European Economic Review 70 (2014) 276–298

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Market size, structure, and access: Trade with capacity constraints

Anson Soderbery

Department of Economics, Purdue University, West Lafayette, IN 47907-2056, United States

ARTICLE INFO

Article history:
Received 12 November 2013
Accepted 3 June 2014
Available online 16 June 2014

JEL classification:
F12
F13
L11
L13

Keywords:
Trade
Heterogeneous firms
Capacity constraints
Market access

ABSTRACT

This paper develops a model of international trade where firms are heterogeneous across capacity and productivity. A binding capacity constraint induces firms to raise prices in order to take advantage of access to new markets. This generates markets with a flexible competitive structure giving rise to instances where trade and trade liberalization negatively impact welfare. Its key predictions can be identified by observing the presence of small yet highly productive firms and substitution by firms across markets as accessibility evolves. Using Thai firm-level data I establish the prevalence of these anomalous firms and demonstrate they face capacity constraints.

1. Introduction

With finite resources firms face physical constraints on production. The question is not whether these capacity constraints exist, but rather how firm behavior changes when they bind. Overcoming a capacity constraint is costly, either explicitly through investment to produce beyond the constraint or implicitly by weakening a firm's competitiveness. It is well known that firms adapt their pricing and production decisions when capacity constraints bind. However, the implications of firms facing capacity constraints on international trade have been widely ignored by economists.

Conversely, policy makers and trade commissions are well aware of how capacity constraints hinder gains from trade. Each annual report from 2011 to 2013 produced by the *United Nations Economic and Social Commission for Asia and the Pacific* highlights the role of production capacity on the ability of developing industries to succeed in global markets.[1] Also, recent *Organization for Economic Cooperation and Development* reports, including OECDReport (2012) and Hallaert and Munro (2010), focus on the importance of understanding and addressing the impact of production capacity on trade by developing industries in low income countries. This paper models the welfare implications of firms facing capacity constraints with international trade and utilizes firm level data from Thailand to support the theoretical predictions. The results presented here corroborate the claims from institutions tasked with promoting global development that capacity plays a significant role in the ability of countries to fully benefit from globalization. The model also points to gains to the trading partners of

E-mail address: asoderbe@purdue.edu

[1] UNESCAP (2011) devotes an entire chapter to the discussion of the need to address capacity constraints in the least developed industries if they are to succeed in global markets.

http://dx.doi.org/10.1016/j.euroecorev.2014.06.003

constrained countries. Capacity constrained firms are shown to have strong internal linkages across destinations and thus pass any distortionary behavior to each of their markets.[2]

In standard heterogeneous firm models of trade, production decisions are completely flexible. In reality, capacity costs put real constraints on firms' production and pricing decisions. This paper adds the real-world feature of costly capacity by developing a model of international trade where firms are heterogeneous not only in productivity, but also capacity. Firms interact in a monopolistically competitive market for their differentiated varieties. They set markups endogenously given their production capabilities and demand for their good. A firm cannot produce beyond its underlying capacity without undertaking costly expansion. When capacity constraints do not bind, the firm freely sets price and quantity to maximize profits given market conditions (number of competitors, trade costs, and market size in origin and destination countries). On the other hand, firms facing a binding capacity constraint cannot freely expand production to access markets. Their only margin of adjustment comes through raising prices or making costly investment. This limits the competitive responses available to the constrained firm as market conditions change.

As markets grow firms approach their underlying capacity constraint. Since firms producing at capacity are unable to expand production without incurring additional marginal costs, they take advantage of access to larger markets by raising prices. The presence of capacity constrained firms leads to a market-wide softening of competition such that average prices can increase with market size even in the face of tougher selection on productivity.

International trade instantly increases the size of the market available each firm. Consider an unconstrained firm utilizing 70% of its capacity to serve the domestic market. Suppose this firm is efficient enough to export and would optimally double production to serve the international market. This firm's capacity constraint binds before satisfying export demand. Since it cannot access the market by expanding sales, the firm adjusts its pricing strategy. Optimally the firm raises the price of its variety in each market as it either passes through the explicit cost of capacity expansion or absorbs the implicit cost of capacity to slow production.

This logic implies that access to international markets generates an opportunity cost – foregoing foreign sales – associated with domestic sales by capacity constrained firms. This opportunity cost implies substitution across markets as global conditions change. A growing foreign market (or a decrease in trade costs) causes the marginal benefit to constrained firms from selling abroad to rise. As the marginal benefit of exporting grows, constrained firms absorb the opportunity cost of foregoing domestic sales and substitute sales from the domestic market to the foreign market. Sudden shocks to market size from opening trade thus allow us to understand the linkages across markets as they are transmitted through capacity constraints.

Constrained firms substituting across destinations shift the weight of their pricing distortions across markets. Trade thus amplifies the distortions caused by firms facing capacity constraints due to the immediate demand for exporters to increase production. These results give a palatable explanation for why many countries are resistant to trade liberalization. Firms facing binding capacity constraints erode the expected gains from trade that come through enhanced price competition. This result calls back to the “second best results” of Bhagwati and Ramaswami (1963). As in the second best literature, I demonstrate that markets may be distorted to such an extent that consumer welfare falls with trade and trade liberalization.

The key predictions of the model can be identified by observing the presence of small yet highly productive firms and substitution by firms across markets as the relative ease to access markets changes at the firm-level. Firm-level data from Thailand allow me to estimate the substitution patterns of firms in light of changes in implicit marginal costs and relative accessibility of markets from tariff movements. My model demonstrates that declining trade barriers lead to increasing opportunity costs of domestic sales for capacity constrained exporters. Implicit opportunity costs imply a feedback effect of trade liberalization which decreases domestic sales that is magnified for firms facing tight constraints. I explore these predictions by decomposing the impact of tariff changes on sales across domestic and foreign markets. I present evidence that the impact of tariffs on domestic sales depends on both the tightness of a firm's capacity constraint and its presence in export markets. Furthermore, I demonstrate throughout my analysis that more productive firms have higher sales after controlling for the effects of capacity. My model also explains the existence of high productivity firms with relatively low sales. I show these firms are prevalent in the data and that their existence can be explained by capacity constraints.

By allowing firms to endogenously set markups, my model demonstrates the impact of capacity constraints on market structure. Melitz and Ottaviano (2008) provide a tractable framework where firms set markups endogenously in order to explain some stylized facts linking market size and competition. Syverson (2007) establishes empirically that larger markets are occupied by firms charging relatively lower markups and prices as a consequence of greater goods market competition and tougher selection. Holding the number of capacity constrained firms constant, my model maintains these results. However, in larger markets constrained firms become more prevalent, and since they access markets through prices, it is possible to see higher average prices in larger markets. This result is similar to Staiger and Wolak (1992) who assert that between markets of identical size, the market with scarce capacity will display higher prices and weaker competition.

The industrial organization literature examining market outcomes in the presence of capacity constraints is vast. Levitan and Shubik (1972) establish the effects of fixed capacity levels on Bertrand duopolies. Spence (1977) demonstrates the

[2] Although not explicitly examined here, the linkages developed in this model across markets within firms have the potential to yield insight into our understanding of how market characteristics such as exchange rates and shocks to supply or demand are passed through by firms. Understanding the source of these linkages would aid in investigating how firm behavior shapes fundamentals such as inflation and elasticities of trade.

possible anti-competitive effects of fixed capacity in a Cournot-type model. In the model, an incumbent firm competes with entrants following its choice to install a fixed capacity level.[3] Modeling capacity constraints has continued to be relevant in industrial organization. The recent work of Knittel and Lepore (2009) analyzes tacit collusion and price wars between firms facing endogenous capacity constraints by building on Staiger and Wolak (1992). Empirically, Suslow (1986) and Bresnahan and Suslow (1989) estimate that firms in the primary aluminum and ammonia fertilizer industries face a flat short-run marginal cost curve below capacity that becomes vertical at capacity. The implications of such capacity constraints can profoundly impact firms' optimal strategies when interacting in a competitive marketplace. I further these studies by showing the change in competition induced by capacity constraints has a far-reaching impact on globalization.

Despite the extensive industrial organization literature on capacity constraints, there has been little analysis of the role of capacity constraints in firm-level decisions related to international markets. Maggi (1996) discusses the effects of "soft" capacity constraints in a three-country duopoly model with homogeneous firms. His work demonstrates that the distortion of competition from capacity constraints has substantial impact on market outcomes, especially in light of globalization, even when firms are homogeneous. The recent works of Vannoorenberghe (2012), Blum et al. (2013) and Liu (2012) impose fixed capital on heterogeneous firms with Cobb–Douglas production, which can explain substitution of sales across destinations.[4] However, firms in these models can reach any finite level of production at increasing marginal costs. Consequently, these models cannot generate the small efficient firms or weak price competition in large markets I highlight, which are central to understanding potential losses from trade and trade liberalization.

This paper continues as follows. The next section discusses the theory in autarky and ends by parameterizing and analyzing the model in Sections 2.4 and 2.5. Section 3 opens the model to international trade and elaborates on the implications of capacity constraints on global markets. Trade is parameterized as being between fully symmetric countries in Section 3.3.1. Section 4 discusses the data, empirical strategy, and estimation used to support the model. Section 5 concludes.

2. Closed economy

This section summarizes the demand-side of Melitz and Ottaviano (2008) and develops a supply-side in which heterogeneous producers of differentiated goods face capacity constraints as they compete in a monopolistically competitive market.

2.1. Consumer preferences

The economy is made up of L identical consumers with linear quadratic utility,

$$U^c = q_0^c + \alpha \int_{i\in\Omega} q_i^c\, di - \frac{1}{2}\gamma \int_{i\in\Omega} (q_i^c)^2\, di - \frac{1}{2}\eta \left(\int_{i\in\Omega} q_i^c\, di \right)^2. \tag{1}$$

Consumption of the numeraire and each variety i from the set of all varieties Ω is given by q_0^c and q_i^c. Demand is structured by the parameters α, η and γ, which are all assumed to be strictly positive. The degree of product differentiation is indexed by γ. Varieties are more substitutable as γ approaches zero and in the limit are perfect substitutes. Substitution between differentiated varieties and the numeraire is characterized by α and η.

Since marginal utilities are bounded, positive demand for each variety is not guaranteed.[5] For simplicity, assume positive demand for the numeraire good ($q_0^c > 0$). Utility maximization yields the linear inverse demand across varieties, $p_i = \alpha - \gamma q_i^c - \eta \mathcal{Q}^c$, for $q_i^c \geq 0$. Positive demand along with identical consumers allows the equation to be inverted and aggregated, yielding

$$q_i = Lq_i^c = \frac{\alpha L}{\eta N + \gamma} - \frac{L}{\gamma} p_i + \frac{\eta N}{\eta N + \gamma} \frac{L}{\gamma} \overline{\mathcal{P}}, \quad \forall i \in \Omega^*. \tag{2}$$

The average price of the N consumed varieties in $\Omega^* \subset \Omega$ is $\overline{\mathcal{P}} = 1/N \int_{i\in\Omega^*} p_i\, di$. Varieties satisfying the positive demand pricing condition,

$$p_i \leq \frac{\gamma\alpha + \eta N \overline{\mathcal{P}}}{\eta N + \gamma} \equiv p_{max}, \tag{3}$$

make up the set Ω^*. Firms unable to profitably charge a price below p_{max} will exit without producing.

[3] There is also a vast literature allowing costly capacity expansion and its effects on competition. For example, Dixit (1980) demonstrates similar competitive results to those of Spence (1977) with "soft" capacity levels.

[4] Vannoorenberghe (2012) is foremost concerned with linkages across markets. Blum et al. (2013) introduce mild dynamics to explain firm entry and exit patterns from foreign markets. Liu (2012) focuses heavily on the dynamics of firm adjustment to study investment patterns.

[5] This is a notable departure from traditional new trade theory models where Constant Elasticity of Substitution (CES) preferences are most common (as seen in Melitz, 2003). Foregoing CES preferences in this model allows tractable modeling of endogenous markups. This is beneficial in the theory to follow since it allows firms to adjust their pricing equation as the tightness of their capacity constraints respond to changes in competition.

The effects of aggregate market outcomes on individual welfare are characterized by the indirect utility function,

$$U^c = I^c + \frac{1}{2}\left(\eta + \frac{\gamma}{N}\right)^{-1}(\alpha - \overline{\mathcal{P}})^2 + \frac{1}{2}\frac{N}{\gamma}\sigma_p^2. \tag{4}$$

Consumer c's income is I^c, and $\sigma_p^2 = 1/N\int_{i\in\Omega^*}(p_i - \overline{\mathcal{P}})^2\, di$ measures the variance of prices across varieties. Positive demand for the numeraire implies $I^c > \int_{i\in\Omega^*} p_i q_i^c\, di = \overline{\mathcal{P}}\mathcal{Q}^c - N\sigma_p^c/\gamma$, for all consumers. Utility clearly increases with declines in the average price level and, given some price level, increases with the variance of prices. Intuitively, consumers are better off with more varieties at a wide range of prices and lower average price levels.

2.2. Capacity constraints and firm behavior

Labor is the only input used to produce the differentiated good and is inelastically supplied from a competitive market. The assumptions behind the numeraire good and labor market together imply a unit wage. Differentiated goods are produced using a constant returns to scale technology with marginal cost c.[6] Firms producing differentiated goods must undertake costly product research and development before entering the market.

Firms pay a one-time irreversible investment f_E to enter the differentiated goods market. Entry occurs in two stages. In stage one, before realizing their marginal costs, firms install capacity for production. In stage two, firms complete research and development which produces a blueprint for their variety. The characteristics of their variety include its marginal cost of production and maximum production given the installed capacity. Firm i thus realizes random draws of marginal cost of production (c_i) and plant capacity (K_i) from the common, known and independent distributions $G(c)$ and $H(K)$ over the strictly positive supports $[0, c_M]$ and $[0, K_{max}]$. Capacity defines the quantity a given firm can produce at its constant marginal cost c_i. In the event that firms under-invest in capacity, they can rent capacity at a constant per unit cost r such that each unit produced beyond K_i is done so at marginal cost $c_i + r$.[7] One could also motivate firm heterogeneity in productivity and capacity by assuming that paying f_E provides the firm with a manager whose abilities translate into two unique skills: depth and breadth.

Managerial depth translates into productive efficiency, and breadth translates into scale of production.[8]

This model abstracts away from firms making dynamic endogenous investments in capacity for two specific reasons. Section 5.2 demonstrates that Thai firm-level data is made up of firms that perpetually over- and under- invest in capacity regardless of their industry, level of sales, and other pertinent firm level characteristics. Second, the variation of capacity utilization within firms is extremely persistent, depending upon previous utilization, sales variation, and export status, rather than choice variables such as investment. These observations support the medium-run model where firms face uncertainty in their productivity and production capacity that follows. Understanding the static equilibrium effects of capacity constraints are important to shedding light on the current hurdles faced by countries when their firms are capacity constrained.[9]

The firm's problem translates into three unique cases pictured in Fig. A1. Panel (a) is the canonical case where the firm maximizes profit at its marginal cost c and faces a non-binding capacity constraint. Panels (b) and (c) are firms facing binding capacity constraints and maximizing profits at higher marginal costs. The following should be thought of as a medium-run model, as firms can marginally, but not permanently, adjust their capacities with a long enough horizon to realize a free entry equilibrium.[10] Next I characterize the distributions of each type of firm in the economy, how firms are endogenously sorted by type, and why their competitive behavior differs.

Surviving firms are those which can cover their marginal cost regardless of whether they are constrained by their capacity level or not. All others exit without producing. Surviving firms compete in a monopolistically competitive market for differentiated goods. Consequently, each firm takes the number of competitors N and price level $\overline{\mathcal{P}}$ as given. They choose

[6] I assume that the numeraire good is sold in a perfectly competitive market by an external set of producers unable to access the differentiated goods sector.

[7] This set-up is most closely related to the pioneer models of Spence (1977), Dixit (1979), and Kreps and Scheinkman (1983) where firms must commit to capacity before production. Notably Besanko and Doraszelski (2004) and others have extended the study of capacity to account for dynamic decisions by firms. This paper highlights the consequences of flexible competitive behavior by firms accessing markets conditional on an underlying friction. The focus of the model here is to understand medium-run outcomes where an economy is populated with firms that have over- and under-invested in capacity, as I observe in the data. Adding dynamics would explain how firms transition from the medium-run to the long-run, but would not aid in the observations of the paper.

[8] There are many other ways to motivate the presence of capacity constraints as I have defined them. For instance, a firm's capacity may be randomly determined with incomplete information in the lending market. Lenders uncertain of the firm's productivity may under- or over-invest in the firm's capacity without the firm becoming aware of its allotment until entry. Additionally, randomly realized capacity is in line with the "resource-based view" of production, which has a rich history in managerial economics (Wernerfelt, 1984) and is gaining traction in economics (Nocke and Yeaple, 2006). The mapping of resources to productive capabilities from the "resource-based view" also provides a nice interpretation in the context of developing countries. Adopting this view can motivate capacity constraints resulting from a deficiency in resources inherent in producing at large scale.

[9] In the Technical Appendix, I provide an alternative explanation for capacity that centers around firms endogenously pre-committing to capacity when there is demand uncertainty. The model with firms facing uncertain market outcomes and making endogenous capacity decisions is observationally equivalent to the following.

[10] One could also characterize the run of the model as a function of the capacity adjustment cost. When r is large no firm will adjust capacity, as we would expect in the short-run, and as r approaches zero all capacity constrained firms adjust, as we would expect in the long-run.

price to maximize profit using the residual demand from (2) subject to their realized capacity constraint K and marginal cost c. The profit maximization problem and first order conditions for firm i are given by

$$\arg\max_{p_i}\begin{Bmatrix} (p_i-c_i)q_i \ |q_i=\min\{q_i(p_i),K_i\} & \text{for } q_i\le K_i \\ (p_i-c_i-r)q_i+rK_i|q_i=q_i(p_i) & \text{for } q_i>K_i \end{Bmatrix}$$

$$FOC\begin{cases} p_i=c_i+\frac{\gamma}{L}q_i & \text{for } q_i<K_i \\ p_i=c_i+\lambda_i+\frac{\gamma}{L}K_i & \text{for } q_i=K_i \\ p_i=c_i+r+\frac{\gamma}{L}q_i & \text{for } q_i>K_i. \end{cases} \tag{5}$$

Let the implicit marginal costs, the sum of actual marginal costs, and the shadow value (λ_i) or actual value of capacity (r), be defined as $\kappa_i\equiv c_i+\lambda_i$ or $\kappa_i\equiv c_i+r$ for the appropriate value of q_i.

Firms with optimal price above p_{max} will exit. Let $_D$ index the firm that is indifferent between producing and not. This firm's price equals its marginal cost, and faces zero demand for its variety. Consequently, $p(c_D)=c_D=p_{max}$ and $q(c_D)=0$. Firm behavior can be characterized solely by implicit marginal costs and firm D's characteristics. Combining firm i's optimal behavior and residual demand yields, respectively, the optimal price, markup over c_i, and quantity:

$$p(\kappa_i)=\tfrac{1}{2}(c_D+\kappa_i), \tag{6}$$

$$\mu(\kappa_i)=\tfrac{1}{2}(c_D+\kappa_i)-c_i, \tag{7}$$

$$q(\kappa_i)=\frac{L}{2\gamma}(c_D-\kappa_i). \tag{8}$$

The first K_i units sold by adjusting firms are produced with marginal cost c_i, and each subsequent unit is produced at cost c_i+r. Therefore, the profits of a firm choosing to adjust its capacity level reflect the cost savings it receives from possessing the preinstalled capacity level K_i. Conversely, firms that do not adjust face constant *actual* marginal costs c_i regardless of capacity level. Therefore, profits are given by

$$\pi(\kappa_i)=\begin{cases} \frac{L}{4\gamma}(c_D-\kappa_i)(c_D+\kappa_i-2c_i) & \text{if } q_i\le K_i \\ \frac{L}{4\gamma}(c_D-\kappa_i)^2+rK_i & \text{if } q_i>K_i. \end{cases} \tag{9}$$

Lower cost firms charge lower prices, sell higher quantities, and make greater profits. In this model costs may be relatively low due to favorable productivity or capacity draws.

Random draws of cost and capacity imply a random distribution for implicit marginal costs characterized by cutoff restrictive capacity levels $\underline{K_i}$ and $\overline{K_i}$, defined as follows. Any firm drawing $K_i\le\overline{K_i}$ faces a binding capacity constraint. Firms drawing K_i below $\overline{K_i}$ but above $\underline{K_i}$ do not find it optimal to adjust their capacity level. Firms drawing K_i below $\underline{K_i}$ adjust their capacity level to serve more of the market. Each firm's specific productivity, combined with market characteristics, endogenously determines its cut-off capacity levels $\underline{K_i}$ and $\overline{K_i}$, which are characterized by

$$MR\left(\underline{K_i}\right)=c_i+r\Longrightarrow p-\frac{\gamma}{L}\underline{K_i}=c_i+r\Longrightarrow\underline{K_i}(c_i,r)=\frac{L}{2\gamma}(c_D-c_i-r) \quad \text{and}$$

$$MR\left(\overline{K_i}\right)=c_i\Longrightarrow p-\frac{\gamma}{L}\overline{K_i}=c_i\Longrightarrow\overline{K_i}(c_i)=\frac{L}{2\gamma}(c_D-c_i). \tag{10}$$

Fig. A1 illustrates the firm's problem and demonstrates implicit marginal costs strictly exceed explicit marginal cost when the firm is constrained. Panel (a) shows a firm with marginal cost c_i and a capacity level K_i that is not binding. Panel (b) displays a level of capacity where the same firm is *just* capacity constrained ($\underline{K_i}\le K_i<\overline{K_i}$). The just constrained firm faces a cost of adjustment r that precludes it from investing in capacity, as it is not profitable to serve a greater portion of the market at marginal cost $c+r$. Panel (c) presents the case of a firm with the same cost level c_i, but a draw of K_i that is particularly unfavorable ($K_i<\underline{K_i}$). This firm is also capacity constrained, but will profitably adjust its capacity level by paying the added per unit cost r in order to sell $q(c_i+r)$ rather than K_i units. Notice from (10), firms with high productivity in large industries where differentiated goods are less substitutable face tighter capacity constraints.

Implicit marginal costs (κ_i) are endogenously determined by the cut-off cost level (c_D), capacity (K_i), and marginal and adjustment costs (c_i and r) faced by the firm. All else equal, firms facing a binding constraint charge higher prices and set higher *actual* markups than their unconstrained counterparts. However, they make lower profits as the quantity restriction from the constraint overshadows any pricing gains. This can be seen immediately from the implicit marginal cost equation, which can be written in the piecewise linear form as a function of the cost cut-off and the firm's capacity and cost draws,

$$\kappa_i\left(c_i, c_D, K_i, \underline{K_i}, \overline{K_i}, r\right) \equiv \begin{cases} c_i + \lambda_i = \begin{cases} c_i & \text{for } \overline{K_i} \le K_i \\ c_D - \dfrac{2\gamma}{L} K_i & \text{for } \underline{K_i} \le K_i < \overline{K_i} \end{cases} \\ c_i + r & \text{for } K_i < \underline{K_i}. \end{cases} \tag{11}$$

Clearly marginal costs are increasing as the capacity constraint tightens (K_i falls). Since prices and markups are monotonically increasing in a firm's implicit marginal costs, constrained firms sell at higher prices and receive larger actual markups. It is also straightforward to show profits are strictly decreasing in implicit marginal costs. Consequently, constrained firms make lower profits, which further decrease as their constraint tightens, than unconstrained firms.

Capacity constraints explain the existence of highly efficient producers that are smaller and less profitable than their productivity suggests. The largest firms are thus the most efficient in terms of both costs and capacity. This outcome is similar to Clementi and Hopenhayn (2006), where with contracting uncertainty in lending, efficient firms may produce at less than optimal scale due to an inability to acquire their optimal level of capital. In my model, firms producing at suboptimal scale face the additional hinderance that they cannot set competitive prices for their output.

2.3. Free entry equilibrium

In order to understand the welfare consequences of capacity constraints beyond the existence of firms producing at suboptimal scale, the model now describes entry and exit in response to capacity. Given (9) and (11), the expected profit of a firm prior to entry is defined as

$$\begin{aligned} \overline{\Pi} &\equiv \int_0^{c_D} \int_0^{K_{max}} \pi(\kappa)\, dH(K)\, dG(c) \\ &= \int_0^{c_D} \pi(c)(1 - H(\overline{K}(c)))\, dG(c) + \int_0^{c_D} \int_{\underline{K}(c,r)}^{\overline{K}(c)} \pi(\kappa)\, dH(K)\, dG(c) + \int_0^{c_D - r} \int_0^{\underline{K}(c,r)} \pi(c+r)\, dH(K)\, dG(c), \end{aligned} \tag{12}$$

which takes into account the possibility of making suboptimal profits due to drawing a restrictive capacity level. In a free entry equilibrium (12) equals the sunk entry cost f_E. A specific parameterization is considered in the following subsection. However, the solution to the equilibrium condition exists and is unique regardless of the distribution of K.[11]

The solution to the free entry condition determines the maximum cost level c_D. In turn, $c_D = p_D$ must also equal the zero demand threshold,

$$c_D = p_D = p_{max} = \frac{\gamma\alpha + \eta N \overline{\mathcal{P}}}{\eta N + \gamma}. \tag{13}$$

Simplifying the average price level results in

$$\begin{aligned} \overline{\mathcal{P}} &\equiv \frac{1}{N}\int_{i \in \Omega^*} p_i\, di = \frac{1}{2}(c_D + \overline{\kappa}), \quad \text{where} \\ \overline{\kappa} &= \frac{1}{G(c_D)}\left(\int_0^{c_D} c(1 - H(\overline{K}))\, dG(c) + \int_0^{c_D} \int_{\underline{K}(c,r)}^{\overline{K}(c)} \kappa\, dH(K) dG(c) + \int_0^{c_D - r} (c+r)(H(\underline{K}))\, dG(c)\right). \end{aligned} \tag{14}$$

The average implicit marginal cost $\overline{\kappa}$ has a relatively simple interpretation. It represents the weighted average of the marginal costs faced by constrained and unconstrained producers with marginal cost draws favorable enough to produce.

By combining (13) and (14) the zero cut-off profit condition is

$$N = \frac{\gamma}{\eta} \frac{\alpha - c_D}{c_D - \overline{\mathcal{P}}} = \frac{2\gamma}{\eta} \frac{\alpha - c_D}{c_D - \overline{\kappa}}. \tag{15}$$

A relatively high proportion of constrained firms raises the average price level. This softening competition increases the number of surviving firms. Therefore, an economy with relatively expensive and scarce capacity will present with an abundance of constrained firms, higher prices, and more varieties. The channels driving this result, along with their welfare consequences, are presented in more detail in the parameterization to follow.

2.4. Cost and capacity parameterization

Thus far the model is consistent for any specification of capacity and cost distributions ($H(K)$ and $G(c)$). In order to analyze the equilibrium outcomes in more depth, this section parameterizes the distributions of capacity and marginal costs and solves the ensuing equilibrium conditions. To focus on the effect of binding capacity constraints, I impose persistent constraints by assuming that the cost to adjust capacity is fully prohibitive, $r \ge c_D$. In this case, no firm will be able to expand

[11] $\overline{\Pi}$ is strictly increasing in c_D. With an application of the Leibniz rule since $\pi(\cdot)$ is increasing in c_D and, under relatively unrestrictive assumptions regarding the support of capacity, since $\lim_{c_D \to \infty} \overline{\Pi} = \infty$ and $\lim_{c_D \to c} \overline{\Pi} = 0$, a unique solution exists.

capacity without pricing above the market choke price. This assumption will produce the starkest impact of capacity constrained firms on competition. In Section 2.6 I relax this assumption and demonstrate that while the anti-competitive effects of capacity constrained firms are softened, the qualitative results are identical.

Suppose each firm draws K from a uniform distribution over the positive support $[0, K_{max}]$. Let firm productivity $(1/c)$ be drawn from a Pareto distribution with shape parameter $\beta > 1$ and minimum productivity bound $1/c_M$. This implies marginal costs are drawn from the distribution, $G(c) = (c/c_M)^\beta,\ c \in [0, c_M]$. This parameterization is consistent with recent empirical evidence.[12]

Solving the free entry condition (12) yields,

$$(c_D)^{\beta+2}\left(1 - \frac{L}{\gamma(\beta+3)}(K_{max})^{-1}c_D\right) = \frac{2(\beta+1)(\beta+2)(c_M)^\beta \gamma f_E}{L}. \tag{16}$$

For notational convenience, define the equilibrium statistic $\chi \equiv K_{max}/c_D$, which is a measure of the ease of capacity acquisition relative to the degree of market competition. Relatively large χ implies a generous capacity acquisition process (high K_{max}) or a lower premium on capacity due to stark competition limiting sales (low c_D). Positive expected profits are guaranteed as $\chi \geq L/2\gamma > L/\gamma(\beta+3)$. This imposes an environment governing capacity that is not overly restrictive.

The implicit solution of the free entry condition in (16) is

$$c_D = \left[\frac{\phi\gamma}{L}\left(\frac{1}{1 - \underbrace{\frac{L}{\gamma\chi}\frac{1}{\beta+3}}_{\text{Effectof composition}}}\right)\right]^{1/\beta+2} \tag{17}$$

where $\phi \equiv 2(\beta+1)(\beta+2)(c_M)^\beta f_E$ is a technology index. Eq. (17) is not a closed form solution as χ is endogenous, but it allows for convenient interpretation. With an application of L'Hospital's rule, as $\chi \to \infty$ (increasing the ease of acquiring capacity) expected profit, and the model in its entirety, approaches the parameterization of Melitz and Ottaviano (2008).[13] This result is useful for understanding how market structure transitions between Melitz and Ottaviano (2008), where firms access markets through quantities, and this model as constrained firms are introduced and access markets through prices. Finite χ ensures distinct groups of firms possessing managers that over- and under-provide capacity. As the probability of being constrained increases, expected profits fall. This is due to an entrant's growing potential of making suboptimal profits due to a binding capacity draw. I call the degree to which constrained firms weaken competition in the market the *effect of composition*.[14] This channel depends on the tightness of capacity acquisition (χ) relative to market size (L) and parameters (γ and β).

Combining (17) with the zero cut-off profit condition (15) characterizes the competitive equilibrium. Simplifying the average price level gives

$$\overline{\mathcal{P}} = \left(\frac{2\beta+1}{2(\beta+1)} + \frac{L}{4\gamma\chi}\frac{1}{(\beta+1)(\beta+2)}\right)c_D. \tag{18}$$

With any nonzero probability of the constraint binding (finite χ), the average price level unambiguously exceeds the average price in the model without capacity constraints and rises when the effect of composition strengthens. The zero cut-off profit

[12] Corcos et al. (2012) estimate average productivity to be approximately Pareto with shape parameter near 2 using manufacturing data from European Union countries.

[13] At each stage of equilibrium, even in the subsequent introduction of trade, this result will hold. This is an important feature of the model as it points to the fact that the competitive structure of Melitz and Ottaviano (2008) is qualitatively embedded in the model.

[14] A direct way to highlight this effect of composition is to consider a model where firms have identical marginal costs but differ in their capacity level. In this example the analogous solution to Eq. (16) in the paper is

$$\frac{L}{4\gamma}(c_D - c)^2\left(1 - \frac{L}{6\gamma}(c_D - c)K_{max}^{-1}\right) = f_E.$$

Recall that the just constrained cutoff is $\overline{K} \equiv \frac{L}{2\gamma}(c_D - c)$, then we can rewrite the above as

$$c_D = \left[\frac{4\gamma f_E}{L}\left(\frac{1}{1 - \frac{1}{3}\frac{\overline{K}}{K_{max}}}\right)\right]^{1/2} + c.$$

The cutoff is directly scaled up by firms' expectations of facing a binding capacity constraint $\overline{K}/K_{max}$. Additionally, we can see how heterogenous capacity constraints impact markups set by all firms. The markup for unconstrained firms is always one half the difference between the cutoff and marginal costs. In this example, we can see precisely the markup premium an unconstrained firm can charge when its competitors are constrained (and hence soften competition). This markup is exactly $\frac{1}{2}\left[\frac{4\gamma f_E}{L}\left(\frac{1}{1-\frac{1}{3}\frac{\overline{K}}{K_{max}}}\right)\right]^{1/2}$ for every unconstrained firm, and varies for constrained firms as a function of the tightness of their constraint. This channel is the effect of composition.

condition (and hence the number of firms) in equilibrium is then,[15]

$$N = \frac{1}{1 - \frac{L}{2\gamma\chi}\frac{1}{\beta+2}} \frac{2(\beta+1)\gamma}{\eta} \frac{\alpha - c_D}{c_D}. \tag{19}$$

Markets where the probability of facing a binding capacity constraint is high contain firms selling at inflated prices and support a greater number of firms (varieties) in equilibrium. This prediction is somewhat unintuitive. One would expect a market with many firms to have a low price level due to a higher degree of competition. Here, the subset of capacity constrained firms counteract the expected intensification of competition when the number of firms grows. Constrained firms buoy the price level since they respond to the tightness of their constraint rather than the competition from other producers. In essence, constrained firms create excess demand, which is allowing more firms to exist in the market at higher prices.

Using the equilibrium conditions to solve for consumer welfare in (4) yields,[16]

$$U = 1 + \frac{1}{2\eta}(\alpha - c_D)\left(\alpha - \frac{\frac{\beta+1}{\beta+2} - \frac{L}{2\gamma\chi}\frac{\beta+1}{(\beta+2)(\beta+3)}}{1 - \frac{L}{2\gamma\chi}\frac{1}{\beta+2}} c_D\right). \tag{20}$$

As with the preceding equilibrium conditions, welfare is uniquely determined by the cost cut-off and the capacity acquisition parameter. The effect of composition ambiguously impacts welfare. Constrained firms soften competition in the market, increasing average prices and decreasing welfare. However, this softer competition allows more firms to produce, which provides consumers with more varieties to choose from and increases welfare. Fluctuations in welfare are ultimately dominated by changes in the degree of competition c_D.

2.5. Market size, selection, and the role of capacity

Changes in market size (L) operate through countervailing channels in this model. Unconstrained firms increase sales as markets grow, thereby strengthening competition through the selection effect (lower cut-off c_D) in large markets. However, constrained firms simultaneously access large markets by raising prices. Holding the cut-off constant, the rotation of the demand curve up the quantity axis that is induced by an increase in market size pushes a number of capacity constrained firms onto their constraint (strengthening the effect of composition). The composition of firms consequently shifts to a higher fraction of constrained firms making the market relatively less competitive. The overall effects on equilibrium are contingent upon the ease of capacity acquisition, as described in the following proposition:

Proposition 1. *Firms serving larger markets face tougher selection ($\partial c_D/\partial L < 0$). In such markets, varieties grow ($\partial N/\partial L > 0$) but the effect on the average price level is ambiguous ($\partial\overline{\mathcal{P}}/\partial L \gtrless 0$). Ultimately, consumer welfare rises ($\partial U/\partial L > 0$).*

Proof. See Technical Appendix.

Proposition 1 demonstrates that larger markets produce tougher selection on productivity. Here the demand effect of producing in a larger market dominates the effect of composition. High market demand invites potential entrants, which reflects itself in the increase of the number of firms ultimately serving the market. Lower relative competitiveness of firms facing binding capacity constraints in the larger market counteracts the immediate decrease in the average price level caused by a stronger selection effect. The result is ambiguity in the change in the average price level as L approaches K_{max}. Consumer welfare is dominated by the growth in the number of available varieties and toughening of selection. These positive channels counteract the flattening of average prices due to the effect of composition, and we see welfare rise as the market grows.

Fig. A2 provides a numerical example of a parameterized equilibrium as market size varies. Notice that the average price levels flatten as more firms become constrained in larger markets even while competition (as seen by the fall in c_D) intensifies in larger markets. Panel (b) shows how the number of firms and welfare move with L. Since capacity is relatively abundant for low L, the effect of composition is weak. Therefore, the number of firms responds foremost to changes in market potential which drives the intensification of competition described in Proposition 1. Ultimately welfare climbs throughout Proposition 1, even though the average price level may rise or fall, but flattens out as prices approach the cost cutoff.

2.6. Endogenous capacity adjustment

The literature on capacity constraints is often concerned with allowing firms to make endogenous adjustments to capacity when constraints bind or are expected to bind. Here I allow firms to make endogenous adjustments to capacity, and show that the model differs quantitatively but not qualitatively from the preceding.

[15] I have already assumed $\chi > L/2\gamma$, which ensures $1 - L/2\gamma\chi 1/(\beta+2) > 0$ and $N > 0$ since $\beta > 1$.

[16] Noting that $\sigma_p = \frac{1}{4}\sigma_\kappa$ (see Melitz and Ottaviano, 2008), simplifies the welfare calculation.

Allow firms to adjust capacity at a viable per unit cost $r \in [0, c_D]$. Assume again that productivities and initial capacity levels are drawn from a Pareto and uniform distribution, respectively. Solving (13) yields the implicit solution for the cost cut-off,

$$c_D = \left[\frac{\phi\gamma}{L} \left(\frac{1}{1 - \frac{L}{\gamma\chi}\frac{1}{\beta+3}\left(1-(1-\Gamma)^{\beta+3}\right)} \right) \right]^{1/(\beta+2)}. \tag{21}$$

The cut-off is driven by the endogenous market characteristic $\chi \equiv K_{max}/c_D$ along with an additional endogenous statistic, $\Gamma \equiv r/c_D \in [0,1]$ capturing the cost of capacity adjustment relative to the degree of competition in the market. When the marginal cost of adjustment (r) is relatively large, there is a *low* premium for expandable capacity as the market is either highly competitive or capacity relatively costly. In the limit, when Γ equals one the cutoff is identical to Section 2.4 since no firm finds it profitable to adjust its capacity level. When r is low, adjusting capacity to better serve the market is relatively cheap. Cheap adjustment entices constrained firms to increase their production capabilities, which softens or even nullifies the effect of composition. If $\mathbf{r}$ equals zero, adjustment is free, and the market is without constrained firms.[17] Fig. A3 illustrates the evolution of the cost cut-off for various values of Γ. Fig. A3 demonstrates that while the quantitative results differ from the preceding section, allowing firms to make costly adjustments to capacity does not impact the qualitative outcomes unless capacity can be added freely.

3. Capacity constraints and international trade

Suppose there are two countries, H and F. Each country's consumers have identical preferences as specified by (1), but differ in size (L^H and L^F are the number of domestic and foreign consumers, respectively). Firms exporting to country $h = H, F$ incur an additional per unit iceberg transport cost of τ^h, which captures the barriers to trade between the two countries. To maintain the focus on the effect of capacity, I assume the distribution of marginal costs is identical and constant across firms and countries but allow the distributions of capacity to potentially differ across countries. The explicit marginal costs of exports are $\tau^h c$ for each firm located in country $l \neq h$ that does not adjust capacity, and $\tau^h(c+r^l)$ for firms paying the country specific cost of adjustment r^l.[18]

Iceberg transport costs require producing $\tau^h q_X^l$ units for export in country l in order to deliver q_X^l units to country h. Consequently, the total production is $\mathbf{q}^l = q_D^l + \tau^h q_X^l$ for firms producing in country l. Intuitively, maintaining retail networks and packaging products for shipment abroad also affects the rate of capacity utilization. With the possibility of trade and capacity expansion each firm located in country l faces the profit maximization problem,

$$\underset{p_D^l, p_X^l}{\arg\max} \begin{Bmatrix} (p_D^l - c)q_D^l + (p_X^l - \tau^h c)q_X^l \,|\, \mathbf{q}^l = \min\{q_D^l(p_D^l) + \tau^h q_X^l(p_X^l), K\} & \text{for } \mathbf{q}^l \le K \\ (p_D^l - (c+r^l))q_D^l + (p_X^l - \tau^h(c+r^l))q_X^l + r^l K \,|\, \mathbf{q}^l = q_D^l(p_D^l) + \tau^h q_X^l(p_X^l) & \text{for } \mathbf{q}^l > K \end{Bmatrix}.$$

The firm's problem yields first order conditions,

$$FOC \begin{cases} p_D^l = c + \lambda + \dfrac{\gamma}{L^l} q_D^l \\ p_X^l = \tau^h(c+\lambda) + \frac{\gamma}{L^h} q_X^l \end{cases} \text{for } \mathbf{q}^l \le K \qquad \begin{cases} p_D^l = c + r^l + \frac{\gamma}{L^l} q_D^l \\ p_X^l = \tau^h(c+r^l) + \frac{\gamma}{L^h} q_X^l \end{cases} \text{for } \mathbf{q}^l > K.$$

Define the implicit marginal costs for firms locating in country l as follows:

$$\kappa^l\left(c, c_D^l, c_X^l, K, \underline{K}^l, \overline{K}^l, r^l\right) \equiv \begin{cases} c + \lambda = \begin{cases} c & \text{for } \overline{K}^l \le K \\ \tilde{\mathbf{c}}_D^l - \dfrac{2\gamma}{\tilde{L}^l} K & \text{for } \underline{K}^l \le K < \overline{K}^l \end{cases} \\ c + r^l & \text{for } K < \underline{K}^l, \end{cases} \tag{22}$$

for a firm drawing marginal cost c and capacity K. The implicit costs embody the opportunity costs of serving a particular market and are captured by $\tilde{\mathbf{c}}_D^l \equiv [L^l c_D^l + (\tau^h)^2 L^h c_X^l]/\tilde{L}^l$ and $\tilde{L}^l \equiv L^l + (\tau^h)^2 L^h$, which measure effective global competition and market size, respectively. The domestic and export cost cut-offs c_D^l and c_X^l arise from the positive demand threshold in each country,

$$p_{max}^l = \frac{\gamma\alpha + \eta N^l \overline{\mathcal{P}^l}}{\eta N^l + \gamma}, \quad l = H, F,$$

where N^l is the total number of firms (domestic plus foreign) serving market l, and $\overline{\mathcal{P}^l}$ is the average price charged by these firms. The capacity constrained thresholds are endogenously determined by the cost cut-offs in each country and the opportunity cost

[17] The market with free adjustment is identical to Melitz and Ottaviano (2008).

[18] To save on notation I suppress the firm-specific subscript here and in the following.

of sales. For a firm with marginal cost c the thresholds are

$$\underline{K}^l\left(c, r^l\right) = \frac{\tilde{L}^l}{2\gamma}\left(\tilde{\mathbf{c}}_D^l - c - r^l\right) \quad \text{and} \quad \overline{K}^l(c) = \frac{\tilde{L}^l}{2\gamma}\left(\tilde{\mathbf{c}}_D^l - c\right) \tag{23}$$

for the adjusting and just constrained firm, respectively. This threshold is the level of capacity at which a firm's capacity constraint "bites" if its capacity draw is $K < \overline{K}^l$ and is prohibitive enough to induce adjustment if $K < \underline{K}^l$. The just constrained firm faces the tradeoff between foregoing sales in a given destination for the other. Therefore, constrained firms equate their domestic and export marginal profits in order to maximize total profits. Unconstrained and adjusting firms maximize profits independently in each destination at marginal cost c and $c + r^l$, respectively.

Written in terms of the domestic and export cut-off cost levels (c_D^l and c_X^l), firms producing in country l optimally choose prices and quantities,

$$p_D^l = \frac{1}{2}\left(c_D^l + \kappa^l\right), \quad p_X^l = \frac{1}{2}\tau^h\left(c_X^l + \kappa^l\right),$$

$$q_D^l = \frac{L^l}{2\gamma}\left(c_D^l - \kappa^l\right), \quad q_X^l = \frac{L^h}{2\gamma}\tau^h\left(c_X^l - \kappa^l\right).$$

Holding explicit marginal cost constant, constrained firms sell less in each destination at higher prices than unconstrained firms.

Profits in each market embody the optimal substitution of sales across markets by constrained producers, and are given by

$$\pi_D^l = \begin{cases} \frac{L^l}{4\gamma}\left(c_D^l - \kappa^l\right)\left(c_D^l + \kappa^l - 2c\right) \\ \frac{L^l}{4\gamma}\left((c_D^l - \kappa^l)^2 + \frac{4\gamma r^l K}{\tilde{L}^l}\right) \end{cases}, \quad \pi_X^l = \begin{cases} \frac{L^h}{4\gamma}(\tau^h)^2\left(c_X^l - \kappa^l\right)\left(c_X^l + \kappa^l - 2c\right) & \text{if } \underline{K}^l \le K \\ \frac{L^h}{4\gamma}(\tau^h)^2\left((c_X^l - \kappa^l)^2 + \frac{4\gamma r^l K}{\tilde{L}^l}\right) & \text{if } K < \underline{K}^l. \end{cases}$$

A constrained firm's optimal decisions in a particular market respond to fluctuations in the relative accessibility of alternate markets. For instance, a constrained firm will increase the price it charges and decrease the quantity it sells domestically in response to the foreign market becoming more favorable (c_X^l increasing). This is a consequence of the fact that the implicit marginal costs of a constrained firm increase when a particular market becomes more appealing, as they factor in the opportunity cost of selling in one destination rather than another.

3.1. Free entry equilibrium

Firms deciding to locate in country l face expected profits,[19]

$$\begin{aligned} \overline{\Pi}^l = & \int_0^{c_D^l} \pi_D^l(c) l(1 - H^l(\underline{K}^l)r)\, dG(c) + \int_0^{c_D^l}\int_{\underline{K}^l}^{\overline{K}^l} \pi_D^l(\kappa^l)\, dH^l(K)\, dG(c) + \int_0^{c_D^l - r^l}\int_0^{\underline{K}^l} \pi_D^l(c + r^l)\, dH^l(K)\, dG(c) \\ & + \int_0^{c_X^l} \pi_X^l(c) l(1 - H^l(\underline{K}^l)r)\, dG(c) + \int_0^{c_X^l}\int_{\underline{K}^l}^{\overline{K}^l} \pi_X^l(\kappa^l)\, dH^l(K)\, dG(c) + \int_0^{c_X^l - r^l}\int_0^{\underline{K}^l} \pi_X^l(c + r^l)\, dH^l(K)\, dG(c). \end{aligned}$$

In a free entry equilibrium expected profits equal the sunk entry cost f_E.[20] The domestic and export cost cut-offs in country l must satisfy $c_D^l = p_{max}^l$ and $c_X^l = p_{max}^h/\tau^h$, which implies $c_X^l = c_D^h/\tau^h$. This relationship shows the added efficiency required by firms to break into the export market and is used to simplify the system of expected profits in each country. The expected profit equation is simply interpreted as the expected weighted average of constrained and unconstrained domestic and export profits.

The average price level in country l is the average of the prices of domestic firms and exporters from h. Exporters to country l facing capacity constraints may drive up the price index in market l even beyond the degree to which constrained domestic firms do. This can be seen in the average price level, $\overline{\mathcal{P}^l} = \frac{1}{2}(c_D^l + \overline{\kappa}^l + \tau^l\overline{\kappa}^h)$. Average prices may increase relative to autarky if imports are supplied by a significant number of capacity constrained firms driving up the average implicit

[19] To ensure expected profits are well defined, I assume $K_{max}^l > \tilde{L}^l/2\gamma\tilde{\mathbf{c}}_D^l$ in each country.

[20] For simplicity I assume entry costs and the distribution of marginal costs are identical across countries. I will allow the support of the capacity distribution to differ across locations and assume firms do not endogenously choose to locate in a particular country.

marginal costs,

$$\overline{\kappa}^l = \frac{1}{G(c_D^l)+G(c_X^h)} \left(\begin{array}{l} \int_0^{c_D^l} c\left(1-H^l\left(\frac{\tilde{L}^l}{2\gamma}\left(\tilde{\mathbf{c}}_D^l-c\right)\right)\right) dG(c)+\frac{\tilde{L}^l}{2\gamma}\int_{c_D^l-r^l}^{c_D^l}\int_c^{c_D^l}\kappa^l h^l\left(\frac{\tilde{L}^l}{2\gamma}\left(\tilde{\mathbf{c}}_D^l-\kappa^l\right)\right)d\kappa^l\, dG(c) \\ \quad + \int_0^{c_D^l-r^l}(c+r^l)H^l\left(\frac{\tilde{L}^l}{2\gamma}\left(\tilde{\mathbf{c}}_D^l-c-r^l\right)\right) dG(c) \end{array} \right).$$

Relaxing the capacity constraints for domestic and foreign firms serving country l forces the price level down as competition strengthens. The number of firms serving country l (domestic and foreign),

$$N^l = \frac{2\gamma}{\eta}\frac{\alpha-c_D^l}{c_D^l-\overline{\kappa}^l} - \tau^l \overline{\kappa}^h, \tag{24}$$

corresponds with the average price level. The price level falls as competition heightens from the introduction of a number of efficient foreign producers serving the market. If a high proportion of these foreign firms serving the market are capacity constrained, the market will contain a greater fraction of producers selling at high prices on average than in autarky. Therefore, the impact of trade on varieties, prices, and competition cannot be deduced in general as they depend greatly on the rigidity of capacity in each market.

3.2. *Cost and capacity parameterization*

The following parameterizes the model with fully symmetric international trade.[21] I demonstrate that constrained firms optimally split their sales across destinations which brings rise to a substitution effect. This substitution effect has consequences for the magnitude (and even sign) of the effect of composition.

3.2.1. *Competition under symmetric trade*

Suppose there are two identical countries l and h engaging in costly but symmetric international trade, such that $L^l = L^h = L$ and $\tau^l = \tau^h = \tau \geq 1$, and consequently $\tilde{L} = (1+\tau^2)L$. Producers in each country face a potentially binding capacity constraint K drawn from the uniform distribution $H(K) \sim U[0, K_{max}]$. Let r be prohibitive in each country.[22] Firm efficiency $(\frac{1}{c})$ is drawn from a Pareto distribution with shape parameter $\beta > 1$ regardless of production location. Since the model is fully symmetric, it is immediate that the cost-cutoffs in each country are identical ($c_D^l = c_D^h = c_D$). As in autarky, define the parameter $\chi \equiv K_{max}/c_D$. The solution to the free entry equation in each country is then given by

$$c_D = \left[\frac{\gamma\varphi}{L}\left(\frac{1}{1+\rho-\frac{\tilde{L}}{2\gamma\chi}\left(1+\frac{\rho}{\tau}\right)\left(\frac{1+\tau}{1+\tau^2}-\frac{\beta+2}{\beta+3}\right)}\right)\right]^{1/(\beta+2)},$$

where $\rho \equiv \tau^{-\beta}$ is a measure of the freeness of trade, and $\varphi \equiv 2(\beta+1)(\beta+2)(c_M)^\beta \gamma f_E$ is a technology index. The effective increase in country size under trade strengthens the selection effect (as expected given Proposition 1). As was the case in autarky, capacity constraints soften selection through an effect of composition that depends on the ease of capacity acquisition. The effect of composition is magnified by the presence of additional constrained firms abroad. Constrained firms split their sales between destinations in light of trade costs until marginal profits are equalized. Depending on the degree of this substitution effect, the effect of composition in each country can be amplified or reduced. In fact, if the substitution of constrained exporters away from the domestic market is strong due to low enough transport costs, the effect of composition magnifies rather than softens competition in each country. Propositions to follow discuss the mechanism driving these effects in greater detail.

The average price level in each country is the average of the prices charged by constrained and unconstrained domestic and foreign firms, which can be written as

$$\overline{\mathcal{P}} = \left(\frac{2\beta+1}{2(\beta+1)} + \frac{\tilde{L}}{4\gamma\chi}\frac{1+\frac{\rho}{\tau}}{1+\rho}\underbrace{\left(\left(\frac{1+\tau}{1+\tau^2}\right)^2 - \frac{2\beta}{\beta+1}\frac{1+\tau}{1+\tau^2}+\frac{\beta}{\beta+2}\right)}_{\text{Substitution Effect}}\right)c_D \tag{25}$$

Constrained firms boost the average price level through the effect of composition. Trade introduces a larger set of constrained firms to each market through the substitution effect, which magnifies the effect of composition. The degree of price magnification from constrained domestic and foreign firms depends on the proclivity of constrained exporters to serve

[21] The channels, and their effects, that arise from symmetric trade also drive asymmetric trade. Therefore, I leave the case of asymmetric trade for the Technical Appendix.

[22] Specifically, r takes on a value greater than or equal to $\max\{c_D^l, c_X^l\}$.

the foreign market. Symmetry implies that transport costs are the main driver of this substitution effect, which noticeably embeds the composition term.

Combining average prices with the zero cut-off profit condition in (24) yields the number of firms serving each market. Letting

$$S(\tau,\beta) \equiv \left(\left(\frac{1+\tau}{1+\tau^2}\right)^2 - \frac{2\beta}{\beta+1}\frac{1+\tau}{1+\tau^2} + \frac{\beta}{\beta+2} \right)$$

be the substitution effect from (25),

$$N = \frac{1}{1-\frac{\tilde{L}}{2\gamma\chi}\frac{1+\frac{\rho}{\tau}}{1+\rho}S(\tau,\beta)(\beta+1)} \frac{2(\beta+1)\gamma}{\eta}\frac{\alpha-c_D}{c_D}.$$

The effect of composition unambiguously boosts the average price level, which in turn leads to a greater number of firms serving the market. Combining the average and variance of prices with the number of varieties available in the market post trade yields the following expression for consumer welfare,

$$U = 1+\frac{1}{2\eta}(\alpha-c_D)\left(\alpha - \frac{\frac{\beta+1}{\beta+2} - \frac{\tilde{L}}{2\gamma\chi}\frac{1+\frac{\rho}{\tau}}{1+\rho}\left(\beta S(\tau,\beta) - \frac{\beta-1}{(\beta+1)(\beta+2)(\beta+3)} + \frac{\tilde{L}}{2\gamma\chi}\frac{1+\frac{\rho}{\tau}}{1+\rho}\left(\frac{\beta+1}{2}S(\tau,\beta) - \frac{1}{\beta+2}\right)\right)}{1-\frac{\tilde{L}}{2\gamma\chi}\frac{1+\frac{\rho}{\tau}}{1+\rho}S(\tau,\beta)(\beta+1)} c_D\right).$$

The degree prices and varieties are elevated through the effect of composition is governed by the substitution effect. Therefore, comparisons between trade and autarky levels of competition, prices, varieties, and consumer welfare in the market depend on how transport costs transform the selection effect and effect of composition.

Proposition 2. *Symmetric trade intensifies competition ($c_D^{Symm} < c_D^{Autarky}$), ambiguously impacts average prices ($\overline{\mathcal{P}}^{Symm} \gtrless \overline{\mathcal{P}}^{Autarky}$), and leads to an increase in the number of firms serving each market ($N^{Symm} > N^{Autarky}$).*

Proof. See Technical Appendix.

Ambiguity in the direction of prices in the transition from autarky to trade is a direct consequence of the added option available to exporters. Constrained firms must choose how much of their limited production to allocate to each destination. Since these firms drive the elevation of prices, whichever country is most accessible faces softer price competition from constrained exporters. Symmetry in this problem implies that there are not disproportionate substitutions across destination, so that each country realizes the same strengthening of the effect of composition.

The softening of prices by constrained exporters coupled with the hardening of competition by unconstrained exporters ultimately means comparing the effect of trade on consumers between autarky and trade depends upon parameter choices.

Proposition 3. *When capacity is abundant in both countries ($K_{max} > L/2\gamma\ \nu(\beta,\gamma,\tau,c_D^{Autarky},c_D^{Symm})$) consumer welfare rises under symmetric trade ($U^{Symm} > U^{Autarky}$). When capacity is scarce ($K_{max} < L/2\gamma\ \nu(\beta,\gamma,\tau,c_D^{Autarky},c_D^{Symm})$), consumer welfare can fall ($U^{Symm} < U^{Autarky}$) with trade.*

Proof. Generally, $\nu(\beta,\gamma,\tau,c_D^{Autarky},c_D^{Symm})$ is the solution to a quadratic equation. In particular cases, $\nu(\beta,\gamma,\tau,c_D^{Autarky},c_D^{Symm})$ can be simplified. The general and simplified solutions are presented in the Technical Appendix along with numerical simulations. □

Limited capacity in each country creates an abundance of constrained domestic and foreign firms. Suppose for a moment that there are zero trade costs between countries ($\tau=1$). Firms efficient enough to produce are also able to export (implied by symmetric free trade). Constrained exporters perfectly split their domestic sales and exports in order to equalize marginal profits in each destination. This division of sales elevates the price level in each country as constrained firms raise prices in each destination to slow sales. As the global market becomes saturated with constrained firms, average prices begin to converge to the cut-off cost level. Consequently, the variance of prices in each country is considerably lower than what is realized in autarky. The channel by which these effects work on consumer welfare is through a decrease in the surplus consumers extract from each variety. As constrained firms increase prices consumers purchase lower quantities from these firms and extract less surplus from each variety consumed when the market becomes saturated by constrained firms. The reduction in the surplus extracted by consumers for each variety can be strong enough to counteract the variety and efficiency gains seen in Proposition 2 from a larger market. The impact is a fall in welfare from trade.

The definition of capacity abundance or scarcity, $\nu(\beta,\gamma,\tau,c_D^{Autarky},c_D^{Symm})$, depends on model parameters. The range through which welfare falls after trade can be narrow or wide. If the fall in the cost cutoff after trade is relatively large, the range is relatively narrow since efficiency gains from stronger selection after trade dominate the distortions of constrained firms. If the efficiency gains are relatively small, this range is considerably wider as the pricing distortions caused by domestic along with foreign constrained firms dominates the variety and efficiency gains. Higher transport costs magnify the distortions and widen the range of potential welfare losses. As τ increases, all exporters utilize more capacity. As these exporters become constrained, they raise both domestic and export prices and scale back sales in each destination, which leads to welfare losses after trade through the substitution effect. Similar effects occur when varieties are more differentiated (i.e., large γ) or when productivities are relatively disperse (i.e., large β).[23]

These results call back to the theory of "second best" in trade made fashionable by Bhagwati and Ramaswami (1963).[24] In this model, symmetric trade amplifies the price distortion caused by constrained firms in precisely a "second best" manner when capacity is sufficiently scarce. Therefore, trade liberalization can have ambiguous effects. It can either toughen competition and improve welfare or lead countries to a "second best" equilibrium.

4. Empirical evaluation

Using Thai firm-level data I identify some key predictions of the model. I begin by supporting some underlying assumptions of the model. I then estimate the substitution patterns of firms in light of changes in implicit (taking into account capacity tightness) marginal costs from tariff movements. The model demonstrates that declining trade barriers lead to increasing opportunity costs of domestic sales for capacity constrained exporters. Implicit opportunity costs imply a feedback effect of trade liberalization which decreases domestic sales and is magnified for firms facing tight constraints. I explore these predictions by decomposing the impact of tariff changes on sales across domestic and foreign markets across. I present evidence that the impact of tariffs on domestic sales depends on both the tightness of a firm's capacity constraint and its presence in export markets. Lastly, the model explains the existence of high productivity firms with relatively low sales when capacity is prohibitive. I show these firms are prevalent in the data, and that the probability of observing such firms is explainable by the intensity of their capacity utilization.

4.1. Data

To my knowledge, the only readily accessible firm-level survey documenting capacity is the *Productivity and Investment Climate Study: Thailand*.[25] This survey was administered in 2004 and 2007 for select Thai manufacturers.[26] The survey's questionnaires allow me to construct an unbalanced panel of Thai establishments from 2001 to 2006 consisting of the variables required to estimate my empirical model. The data are at the 4-digit International Standard Industrial Classification (ISIC) Revision 2 level.

The key variables for the following analysis include the following. Domestic and export revenue document the total revenues earned by the firm for domestic and export sales, respectively. Capacity utilization is self reported. Investment includes all investment by the firm in billions of Baht. Total Assets are total fixed assets. Domestic sales are the total sales by all domestic firms in an industry. Industry Imports are total imports by an industry. Credit constraint is an indicator for whether the firm considers "access to credit to be an obstacle to doing business." Ownership is the individual or group with the largest share of ownership of the firm. Most importantly for this study, firms respond to the question: "*III.8. Capacity utilization is the amount of output actually produced relative to the maximum amount that can be produced. At what capacity did this plant produce in* [*year*]?" Table A1 presents summary statistics of all of the variables used in the subsequent analysis across firms that sell only domestically and all other firms that export. On average, we can see that exporters have higher sales, are more productive, and utilize more capacity precisely as the preceding model predicts.

A total of 5493 firm-year observations for 1874 distinct firms populate the data.[27] I supplement the firm-level data with 4-digit industry level imports and tariffs from the UNCTAD Trade Analysis and Information System (TRAINS) database.[28] Between 2003 and 2006, Thailand undertook a significant program to restructure and reduce tariffs (WTO, 2007). In kind, the majority of Thailand's trading partners also decreased their tariffs against Thailand. Controlling for industry differences, the data exhibit an average decline of around 20% per year relative to 2001. The variation across industries and over time are particularly useful in identifying the affect foreign tariffs have on capacity constrained exporters in what follows. Given the coinciding decline in tariffs by and on Thailand, it will be important to control for changing domestic market structure. Total domestic sales (*DomesticSales*) will capture competition amongst Thai firms and total imports (*IndustryImports*) will control for any increases in foreign competition and import penetration in what follows.

[23] When varieties are more differentiated, consumers extract relatively less surplus from each firm as they price higher on their demand curve. This surplus is further decreased after trade as firms are pushed onto their constraint. When productivities are relatively disperse, more firms hit their constraint after trade, and the loss of consumption from these firms erodes the surplus consumers extracted before trade.

[24] A thorough analysis of the "second best" literature is presented by Krishna and Panagariya (2000), which provides great detail as to how trade can magnify distortions resulting in suboptimal outcomes.

[25] Conducted by The Foundation of Thailand Productivity Institute and the World Bank.

[26] A more detailed description of this survey and its positive and negative qualities can be found in Dhingra (2010).

[27] The subsequent empirical analysis utilizes the "sales generating function" to estimate total factor productivity (TFP). Therefore, I drop firms with negative value added. Before dropping a firm on the basis of inaccurate value added measures I construct value added using sales and inputs. This measure is highly correlated with reported value added, and I use it in place of recorded value added less than zero. After eliminating observations on this basis, I am left with 1759 firms and 5453 observations.

[28] Since the Thai data only record total exports at the firm level, my tariff measure is the average of tariffs across all Thai export destinations weighted by total exports at the industry level. Notably, variation in this tariff variable is not driven by variation of total exports within the industry, as demonstrated by within industry regressions and raw correlations. The variation rather is driven by changes to the structure of tariffs. This measure is intended to capture the desirability of the aggregate export market, which constrained firms use to substitute across destinations.

4.2. Analyzing capacity utilization

Capacity utilizations within each 4-digit industry follow distributions consistent with the notion that there are sets of unconstrained and just constrained firms. Fig. A4 presents the distributions of capacity utilization for four of the largest industries in Thailand.[29] This figure documents a grouping between 50% and 80% utilization of capacity and a spike in the density of firms utilizing 90% or more of their capacity. The discontinuity in capacity utilization is consistent with a mass of firms facing binding (or nearly binding) capacity constraints. Furniture displays a low average rate of utilization but has the highest mass of firms facing tight capacity constraints of the group. Conversely, wearing apparel generates the highest average rate even with a lesser mass of constrained firms. The heterogeneity across industries, and discontinuity in the distributions of capacity utilization provide identification in my model.

Before testing the particular predictions of the model, it is useful to describe the variation of capacity utilization within firms and provide support for some of the model's key underlying assumptions. The model abstracts away from two potential regularities: firms make endogenous capacity decisions *ex ante* and dynamic capacity adjustment decisions *ex post*. Firms not engaging in strategic entry deterrence or facing uncertainty about demand or productivity would fully utilize capacity. However, as seen in Fig. A4 there are distinct sets of firms that over- and under- invest in capacity. My model assumes that capacity is randomly assigned, and distributions of these firms should not differ significantly by fundamentals such as sales and productivities. Fig. A5 documents this feature of the data. The distributions of the capacity utilization do not differ statistically from one another across sales or productivity quantiles.

Finally, if firms were making frequent permanent adjustments to capacity, the data should show a strong relationship between investment and capacity utilization. Table A2 looks within firms to examine dynamic capacity adjustments. Columns (1)–(4) present a series of fixed effects regressions of current capacity utilization. In the first two columns, we see that capacity responds to contemporaneous shocks to firm outcomes such as sales and export status. It is possible that firms face time to build constraints when responding to tight capacity. Column (3) includes lagged investment to account for this possibility. The coefficients on investment become positive and significant. This regression, however, does not account for the reality that investment is lumpy, and capacity utilization is likely persistent.

Columns (5)–(8) examine the dynamic nature of capacity utilization using the method developed by Blundell and Bond (1998).[30] It is evident that capacity is highly persistent, suggesting firms do not easily adjust their utilization period to period. Additionally, variation in capacity utilization within the firm depends upon sales at the firm level rather than investment, precisely as the model assumes. Noting the lumpiness of investment and that the capacity of firms making positive investment may be fundamentally different than for firms with little or no investment, Columns (4) and (8) restrict the sample to firms investing more than 1 Million Baht. The estimates further support the idea that the capacity utilization of firms depends foremost on fluctuations in sales rather than dynamic adjustments to capacity in the short to medium run.

4.3. Testing model predictions

I focus here on establishing the presence of the substitution of firms across sales destinations in light of changes to their implicit cost structure when capacity constraints tighten. The substitution effect described can be explained by any arbitrary specification where firms face increasing marginal costs. However, since marginal costs are generally unobservable these effects have remained difficult to characterize. My model provides the clean prediction that firms substituting across destinations are specifically capacity constrained exporters. This substitution provides insight into how foreign market conditions transmit to domestic outcomes.

According to the preceding theory, domestic revenues directly depend on the opportunity cost of selling limited production domestically rather than abroad if the firm is constrained. In the model with prohibitive capacity cost, firms are either unconstrained or just constrained and the log of domestic revenues is

$$\log\left(p_{ij}^l q_{ij}^l\right) = \begin{cases} \log\left(\frac{L^l}{4\gamma}\right) + \log\left(c_D^2 - \left(\tilde{\mathbf{c}}_D^l - \frac{2\gamma}{\tilde{L}^l} K_{ij}\right)^2\right) & \text{if } K_{ij} < \overline{K}^l \\ \log\left(\frac{L^l}{4\gamma}\right) + \log(c_D^l - c_{ij}) + \log(c_D^l + c_{ij}) & \text{if } K_{ij} \geq \overline{K}^l \end{cases}.$$

It is readily confirmed that the derivative of domestic revenue is positive for constrained firms ($K_{ij} < \overline{K}^l$) and zero for unconstrained firms ($K_{ij} \geq \overline{K}^l$).[31] In the model as τ increases, the foreign market becomes relatively less profitable.

[29] Notably, each of the other industries in the data produce similarly shaped distributions.

[30] Repeating Columns (5)–(8) using Arellano and Bond (1991) yields qualitatively identical results to the Blundell and Bond (1998) strategy reported.

[31] Under symmetric trade the derivative of domestic revenues with respect to trade costs is

$$\frac{\partial p_{ij}^l q_{ij}^l}{\partial \tau_j} = \begin{cases} 2\left(\left(1+\frac{1}{\tau_j}\right)c_D - \frac{2\gamma K}{(1+\tau^2)L}\right)\left(-\frac{1}{\tau_j^2}c_D + \frac{2\gamma K}{(1+\tau^2)L}\right) & \text{if } K_{ij} < \overline{K}^l \\ 0 & \text{if } K_{ij} \geq \overline{K}^l \end{cases}. \quad (26)$$

Constrained exporters respond by substituting their limited capacity away from the export market and into the domestic market. The unconstrained firm is unaffected by the increase in trade costs.[32] The differences between how firms respond to exogenous changes in trade costs given the tightness of their capacity constraints is the vehicle by which the following empirics identify substitution by firms across destinations.

Effectively investigating the impact of capacity constraints in the model requires controlling for the latent efficiency (here TFP) of each firm in order to control for the expected level of production across firms. Separately across industries, suppose firm i's "sales generating function" in time t can be written as, $va_{it} = \theta_l l_{it} + \theta_k k_{it} + \omega_{it} + \epsilon_{it}$, where va, l, and k represent value added, labor (total employees), and capital (fixed assets), respectively, in logs. Productivity terms ω and $\epsilon$are unobserved by the econometrician.[33] Olley and Pakes (1996) propose a two-step structural estimation method using a firm's investment decisions as a proxy for capital while also accounting for selection in order to estimate the production function. I rely on their procedure to estimate productivity, as their methodology and intuition is internally consistent with the intuition in this paper of firms facing capacity constraints.[34]

The estimating equation for each Thai firm decomposes domestic revenues, and is given by,

$$\log(R_{ijt}) = \delta_j + \delta_t + \delta_1 tfp_{ijt} + \delta_2 CU_{ijt} + \delta_3 \tau_{jt} + \delta_4 X_{ijt} + \delta_5 \tau_{jt} * X_{ijt} + \delta_6 \tau_{jt} * CU_{ijt} + \delta_7 X_{ijt} * CU_{ijt} \\ + \delta_8 X_{ijt} * CU_{ijt} * \tau_{jt} + \delta_9 IndustryImports_{jt} + \delta_{10} DomesticSales_{jt} + \varepsilon_{ijt}, \tag{27}$$

for each firm i selling its variety in the 4-digit industry j in year t. Year and product fixed effects (δ_t and δ_j) will absorb average differences across industries and time. The log of firm productivity is given by tfp_{ijt}. The reported utilization of capacity is CU_{ijt}.[35] To control for relative differences between the desirability of exporting versus selling domestically, I include the log of the average foreign tariff faced by firms (τ_{jt}) and combinations of imported ($IndustryImports_{jt}$) and domestic ($DomesticSales_{jt}$) consumption by Thailand within a 4-digit industry.

Standard theory would predict no direct impact of foreign tariffs on domestic sales (holding the cut-offs constant, we see this in unconstrained firms' revenues). When firms face capacity constraints, the response from constrained firms depends on the relative desirability of each destination ($c_D^l - c_X^l$). Market characteristics impact firms differently depending upon the tightness of its constraint and its presence in the export markets ($X_{ijt} = 1$ if a firm exported in t). Therefore, I interact CU_{ijt} with tariffs and exporter status. My model predicts that a capacity constrained exporter is foregoing domestic sales to serve the foreign market, thus a negative coefficient on $X_{ijt} * CU_{ijt}$ is expected.

The triple interaction of tariffs, capacity, and exporter status ($X_{ijt} * CU_{ijt} * \tau_{jt}$) is the key variable for identifying substitution by constrained exporters into the domestic market as accessibility abroad changes. My model predicts a positive coefficient on this triple interaction as a constrained exporter will substitute away from foreign markets as tariffs rise.

Table A3 presents the results of various log-log fixed effects regressions of Eq. (27). Column 1 estimates the correlation of productivity and capacity with revenues including a full set of industry and year fixed effects. The coefficients are thus being identified off of variation across firms within an industry. Initial results show that firms with high productivity utilizing more of their capacity have higher revenues. Quickly looking across the first row of Table A3 demonstrates that more productive firms obtain higher revenues. The positive correlation between capacity utilization and revenues throughout Table A3 demonstrates that high capacity utilization (production) corresponds with higher sales.

Column 2 of Table A3 sets out to establish the substitution of firms away from the domestic market in response to tariff changes. The estimate suggests that increasing the average tariff faced by a firm 10% leads to a 2% increase in domestic revenue.[36] My model predicts that it is capacity constrained exporters which substitute into the domestic market as tariffs rise. In order to establish the differential effect of tariffs on firms depending on their capacity utilization and presence in foreign markets, I decompose the tariff, capacity utilization, and export status by including all possible interactions of these variables. Column 3 presents the results of this specification. Controlling for tariff driven shocks to export supply and the substitution of firms between exporting and domestic sales isolates the behavior of constrained firms. Signs of the decomposition estimates correspond directly with the model's predictions. Capacity constrained exporters have lower domestic sales ($\delta_8 < 0$) on average, and as tariffs rise they substitute into the domestic market ($\delta_9 > 0$).[37]

(*footnote continued*)
Focusing on constrained firms, it is trivial to show $(1+1/\tau_j)c_D - 2\gamma K/(1+\tau^2)L > 0$ and $-1/\tau_j^2 c_D + 2\gamma K/(1+\tau^2)L > 0$ when $K_{ij} < \overline{K}^l$ so that domestic revenues for constrained firms are increasing in trade costs.

[32] Naturally, this is as long as they remain unconstrained. An additional prediction of the model is that capacity constraints are tightening for all firms as the derivative of the constrained cutoff ($\overline{K}^l$) is increasing with respect to τ.

[33] There are some caveats between interpreting the "sales generating function" I estimate and the classic production function. These issues are discussed at length by Fernandes and Pakes (2008).

[34] Levinsohn and Petrin (2003) propose an alternate method to estimate productivity. However, averaged across all industries, the within industry correlation between Olley and Pakes (1996) and Levinsohn and Petrin (2003) TFP estimates here is 0.95. The following estimates are nearly identical regardless of methodology, as productivity differences are intended as controls for the level of production.

[35] I have investigated many variations of capacity in this estimation. These robustness checks include using within industry capacity quintiles or an indicator for "tight" capacity and all yield qualitatively identical results to the following results.

[36] Supporting this evidence, Vannoorenberghe (2012) demonstrates strong negative correlation between domestic and foreign sales growth rates using French firm level data. Using similar data, Berman et al. (2011) argue that a 10% increase in exports generates a 1.5–3% decrease in domestic sales and discuss larger effects in industries where capacity likely plays a more significant role.

[37] Additional empirical evidence of the implications of capacity is presented by Ahn and McQuoid (2012) who show capacity constraints explain muted responses by Indonesian firms facing shocks to demand. A corollary prediction of the model is that exporters' capacity constraints tighten as tariffs

The model suggests that the relative desirability of foreign and domestic markets influences how firms ration their limited production across destinations. Columns (4) and (5) of Table A3 include controls for total consumption in Thailand of imported products ($TotalImports_{jt}$) and products produced domestically ($DomesticSales_{jt}$) in order to capture the relative competitiveness of the domestic market. The previous results are unaffected.

Outside of the model, there may be competing explanations for firm substitution as described. One may be concerned that Thai firms owned by foreign entities are inherently highly productive and produce at high capacity levels. This is an issue if these firms produce in order to serve the foreign affiliates rather than compete in the Thai market. In order to control for this, I include fixed effects for the 5 modes of ownership described in the survey in Column (6). Including ownership fixed effects increases the R^2 significantly, but does not impact the core results. Finally, a competing theory of firm constraints affecting firm behavior is the idea that firms facing credit constraints cannot readily access markets. To address this issue Column (7) includes an indicator of whether the firm found access to credit to be an "obstacle to doing business in Thailand." Credit constrained firms sell less domestically, as expected, and the main results of the model are maintained. We may still be concerned that there are industry by year unobservables, such as sectoral demand differences, influencing the results. Column (8) includes IndustryXYear fixed effects to further absorb any of this unobserved variation. The R^2 increases to nearly 0.4, and the key triple interaction becomes statistically significant at the 1% percent level.

Calculating the net effect of the variables of interest on domestic sales through direct marginal effects is cumbersome in Column 4 of Table A3. To alleviate these calculations, Fig. A6 provides the distributions of firms' predicted domestic revenues in Specification 3 across apparent low, middle, and high tariff regimes.[38] Panel (a) suggests a modest substitution by unconstrained exporters away from the domestic market in response to trade liberalization. Panel (b) highlights the predictions of the model by focusing on capacity constrained exporters. As predicted, there is a monotonic increase in domestic sales as firms substitute away from markets abroad when trade barriers increase. The predicted impact on the average capacity constrained exporter of a 10% *ad valorem* increase in tariffs is approximately a 20% increase in domestic revenue.

The model can also explain the presence of high productivity firms with relatively low sales. This is an anomaly that the majority of new trade theory models cannot reconcile, but an important window into the potential for growth of this developing country. Fig. A7 documents the proportion of firms with TFP in the top two quintiles in each revenue quintile within their respective industries.[39] This fraction of high TFP firms increases as the "Standard Theory" of Melitz (2003) and Melitz and Ottaviano (2008) would predict with sales. However, within each of the lowest sales quintiles a considerable fraction of firms do not coincide with standard predictions. The fraction of highly productive firms with low sales make up nearly 20% of all firm-year observations.[40] My model asserts that the existence of these anomalous firms is explainable by capacity constraints.

To test this prediction, I produce a dummy variable indicating the firms violating "standard theory," which I use in various conditional logit regressions. The results shown in Table A4 demonstrate the probability of a firm being highly productive yet small increases substantially as its capacity constraint tightens. The first four columns look within firms to identify the effects. The estimates suggest that the tighter is a firm's capacity constraint the more likely the firm is to violate standard theory. I estimate that for a 10% increase in capacity utilization, a firm is around 30% more likely to violate standard theory. The final three columns look across firms for identification. Here, we want to control for potentially convoluting factors such as firm ownership and credit constraints as we did to identify substitution effects. Additionally, Holmes and Stevens (2010) discuss the existence of "boutique firms." These firms are potentially highly productive, produce at high capacity, and face inherently low demand for their niche products. These firms also likely possess relatively few fixed assets. The results in the final three columns do suggest that firms with less assets (e.g., boutique firms) or those which are family owned are more likely to violate standard theory. Even with these controls, capacity utilization has a strong positive correlation with the probability of a firm violating standard theory. The existence of these highly productive constrained firms suggests potential for growth in Thailand from policy that alleviates the capacity constraints of these promising firms.

5. Conclusion

By introducing capacity constraints into a model of international trade with heterogeneous firms, this paper produces a tractable framework that can explain various occurrences in the data anomalous to new trade theory. The model highlights a general equilibrium where price competition is softened in markets with high concentrations of capacity constrained firms. Changes in the structure of firms' implicit marginal costs through fluctuating trade barriers and market size endogenously impact their production and pricing decisions. This introduces relative market characteristics directly into competitive equilibrium conditions, thus providing an explanation for the substitution patterns of exporters across

(*footnote continued*)
increase. This holds in the data as well. Fixed effect regressions including the interaction of tariffs with export status yield a positive and significant coefficient on the interaction term.

[38] Each tariff regime corresponds to approximately 30% of all plant-year observations in the data and are defined as: Low 0–4.5%, Middle 4.5–8%, and High 8–10%.

[39] The industries focused on in Fig. A7 were chosen because they are of the largest in Thailand. The results presented by this figure are not sensitive to the choice of industry.

[40] Low sales is defined as any of the lowest three revenue quintiles, and high productivity is the top two TFP quintiles at the industry level.

destinations. Additionally, I empirically identified the prevalence of these substitution patterns within firms and attributed them to capacity constraints as predicted by the model.

The flexibility of the model's results allows it to be readily applied to a host of empirical studies. For instance, unintuitive responses by firms to changes in global demand may suggest a capacity acquisition process hindering firms' production capabilities. Without reform to capacity acquisition (be it through lowering the cost of expansion or ease of acquisition) a country heavily impacted by capacity constraints will face considerable limitations when trying to compete in global markets. The impact of constrained exporters may resonate globally as the shocks across markets are now linked within firms. These linkages could yield insight into our understanding of how market characteristics such as exchange rates and shocks to supply or demand are passed through by firms to shape fundamentals such as inflation and elasticities of trade.

Acknowledgments

I am exceedingly grateful to Bruce Blonigen, Swati Dhingra, Robert Feenstra, David Hummels, Christopher Knittel, John Morrow, Stephen Redding, Katheryn Russ, Georg Schaur, Nicholas Sly, Alan Spearot, Deborah Swenson, Victor Stango, David Weinstein, and Greg Wright for their helpful comments and ongoing support of this project. This paper has also benefited from the comments of seminar participants at an Empirical Investigations of International Trade Conference, Midwest International Economics Group Meeting, Purdue University, and the University of California–Davis. Any remaining errors or omissions are my own..

Appendix A

See Figs. A1–A7 and Tables. A1–A4.

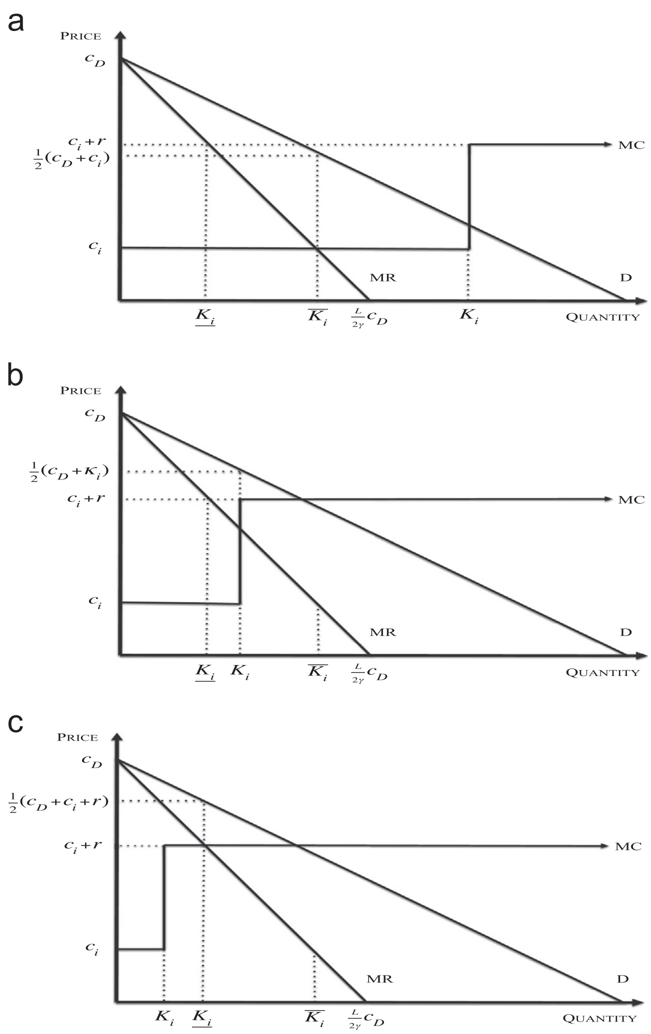

Fig. A1. The firm's problem in autarky. (a) Unconstrained firm, (b) just constrained firm and (c) adjusting firm.

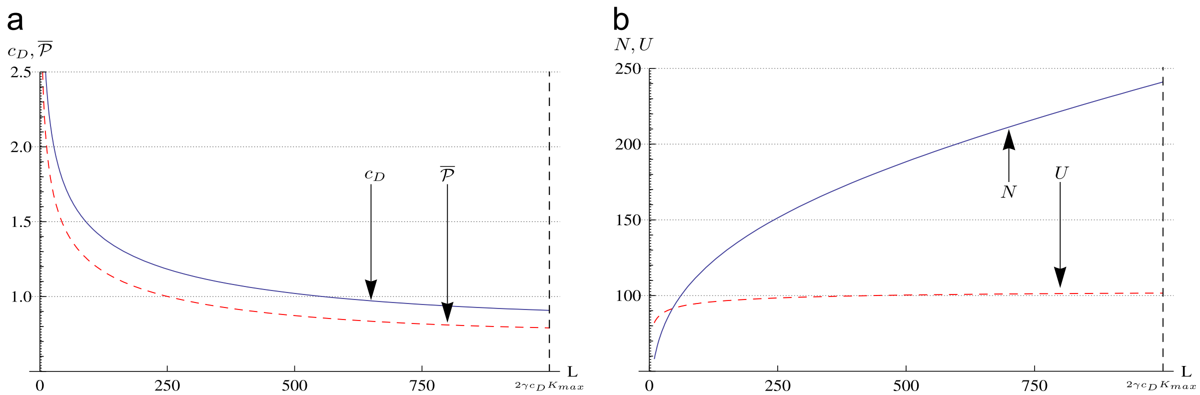

Fig. A2. Model parameterization in autarky. (a) Cost cut-off and average price level and (b) number of firms and consumer welfare. *Notes*: Parameter values are fixed at $\alpha = 15$, $\beta = 2$, $K_{max} = 250$, $\eta = 1$, $\gamma = 2$, $f_e = 1$, $c_M = 3$.

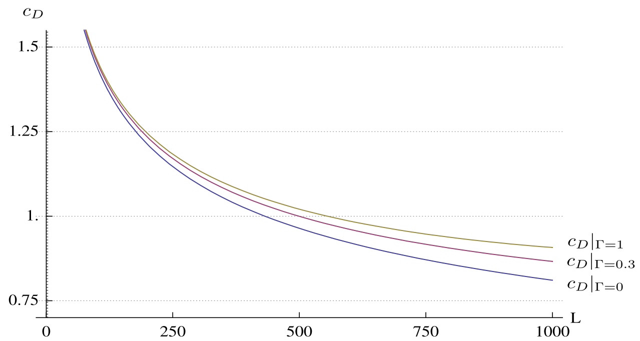

Fig. A3. Model parameterization in autarky with endogenous capacity adjustment. *Notes*: Parameter values are fixed at $\alpha = 15$, $\beta = 2$, $K_{max} = 250$, $\eta = 1$, $\gamma = 2$, $f_e = 1$, $c_M = 3$.

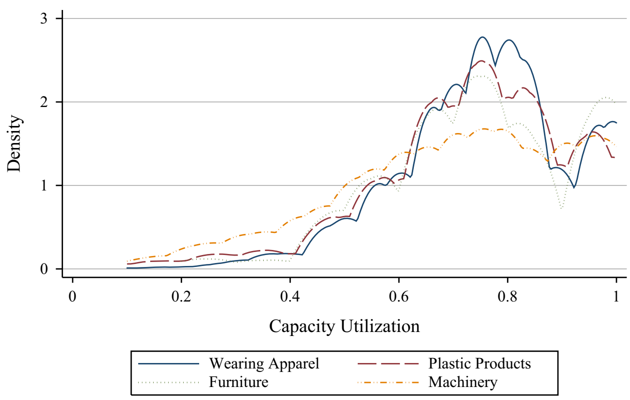

Fig. A4. Capacity utilization kernel: by industry.

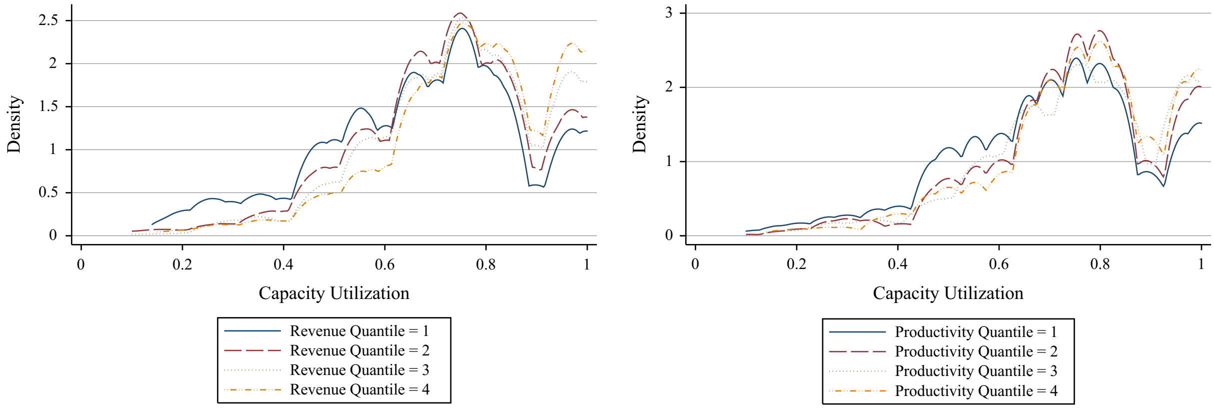

Fig. A5. Capacity utilization kernel: by sales and productivity.

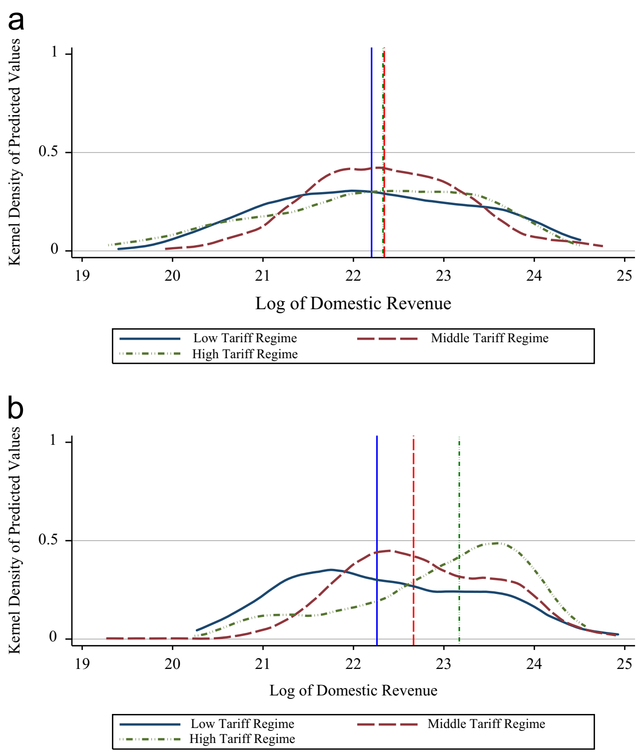

Fig. A6. Firm substitution into the domestic market in response to trade liberalization: the distributions of predicted values from the specification in Table A3.(3). (a) Exporters in the bottom three capacity utilization quintiles and (b) exporters in the top two capacity utilization quintiles.

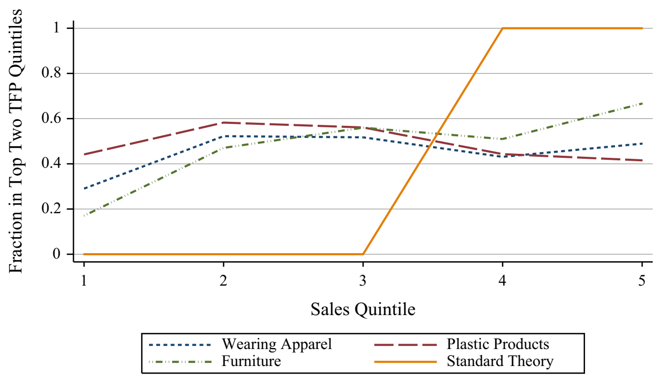

Fig. A7. Investigating high productivity low sales firms.

Table A1
Summary statistics for thai firm-level data.

Variables	Domestic only				Exporters			
	Mean	Min	Max	SD	Mean	Min	Max	SD
Domestic Revenue	21.705	14.955	27.339	1.491	22.436	13.424	27.639	1.957
Export Revenue	0.000	0.000	0.000	0.000	22.332	12.215	28.650	2.393
Tariff	1.656	−2.526	3.422	0.992	1.678	−2.526	3.422	1.007
TFP	0.811	−5.675	3.936	0.857	0.898	−3.624	3.479	0.728
Capacity utilization	0.736	0.100	1.000	0.189	0.779	0.100	1.000	0.180
Investment	0.005	0.000	2.740	0.064	0.050	0.000	56.600	1.090
Total assets	16.676	11.382	21.944	1.582	18.618	12.698	23.884	1.818
Domestic sales	20.364	11.849	23.173	1.418	20.094	11.849	23.173	1.883
Industry imports	27.434	19.781	29.216	1.403	27.400	20.098	29.216	1.496
Credit constraint	0.156	0.000	1.000	0.363	0.108	0.000	1.000	0.311
Ownership								
Family	0.219	0.000	1.000	0.414	0.159	0.000	1.000	0.366
Private investors	0.385	0.000	1.000	0.487	0.379	0.000	1.000	0.485
Private corp	0.039	0.000	1.000	0.195	0.078	0.000	1.000	0.269
Multinational corp	0.030	0.000	1.000	0.171	0.204	0.000	1.000	0.403
Other	0.285	0.000	1.000	0.451	0.166	0.000	1.000	0.372
N	2731				2759			
Firms	932				942			

Notes: Variables labeled in italics are reported in logs. Domestic and Export Revenue represent revenues earned by the firm for domestic and export sales, respectively. Tariffs are 4-digit ISIC tariffs averaged across Thai export destinations weighted by total exports. TFP is estimated using Olley and Pakes (1996). Capacity utilization is self reported. Investment includes all investment by the firm in billions of Baht. Total Assets are total fixed assets. Domestic Sales are the total sales by all domestic firms in an industry. Industry Imports are total imports by an industry. Credit Constraint is an indicator for whether the firm considers "access to credit to be an obstacle to doing business." Ownership is the individual or group with the largest share of ownership of the firm.

Table A2
Exploring the variation in capacity utilization.

CU_{ijt}	Fixed effects				Blundell-bond			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CU_{ijt-1}					0.736***	0.735***	0.730***	0.612***
					(0.029)	(0.032)	(0.032)	(0.047)
$TotalRevenue_{ijt}$	0.018**	0.017**	0.028**	0.034***		0.042**	0.041**	0.035**
$(BillionBaht)$	(0.007)	(0.007)	(0.012)	(0.010)		(0.017)	(0.017)	(0.017)
$Investment_{ijt}$	−0.000	−0.000	0.002***	−0.002		−0.002	0.001	0.001
$(BillionBaht)$	(0.000)	(0.000)	(0.001)	(0.003)		(0.003)	(0.001)	(0.001)
$Investment_{ijt-1}$			0.001*	0.001			0.001	0.000
$(BillionBaht)$			(0.001)	(0.001)			(0.001)	(0.001)
X_{ijt}		0.032*	0.019	0.027		0.008	0.007	−0.003
		(0.018)	(0.019)	(0.030)		(0.018)	(0.018)	(0.028)
Invest $< 1Mill$	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firms	1764	1764	1733	885	1741	1733	1733	885
N	5493	5493	3724	1596	4516	3724	3724	1596

Notes: Variables labeled in italics are reported in logs. Total Revenue represents revenues earned by the firm for domestic and export sales. Capacity Utilization is self reported. Investment includes all investment by the firm in billions of Baht. X_{ijt} is a dummy indicating a firm's export status. Standard errors clustered by firm are in parentheses.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

Table A3
Testing for substitution patterns.

	Domestic Revenue$_{ijt}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
tfp_{ijt}	0.691***	0.680***	0.683***	0.683***	0.679***	0.677***	0.672***	0.674***
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.045)	(0.045)	(0.029)
CU_{ijt}	0.808***	0.705***	0.917***	0.915***	0.905***	0.721***	0.722***	0.737***
	(0.109)	(0.107)	(0.216)	(0.216)	(0.217)	(0.201)	(0.205)	(0.144)
τ_{ijt}		0.227***	0.154	0.149	0.134	0.144	0.136	0.267**
		(0.082)	(0.102)	(0.105)	(0.106)	(0.104)	(0.104)	(0.128)
$\tau_{ijt}*CU_{ijt}$			−0.101	−0.100	−0.098	−0.047	−0.051	−0.057
			(0.121)	(0.121)	(0.121)	(0.113)	(0.114)	(0.080)
$\tau_{ijt}*X_{ijt}$			0.159	0.160	0.160	0.194**	0.203**	0.195***
			(0.101)	(0.101)	(0.101)	(0.096)	(0.096)	(0.058)
X_{ijt}		0.542***	0.247	0.244	0.244	−0.059	−0.080	−0.070
		(0.075)	(0.191)	(0.191)	(0.191)	(0.187)	(0.187)	(0.114)
$X_{ijt}*CU_{ijt}$			−0.596*	−0.593*	−0.578*	−0.551*	−0.563*	−0.564**
			(0.318)	(0.318)	(0.317)	(0.306)	(0.309)	(0.224)
$X_{ijt}*CU_{ijt}*\tau_{ijt}$			0.303*	0.302*	0.293*	0.304*	0.305*	0.306***
			(0.178)	(0.178)	(0.178)	(0.168)	(0.169)	(0.119)
$IndustryImports_{jt}$				0.032	−0.059	−0.141	−0.138	−0.067
				(0.083)	(0.086)	(0.093)	(0.094)	(0.072)
$DomesticSales_{jt}$					0.125*	0.136*	0.140*	0.054
					(0.071)	(0.080)	(0.082)	(0.061)
Credit constraint$_{ijt}$							−0.182	−0.167***
							(0.114)	(0.062)
Owner = family						0.051	0.059	0.053
						(0.117)	(0.116)	(0.060)
Owner = private invest						0.686***	0.684***	0.694***
						(0.107)	(0.106)	(0.055)
Owner = private corp						1.654***	1.640***	1.651***
						(0.184)	(0.184)	(0.094)
Owner = multinational						1.291***	1.276***	1.281***
						(0.167)	(0.167)	(0.088)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IndustryXYear FE	No	No	No	No	No	No	No	Yes
R^2	0.261	0.284	0.285	0.285	0.286	0.370	0.371	0.377
Firms	1759	1759	1759	1759	1759	1480	1480	1480
N	5453	5453	5453	5453	5453	4903	4903	4903

Notes: Variables labeled in italics are reported in logs. Domestic and Export Revenue represent revenues earned by the firm for domestic and export sales, respectively. Tariffs are 4-digit ISIC tariffs averaged across Thai export destinations weighted by total exports. TFP is estimated using Olley and Pakes (1996). Capacity Utilization is self reported. $DomesticSales_{jt}$ are the total sales by all domestic firms in an industry. $IndustryImports_{jt}$ are total imports by consumers in industry j. Credit Constraint is an indicator for whether the firm considers "access to credit to be an obstacle to doing business." X_{ijt} is a dummy indicating a firm's export status. Standard errors clustered by firm are in parentheses.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Appendix B. Supplementary materials

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.euroecorev.2014.06.003.

Table A4
Estimating the probability of high tfp and low sales.

	Conditional Logit: Pr[Violate Standard Theory]						
	(1)	(2)	(3)	Odds	(4)	(5)	Odds
CU_{ijt}	0.772*	0.805*	1.008**	3.916**	0.608***	0.585**	1.794**
	(0.442)	(0.442)	(0.432)	(1.872)	(0.231)	(0.232)	(0.416)
X_{ijt}		−0.354	−0.236	0.67	0.012	0.010	1.011
		(0.322)	(0.325)	(0.213)	(0.096)	(0.096)	(0.097)
$log(TotalAssets)_{ijt}$			−0.516***	0.496***	−1.162***	−1.169***	0.311***
			(0.147)	(0.081)	(0.042)	(0.042)	(0.013)
Credit constraint$_{ijt}$						−0.323***	0.724***
						(0.118)	(0.085)
Owner=family					0.279**	0.299**	1.349**
					(0.118)	(0.118)	(0.159)
Owner=private invest					−0.087	−0.079	0.924
					(0.106)	(0.107)	(0.098)
Owner=private corp					−0.098	−0.113	0.893
					(0.220)	(0.221)	(0.197)
Owner=multinational					−0.019	−0.033	0.968
					(0.189)	(0.189)	(0.183)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	No	No	No
Industry FE	No	No	No	No	Yes	Yes	Yes
IndustryXYear FE	No	No	No	No	Yes	Yes	Yes
R^2	0.008	0.009	0.019	0.127	0.369	0.370	0.370
N	2050	2050	2050	2053	4848	4848	4848

Notes: Odds stands for the odds ratio of the preceding column specification. This column converts the logit estimates into the impact on probabilities from a change in the dependent variable. For instance the Odds column estimating a coefficient of 3.916 on capacity suggests that a firm utilizing 100% of its capacity nearly 4 times likelier to violate standard theory. Standard errors clustered by firm in parentheses.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

References

Ahn, J., McQuoid, A., 2012. Capacity constrained exporters: micro evidence and macro implications. Mimeo, Columbia University.
Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. 58, 277–297.
Berman, N., Berthou, A., Hericourt, J., 2011. Export dynamics and sales at home. CEPR, Discussion Paper.
Besanko, D., Doraszelski, U., 2004. Capacity dynamics and endogenous asymmetries in firm size. RAND J. Econ. 35, 23–49.
Bhagwati, J., Ramaswami, V., 1963. Domestic distortions, tariffs and the theory of optimum subsidy. J. Polit. Econ. 71, 44–50.
Blum, B., Claro, S., Horstmann, I., 2013. Occasional and perennial exporters. J. Int. Econ. 90, 65–74.
Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. J. Econ. 87, 115–143.
Bresnahan, T., Suslow, V.Y., 1989. Short-run supply with capacity constraints. J. Law Econ. 32, 11–42.
Clementi, G.L., Hopenhayn, H.A., 2006. A theory of financing constraints and firm dynamics. Q. J. Econ. 121, 229–265.
Corcos, G., Gatto, M.D., Mion, G., Ottaviano, G., 2012. Productivity and firm selection: quantifying the new'gains from trade. The Econ. J. 122, 754–798, http://dx.doi.org/10.1111/j.1468-0297.2011.02487.x.
Dhingra, S., 2010. Trading away wide brands for cheap brands. Mimeo, Princeton University.
Dixit, A., 1979. A model of duopoly suggesting a theory of entry barriers. Bell J. Econ. 10, 20–32, http://dx.doi.org/10.2307/3003317.
Dixit, A., 1980. The role of investment in entry-deterrence. Econ. J. 90, 95–106.
Fernandes, A., Pakes, A., 2008. Factor utilization in indian manufacturing: a look at the world bank investment climate surveys data. NBER, Working Paper.
Hallaert, J.-J., Munro, L., 2010. Binding constraints to trade expansion. OECD Trade Policy Working Papers 94.
Holmes, T., Stevens, J., 2010. An alternative theory of the plant size distribution with an application to trade. NBER Working Paper.
Knittel, C.R., Lepore, J., 2009. Tacit collusion in the presence of cyclical demand and endogenous capacity levels. Int. J. Ind. Organ. 28, 131–144.
Kreps, D., Scheinkman, J., 1983. Quantity precommitment and bertrand competition yield cournot outcomes. Bell J. Econ., 326–337.
Krishna, P., Panagariya, A., 2000. A unification of second best results in international trade. J. Int. Econ. 52, 235–257.
Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Rev. Econ. Stud. 70, 317–341.
Levitan, R., Shubik, M., 1972. Price duopoly and capacity constraints. Int. Econ. Rev. 13, 111–122.
Liu, Y., 2012. Capital adjustment costs: implications for domestic and export sales dynamics, Mimeo.
Maggi, G., 1996. Strategic trade policies with endogenous mode of competition. Am. Econ. Rev. 86, 237–258.
Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71, 1695–1725.
Melitz, M.J., Ottaviano, G.I., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75, 295–316.
Nocke, V., Yeaple, S., 2006. Globalization and the size distribution of multiproduct firms.
OECDReport, 2012. Estimating the constraints to agricultural trade of developing countries. OECD Policy Dialogue on Aid for Trade.
Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263–1297.
Spence, A.M., 1977. Entry, capacity, investment and oligopolistic pricing. Bell J. Econ. 8, 534–544.
Staiger, R.W., Wolak, F.A., 1992. Collusive pricing with capacity constraints in the presence of demand uncertainty. RAND J. Econ. 23, 203–220.

Suslow, V., 1986. Estimating monopoly behavior with competitive recycling: an application to Alcoa. RAND J. Econ. 17, 389–403.
Syverson, C., 2007. Prices, spatial competition and heterogeneous producers: an empirical test. J. Ind. Econ. 55, 197–222.
UNESCAP, 2011. Building the Productive Capacity of the Least Developed Countries. United Nations Economic and Social Commission for Asia and the Pacific Annual Report, pp. 153–185.
Vannoorenberghe, G., 2012. Firm-level volatility and exports. J. Int. Econ. 86, 57–67.
Wernerfelt, B., 1984. A resource-based view of the firm. Strateg. Manag. J. 5, 171–180.
WTO, 2007. Trade Policy Review, Report by the Secretariat, Thailand. WTO WT/TPR/S/191.