use a Mellin transform to derive the following distribution for $z = \frac{c_1}{c_2}$ given a Pareto distribution of efficiency draws (a power law distribution of cost draws)

$$\hat{h}(z) = \theta z^{\theta-1}.$$

Again, we specify the (unrestricted) markup as $m = \frac{c_2}{c_1}$, implying that $m = \frac{1}{z}$, and apply the Jacobian transformation:

$$\begin{aligned} h(m) &= \frac{\theta m^{1-\theta}}{m^2} \\ &= \theta z^{-(\theta+1)} \end{aligned}$$

B Free entry under autarky

Since the distribution of markups is the same for all goods j , we drop the goods index below for simplicity. Taking (natural) logs, the expression decomposes into

$$\ln \left(1 + \frac{\beta \delta f}{L} \right) + \ln E[M^{-\sigma}] = \ln E[M^{1-\sigma}]. \quad (\text{B.1})$$

Since the natural log is a concave function, Jensen's inequality implies $E[\ln M^{1-\sigma}] \leq \ln E[M^{1-\sigma}]$ and $E[\ln M^{-\sigma}] \leq \ln E[M^{-\sigma}]$. The function $M^{-\sigma}$ has a greater degree of convexity than $M^{1-\sigma}$, so $\ln E[M^{-\sigma}] - E[\ln M^{-\sigma}] \geq \ln E[M^{1-\sigma}] - E[\ln M^{1-\sigma}]$.²⁹ This last inequality implies that

$$E[\ln M^{1-\sigma}] \geq \ln \left(1 + \frac{\beta \delta f}{L} \right) + E[\ln M^{-\sigma}],$$

as taking the log inside the expectation reduces the right-hand side more than the left-hand side. We note that for any constant k , $E[\ln M^k] = kE[\ln M]$, yielding

$$E[\ln M] \geq \ln \left(1 + \frac{\beta \delta f}{L} \right).$$

B.1 Uniqueness

Standard properties of expectations tell us that $E[M(j)^{1-\sigma}] > E[M(j)^{-\sigma}]$ for $\infty > \sigma > 1$ and $M(j) \geq 1$. In Proposition 1, we showed that $E[M(j)]$ is decreasing in the number of rivals. Thus, $E[M(j)^{1-\sigma}]$ is increasing in r and $E[M(j)^{-\sigma}]$ is increasing even faster. Thus, $E[M(j)^{1-\sigma}]/E[M(j)^{-\sigma}]$ is greater than 1 and decreasing in r toward 1, meaning that there can only be one r for which the ratio equals the constant $(1 + \frac{\beta \delta f}{L})$

²⁹Another way to see this is to note that $E[\ln M^{1-\sigma}]$, $\ln E[M^{1-\sigma}]$, $E[\ln M^{-\sigma}]$, and $\ln E[M^{-\sigma}]$ are all negative numbers, with $|\ln E[M^{1-\sigma}]| < |\ln E[M^{-\sigma}]| < |E[\ln M^{-\sigma}]|$ and $|\ln E[M^{1-\sigma}]| < |E[\ln M^{1-\sigma}]| < |E[\ln M^{-\sigma}]|$. Thus, switching the logs from outside to inside the expectation in equation (B.1) reduces the left hand side more than the right hand side.

B.2 Upper- and lower- bounds for the number of rivals.

The distribution of the markup does not yield a closed-form solution for the expected markup $E[M]$ or for the expected log markup, $E[\ln M]$. However, we know from Proposition 1 that the mean markup $E[M]$ is decreasing in r . Therefore, we determine an upper- and lower- bound for r . Specifically, we can express the minimum number of rivals as a function of the expected log markup and derive a clean closed-form solution for the maximum number of rivals. Let $V = \ln M$. Then the probability density for V is a simple transformation of $h(m)$,

$$\begin{aligned} h_V(v) &= e^v h(e^v) I_{\mathbb{R}_+}(v) \\ &= e^v \frac{r(r-1)\theta (e^v)^{-(\theta+1)}}{\left[(r-1) + (e^v)^{-\theta}\right]^2}. \end{aligned}$$

The probability that $V \geq \bar{m}$ (or any other positive constant) is then

$$\int_{\ln(\bar{m})}^{\infty} e^v \frac{r(r-1)\theta (e^v)^{-(\theta+1)}}{\left[(r-1) + (e^v)^{-\theta}\right]^2} dv = \frac{r}{1 + (r-1)e^{\theta\bar{m}}}.$$

Using a generalized version of Chebyshev's inequality³⁰, we can characterize a lower-bound for the number of rivals:

$$\begin{aligned} \bar{m} \Pr[\ln M \geq \bar{m}] &\leq E[\ln M] \\ \frac{r\bar{m}}{1 + (r-1)e^{\theta\bar{m}}} &\leq E[\ln M] \\ r &\geq \frac{E[\ln M](e^{\theta\bar{m}} - 1)}{E[\ln M]e^{\theta\bar{m}} - \bar{m}}. \end{aligned}$$

As noted previously, the expected markup and the number of rivals is inversely related, a relationship seen here in the lowerbound for r . When $E[M]$ falls, the lowerbound increases, reflecting the fact that more rivals will enter when the

³⁰See Theorem 5 in Mood, Graybill, and Boes (1974, p.71): For a random variable X , a nonnegative function $g(\cdot)$, and a scalar $k > 0$, then $kP[g(X) \geq k] \leq E[g(X)]$.

expected markup is high (and vice versa). We know from equation (10) that the expected log gross markup $E[\ln M]$ must be at least as large as the gross log per-period cost of production, $\ln(1 + \frac{\beta\delta f}{L})$, producing an upperbound for r . Thus, we know that r lies within the following bounds:

$$\frac{\ln\left(1 + \frac{\beta\delta f}{L}\right) (e^{\theta\bar{m}} - 1)}{\ln\left(1 + \frac{\beta\delta f}{L}\right) e^{\theta\bar{m}} - \bar{m}} \geq r \geq \frac{E[\ln M](e^{\theta\bar{m}} - 1)}{E[\ln M]e^{\theta\bar{m}} - \bar{m}}. \quad (\text{C.1})$$

C The distribution of markups under trade

To calculate the unconditional probability that both the first and second best suppliers of a good to country n are from country i , we start from the main text:

$$\psi_{ni} = \int_0^\infty \int_{c_1}^\infty g_{2ni}(c_{2i}) \prod_{u \neq i}^N [1 - G_{1nu}(c_{2i})] dc_{2i}$$

The first step is to derive the marginal distribution $g_{2ni}(c_{2i})$ from the joint distribution, which is analogous to the joint distribution under autarky but including trade costs. Integrating from the lower limit c_{1i} , we have

$$\begin{aligned} g_{2ni}(c_{2i}) &= \int_{c_{1i}}^\infty g_{1ni,2ni}(c_{1i}, c_{2i}) dc_{1i} \\ &= \int_{c_{1i}}^\infty r_i(r_i - 1) [\theta T_i(w_i d_{ni})^{-\theta}]^2 c_{1i}^{\theta-1} c_{2i}^{\theta-1} e^{-T_i(w_i d_{ni})^{-\theta} c_{1i}^\theta} e^{-(r_i-1)T_i(w_i d_{ni})^{-\theta} c_{2i}^\theta} dc_{1i} \\ &= r_i(r_i - 1) \theta T_i(w_i d_{ni})^{-\theta} c_{2i}^{\theta-1} e^{-(r_i-1)T_i(w_i d_{ni})^{-\theta} c_{2i}^\theta} \left(1 - e^{-T_i(w_i d_{ni})^{-\theta} c_{2i}^\theta}\right) \end{aligned}$$

Substituting into the formula for ψ_{ni} yields

$$\begin{aligned}
\psi_{ni} &= \int_0^\infty r_i(r_i - 1)\theta T_i(w_i d_{ni})^{-\theta} c_{2i}^{\theta-1} e^{-(r_i-1)T_i(w_i d_{ni})^{-\theta} c_{2i}^\theta} \left(1 - e^{-T_i(w_i d_{ni})^{-\theta} c_{2i}^\theta}\right) e^{-[\Phi_n - r_i T_i(w_i d_{ni})^{-\theta}] c_{2i}^\theta} dc_{2i} \\
&= \int_0^\infty r_i(r_i - 1)\theta T_i(w_i d_{ni})^{-\theta} c_{2i}^{\theta-1} e^{-[\Phi_n - r_i T_i(w_i d_{ni})^{-\theta} + (r_i-1)T_i(w_i d_{ni})^{-\theta}] c_{2i}^\theta} dc_{2i} \\
&\quad - \int_0^\infty r_i(r_i - 1)\theta T_i(w_i d_{ni})^{-\theta} c_{2i}^{\theta-1} e^{-[\Phi_n - r_i T_i(w_i d_{ni})^{-\theta} + (r_i-1)T_i(w_i d_{ni})^{-\theta} + T_i(w_i d_{ni})^{-\theta}] c_{2i}^\theta} dc_{2i} \\
&= \frac{r_i(r_i - 1)T_i(w_i d_{ni})^{-\theta}}{\Phi_n - T_i(w_i d_{ni})^{-\theta}} (-1) e^{-[\Phi_n - T_i(w_i d_{ni})^{-\theta}] c_{2i}^\theta} \Big|_0^\infty - \frac{r_i(r_i - 1)T_i(w_i d_{ni})^{-\theta}}{\Phi_n} (-1) e^{-\Phi_n c_{2i}^\theta} \Big|_0^\infty \\
&= r_i(r_i - 1)T_i(w_i d_{ni})^{-\theta} \left(\frac{1}{\Phi_n - T_i(w_i d_{ni})^{-\theta}} - \frac{1}{\Phi_n} \right) \\
&= \frac{r_i T_i(w_i d_{ni})^{-\theta}}{\Phi_n} \frac{(r_i - 1)T_i(w_i d_{ni})^{-\theta}}{[\Phi_n - T_i(w_i d_{ni})^{-\theta}]}.
\end{aligned}$$

We can also derive the unconditional probability that the first and second best rivals to supply a good to country n are, respectively, from i and $u \neq i$. We start with the formula (equation (14)) in the main text,

$$\psi_{niu} = \int_0^\infty \int_{c_{1i}}^\infty \left(\int_{c_{1i}}^{c_{2i}} g_{1nu}(c_{1u}) \prod_{v \neq i, u}^{N-1} [1 - G_{1nv}(c_{1u})] dc_{1u} \right) g_{1ni, 2ni}(c_{1i}, c_{2i}) dc_{2i} dc_{1i}.$$

Define A as the inner integral,

$$\begin{aligned}
A &:= \int_{c_{1i}}^{c_{2i}} g_{1nu}(c_{1u}) \prod_{v \neq i, u}^N [1 - G_{1nv}(c_{1u})] dc_{1u} \\
&= \int_{c_{1i}}^{c_{2i}} \theta r_u T_u c_{1u}^{\theta-1} (w_u d_{nu})^{-\theta} e^{-r_u T_u (w_u d_{nu})^{-\theta} c_{1u}^\theta} e^{-[\Phi_n - r_i T_i(w_i d_{ni})^{-\theta} - r_u T_u (w_u d_{nu})^{-\theta}] c_{1u}^\theta} dc_{1u} \\
&= \frac{-r_u T_u (w_u d_{nu})^{-\theta}}{\Phi_n - r_i T_i(w_i d_{ni})^{-\theta}} e^{-[\Phi_n - r_i T_i(w_i d_{ni})^{-\theta}] c_{1u}^\theta} \Big|_{c_{1i}}^{c_{2i}} \\
&= \psi'_{niu} \left(e^{-[\Phi_n - r_i T_i(w_i d_{ni})^{-\theta}] c_{1i}^\theta} - e^{-[\Phi_n - r_i T_i(w_i d_{ni})^{-\theta}] c_{2i}^\theta} \right), \tag{D.1}
\end{aligned}$$

where we define $\psi'_{niu} = \frac{r_u T_u (w_u d_{nu})^{-\theta}}{\Phi_n - r_i T_i (w_i d_{ni})^{-\theta}}$ as in the main text.

We then define B as the integral with the first half of A ,

$$\begin{aligned}
B &:= \psi'_{niu} \int_0^\infty \int_{c_1}^\infty e^{-[\Phi_n - r_i T_i (w_i d_{ni})^{-\theta}] c_{1i}^\theta} g_{1ni, 2ni}(c_{1i}, c_{2i}) dc_{2i} dc_{1i} \\
&= \psi'_{niu} \left(\frac{(r_i - 1) T_i (w_i d_{ni})^{-\theta}}{\Phi_n - T_i (w_i d_{ni})^{-\theta}} \right) \left(\frac{r_i T_i (w_i d_{ni})^{-\theta}}{\Phi_n} \right) \\
&= \psi'_{niu} \psi'_{ni} \pi_{ni}
\end{aligned} \tag{D.2}$$

with $\psi'_{ni} = \frac{(r_i - 1) T_i (w_i d_{ni})^{-\theta}}{\Phi_n - T_i (w_i d_{ni})^{-\theta}}$, also as in the main text. Let C be the integral with the second half of A , given by

$$\begin{aligned}
C &:= \psi'_{niu} \int_0^\infty \int_{c_1}^\infty e^{-[\Phi_n - r_i T_i (w_i d_{ni})^{-\theta}] c_{2i}^\theta} g_{1ni, 2ni}(c_{1i}, c_{2i}) dc_{2i} dc_{1i} \\
&= \psi'_{niu} \left(\frac{r_i T_i (w_i d_{ni})^{-\theta}}{\Phi_n} \right) \\
&= \psi'_{niu} \pi_{ni}
\end{aligned} \tag{D.3}$$

Finally we combine the two components to compute ψ_{niu} ,

$$\psi_{niu} = C - B = \psi'_{niu} \pi_{ni} (1 - \psi'_{ni})$$

Since $\sum_{u \neq i}^N \psi'_{niu} = 1$, it is clear that

$$\begin{aligned}
\sum_{i=1}^N \psi_{ni} + \sum_{i=1}^N \sum_{u \neq i}^N \psi'_{niu} \pi_{ni} (1 - \psi'_{ni}) &= \sum_{i=1}^N \psi_{ni} + \sum_{i=1}^N \pi_{ni} (1 - \psi'_{ni}) \\
&= \sum_{i=1}^N \psi_{ni} + \sum_{i=1}^N \pi_{ni} - \sum_{i=1}^N \psi_{ni} \\
&= \sum_{i=1}^N \pi_{ni} \\
&= 1
\end{aligned}$$

D Gains from trade

To close the model under autarky or trade, we use a market clearing condition. Let λ_D be the share of variable costs in profits for each country, given the vector of trade costs D that it faces when exporting. We can use the free entry condition to show that under autarky, λ_D equals $\frac{1}{1+\frac{\delta\beta f}{L}}$. Similarly, under free trade with symmetric countries, λ equals $\frac{1}{1+\frac{\delta\beta f}{L(1+\frac{N-1}{N})}}$. Given our unit cost specification, the share of labor in these variable costs is β . Then, the labor market clearing condition stipulates that payments to labor equal labor's share in production costs:

$$\omega_n L_n = \beta \lambda P_n Y_n.$$

We use the wage as our numeraire, $\omega \equiv 1$. Then, we can compare output under autarky with output under free trade in a world with N symmetric countries:

$$\frac{Y^t}{Y^a} = \left(\frac{1 + \frac{\delta\beta f}{L}}{1 + \frac{\delta\beta f}{L(1+\frac{N-1}{N})}} \right) \left(\frac{P^a}{P^t} \right).$$

The first term on the right-hand side is less than one and reflects the fact that aggregate revenues and average firm profits fall under trade versus autarky because opening to foreign competition squeezes markups. We already know from Propositions 2 and 3a that the autarkic price level is greater than the price level under free trade. To find out how much greater, we must substitute in our formulas for the aggregate price level under autarky and free trade,³¹

$$\frac{P^a}{P^t} = \left(\frac{(1 + (R-1)\bar{m}^\theta) \left\{ \bar{m}^\theta r^{\frac{\sigma-1}{\theta}+1} + (r-1)(\bar{m}^\theta - 1) \left[r(r-1)^{\frac{\sigma-1}{\theta}} - (r-1)r^{\frac{\sigma-1}{\theta}} \right] \right\}}{(1 + (r-1)\bar{m}^\theta) \left\{ \bar{m}^\theta R^{\frac{\sigma-1}{\theta}+1} + (R-1)(\bar{m}^\theta - 1) \left[R(R-1)^{\frac{\sigma-1}{\theta}} - (R-1)R^{\frac{\sigma-1}{\theta}} \right] \right\}} \right)^{\frac{1}{1-\sigma}}$$

Even under symmetry, the level of gains from trade clearly depends upon the number of domestic rivals before liberalization. In Figure 3, we show that they are lower for countries with a high level of contestability (r^a) to begin with, as

³¹If we do not normalize the wage ω to equal 1, this expression is the ratio of the real wage under trade, relative to the real wage under autarky.

these countries already have lower average markups than their trading partners.