DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY

Gains from Trade When Firms Matter

Citation
Melitz, Marc J., and Daniel Trefler. 2012. Gains from trade when firms matter. Journal of Economic Perspectives 26(2): 91-118.

Published version
https://doi.org/10.1257/jep.26.2.91

Link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:10914282

Terms of use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles (OAP), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility
https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story
The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

Gains from Trade when Firms Matter

Marc Melitz
Daniel Trefler[1]

Introduction

The gains from long-distance international trade have been understood and exploited since prehistoric times. Our pre-urban ancestors were benefitting from long-distance trade in obsidian some 10,000 years ago, Plato's Academy was built on the profits of Athenian silver exports, and Rome was not built in a day partly because goods moved too slowly in the vast Roman trade network.

Much has changed since Roman times. Whereas trade was once dominated by the movement of goods that could only be produced, harvested or mined regionally, today the trade landscape is dominated by two striking facts. The first is the rise of intra-industry trade, that is, two-way trade in similar products. Chinese consumers can now buy a midsize car from Toyota (Japan), Kia (Korea), General Motors (United States), and Chery (China). Ditto for consumers in Japan, Korea, and the United States. The second striking fact is that world trade is dominated by huge and extraordinarily productive firms.

These two facts combined have transformed the way we think about the gains from trade. In the past we focused on gains that stemmed either from endowment differences (wheat for iron ore) or inter-industry comparative advantage (Ricardo's cloth for port). Today we focus both on the variety gains from intra-industry trade and on the efficiencies associated with shifting labor and capital out of small, less productive firms and into large, more productive firms.

In what follows we describe the gains from intra-industry trade with particular emphasis on the consequences for the firms involved, including their workers and owners. The recent and rapid reduction in trade costs ('Globalization') has impacted welfare in many different ways. Here we focus on three sources of gains from intra-industry trade. The first source is driven by economies of scale on the production side along with product differentiation in the eyes of the consumer.[2] Intra-industry trade allows firms to increase the scale of production (lowering average costs) while giving consumers more product choices (as well as lower prices).

The second source of gains stems from a 'reallocation' effect at the firm level that was described by Melitz (2003) and Bernard et al. (2003). Essential to these gains is the notion of firm heterogeneity i.e., that performance varies across firms (because, for example, some firms are more productive than others). Globalization generates both winners and losers among firms

[1] Financial support for Trefler's research is from the Social Sciences and Humanities Research Council of Canada (SSHRC). His research would not have been possible without the tremendous support of John Baldwin and Alla Lileeva.

[2] Paul Krugman earned the Nobel Prize in 2008 in large part for his work highlighting how economies of scale and product differentiation lead to intra-industry trade and this "new" source of gains from trade. This research is developed in Krugman (1979, 1980) and Helpman and Krugman (1985). The press release by the Nobel committee is available at http://nobelprize.org/nobel_prizes/economics/laureates/2008/press.html. Helpman (2011, Ch. 4) provides an accessible overview of this theory and the associated empirical work.

in the same sector and these effects are magnified by heterogeneity. Better performing firms thrive and expand, while worse performing firms contract and even shut down. This generates a new source of gains from trade: as production is concentrated towards better-performing firms, the overall efficiency of the sector improves. An analogy with high school math may help here. Imagine that your teacher allows you to drop your worst test and put more weight on your best test. This raises your average, just as globalization raises average sectoral efficiency. Better performing firms not only expand domestically, they expand internationally because they have a greater incentive to engage in the global economy, either by exporting, by outsourcing some of their intermediate production processes abroad, or by building plants abroad (multinationals). These are options that require up-front fixed costs so that, as in Melitz (2003), only the best performing firms have the sales levels that would justify incurring such fixed costs.

The third source of gains from trade deals with innovation. The development of new productivity- and profit-enhancing products and processes also involves substantial fixed costs. The larger scale of production associated with trade integration makes it more attractive for some firms to incur these up-front investments, thereby raising their productivity and profits, and thereby increasing the gains from trade. See Grossman and Helpman (1991) and Lileeva and Trefler (2010) as well as Bustos (2010) and Aw, Roberts and Xu (2011). Note that this third source of gains deals with *within-firm* efficiency; in contrast, the second source of gains above deals with *between-firm* or allocative efficiency.

This paper reviews these three sources of gains from trade both theoretically and empirically. Our empirical evidence will be centered on the experience of Canada following its closer economic integration with the United States (the largest example of bilateral intra-industry trade in the world!); but we will also describe related evidence for other countries.

The related literature is huge. Here we focus on firms that expand internationally via exporting as in Melitz (2003) and Bernard et al. (2003). Another related research topic analyzes how firm boundaries evolve across borders as firms locate key parts of their production chain abroad (e.g. outsourcing and multinationals). Antras and Rossi-Hansberg (2010) and Helpman (2011, Ch.6) provide recent surveys of this important new research topic.

Intra-Industry Trade: A Few Facts and Examples

The Rise of Intra-Industry Trade

Figure 1 documents the rise of intra-industry trade as a share of overall trade for the world over the last half-century.[3] In order to measure intra-industry trade, one needs to start with an industrial classification system that assigns all documented trade flows to a particular "industry". One can then categorize all bilateral trade flows as either intra-industry (two way trade within the same industry classification code) or inter-industry (imports and exports in separate industry codes). The United Nations uses such a classification system known as the Standard International Trade Classification, or SITC, to categorize all world trade flows that it reports. In its most detailed form, there are 1,161 separate industry codes, but these industries are often aggregated into a smaller subset of industries. Figure 1 shows the time trend for the share of intra-industry trade according to this most detailed classification, and a more aggregated

[3] The data in Figure 1 come from Bruhlhart (2009). That paper documents the many facets of the rise of intra-industry trade from 1962-2006, and describes all the associated measurement issues. We thank Marius Brulhart for generously sharing his data.

version (with 59 separate industry codes). Mechanically, the share of intra-industry rises with the level of aggregation for the industrial classification system (if there was just a single aggregate industry code, then all trade would be "intra" to this aggregated industry). However, the time trends for both series are very similar: intra-industry trade (relative to inter-industry trade) grew rapidly from 1962 to the mid-1990s, before stabilizing at a substantially higher level. Looking across countries, there is a strong correlation between the share of intra-industry trade and development: as countries industrialize, they produce and export differentiated manufactured goods that are similar to other brands of goods that are imported. However, it is not just the richest countries whose trade is dominated by intra-industry trade. Some of the countries with the highest shares of intra-industry trade in 2000 were newly industrializing nations such as the Czech Republic (77%), the Slovak Republic (76%), Mexico (73%), and Hungary (72%). For comparison, the U.S. share of intra-industry trade in 2000 was 69%.[4] Most recently, China's share of intra-industry trade has risen above the 50% mark.

Why does a country both export and import goods that are similar? One major reason is that those goods are similar, but not identical. Consider the world trade in automobiles that was previously mentioned. Consumers in a car-producing country are not limited to buying the car models that are produced domestically: many of those consumers choose to buy different models that are produced elsewhere and imported. Brand proliferation is then moderated by economies of scale: producing the first unit of a new good is substantially more expensive than the production cost for the subsequent ones (especially if the good is produced on a very large scale). We now highlight how the combination of product differentiation and economies of scale generates intra-industry trade and gains from such trade using a highly stylized theoretical example. The goal is to motivate the first source of gains from trade that we listed in the introduction, and to show how such gains arise even when two identical countries open up to trade. This provides a stark contrast with the gains from inter-industry trade which arise only from exploiting differences across countries such as differences in technology (some countries are relatively better than others at producing certain types of goods) or differences in factor supplies (production of some goods requires certain factors of production that are relatively more abundant, hence cheaper, in some countries relative to others).[5]

Gains from Intra-IndustryTrade: An Initial Example

In our highly stylized example, two identical countries produce highly differentiated widget varieties subject to the same technology that features economies of scale. To be specific, assume that one worker can produce 1 widget, but that production of any new variety of widgets requires 4 workers to cover fixed overhead costs: this implies decreasing average costs of production as the fixed cost is spread over an increasing number of output units (hence the economies of scale). Also to be specific, suppose that both countries have a fixed supply of 12 workers. If they do not trade, then each country can produce:

(a) 8 units of 1 variety, or

(b) 2 units each of 2 different varieties.

Allowing countries to trade leads to a new possibility that is better than either (a) or (b). Suppose that each country produces 8 units of 1 variety (as in (a)) and exports 4 of these units to the other

[4] See the OECD Economic Outlook (2002, Ch. 6). All the listed shares of intra-industry trade are based on the aggregated industrial classification code from Figure 1.

[5] Differences in consumer tastes across countries can also be a source of inter-industry trade but most of the world's inter-industry trade is driven by supply-side differences.

country. Consumers are now consuming 4 units of the home variety and 4 units of the foreign variety. This is preferred to either of the no-trade production plans above. Compared to (b), there are the same number of varieties (2), but more of each variety (4 versus 2). Compared to (a), there are the same number of units (8), but more varieties (2 versus 1). Note how trade gives consumers a better tradeoff between more consumption units and more product variety. Economic integration allows production of each individual variety to be consolidated for the whole integrated market; given increasing returns to scale, this reduces average production costs. At the same time, product variety increases because consumers can buy varieties produced anywhere in the integrated market.

As Figure 1 illustrates, the rise in intra-industry trade – two-way trade in varieties of a good – has been a dominant trend in world trade at least since the 1960s. Two historic examples serve to bring home this point about the gains from variety.

Trade Agreements and Economic Integration

One of the most salient examples of economic integration between similar countries has occurred between the United States and Canada over the last 50 years. This integration started in the mid- sixties for a single sector with the implementation of the North American Auto Pact. Prior to 1964 (the signing of the Pact), most car models were produced both in the United States (for U.S. consumers) and in Canada (for Canadian consumers). High tariffs on auto trade made it uneconomical to export most car models across the border. Since the Canadian auto market was roughly one-tenth the size of the U.S. market, this implied substantial scale disadvantages for production in the Canadian market: labor productivity there was about 30% below the U.S. level.[6] The Pact established a free trade area for autos that allowed manufacturers to consolidate the production of particular car models in one country, and export that model to consumers in the other country. For example, General Motors cut in half the number of models assembled in Canada. However, auto production in Canada increased as the remaining models produced in Canada supplied the U.S. market as well as the Canadian one. Canadian automotive exports to the United States increased from $16 million in 1962 to $2.4 billion in 1968. That same year, U.S. automotive exports to Canada were valued at $2.9 billion: intra-industry trade in action. Today, $85 billion worth of automotive products cross the U.S.-Canada border each year – roughly half in each direction: that represents over $350 million of cross-border exchanges per average work day! The productivity gains associated with this consolidation were also substantial: by the early 1970s, the 30% labor productivity shortfall relative to the United States had disappeared.

Later on, this transformation of the automotive industry was extended to include Mexico. In 1989, Volkswagen consolidated its North American operations in Mexico, shutting down its plant in Pennsylvania. This process continued with the implementation of NAFTA (the North American Free Trade Agreement between the United States, Canada, and Mexico). In 1994 Volkswagen started producing the new Beetle for the whole North American market in that same Mexican plant.

This consolidation in response to closer economic integration with the United States was not limited to the auto industry. Baldwin, Beckstead and Caves (2002), Baldwin, Caves, and Gu (2005), and Baldwin and Gu (2006b) document a dramatic reduction in the product

[6] The U.S. market was large enough that assembly lines could be dedicated to one particular car model, while Canadian assembly lines had to switch across models, involving costly down-time and reconfiguration costs. It also forced Canadian plants to hold substantially higher inventory levels.

diversification of Canadian plants across the entire Canadian manufacturing sector following the implementation of the Canada-U.S Free Trade Agreement in 1989 (NAFTA's precursor). This reduction in product diversification implies that Canadian plants concentrated their production on a smaller subset of products. Indeed, Baldwin, Caves and Gu (2005) also report that the decrease in product diversification was accompanied by substantial increases in production runs for individual products.[7]

Another prominent example of economic integration has taken place in Europe over the last half century. In 1957, the major countries of Western Europe established a free trade area in manufactured goods (the European Economic Community or EEC). The result was a rapid growth of trade, especially intra-industry trade. Trade within the EEC grew twice as fast as world trade during the 1960s, and intra-industry trade as a share of EEC trade more than doubled from 1960 to 1990. Economic integration has continued in Europe as more countries have joined the free trade area (which is today called the European Union or EU) and as a subset of EU countries have adopted a common currency (the Euro) since 1999. Eurozone members have experienced strong trade growth (especially intra-industry trade growth) relative to non-EU countries and even relative to EU countries that have not adopted the Euro.

A substantial portion of the increased trade that comes with economic integration also delivers increased product variety to consumers. Around the world, countries that are part of a free trade area (such as NAFTA and the EU, but also lots of other regional agreements around the world) exchange substantially more products with one-another than with countries outside of their free trade area. Even the adoption of the Euro (among countries that had already nearly completed integrated into a single trade area) led to a substantial increase in the number of products traded within the Euro area. The benefits of this additional product variety for consumers are also substantial. Broda and Weinstein (2006) estimate that the number of available products in U.S. imports tripled in the thirty-year time span from 1972 to 2001. They further estimate that this increased product variety for U.S. consumers represented a welfare gain equal to 2.6% of U.S. GDP! Feenstra (2010) estimates that the worldwide consumer gains are equal to between 9% and 15% of world GDP.

Gains from Trade with Heterogeneous Firms

Our stylized example with two countries and identical production technologies highlighted one important channel for the gains from intra-industry trade via the consolidation of production (and the associated expansion of consumed varieties). However, that simple example cannot help us predict how individual firms will respond to trade, and misses the other two sources of gains from trade that we mentioned in the introduction. In order to analyze those other two sources, we first need to develop a model of trade with heterogeneous firms (i.e. where performance varies across different firms). We can then capture how firms with different characteristics respond differently to trade. Consider the case from the previous example where opening to trade leads to a transition from production plan (a) where each country produces 2 varieties to production plan (b) where each country produces 1 variety. In the real world, those products are associated with the firms that produce them. Opening up to trade therefore implies

[7] This process is even evident in the Canadian wine industry, an industry that exclusively produced low-end wines that could not possibly compete with Californian giants such as Gallo. In response to the Agreement, Canadian manufactures dramatically reduced the number of varietals produced and focused on the varietals used to produce ice wine. The industry is now healthier than ever. See Beamish and Celly (2003).

that 1 of the 2 firms in each country shuts down while the remaining firms expand production (from 2 units to 8 units). Which firms expand and which ones exit? In this example, firm labels are inconsequential because all the firms look exactly the same. We now investigate how these predictions change once we introduce performance differences across the firms.[8]

Demand, Cost, and Profit

Consider an industry where many firms compete by offering different products that are nevertheless close substitutes for one another (at least, as compared to products in other industries). For simplicity, we associate a firm with a single product.[9] The demand for each firm's product is represented by a standard downward sloping demand curve as represented in panel (a) of Figure 2. Due to the competition across firms and products, this is a *residual* demand curve: this demand depends on the behavior of other competing firms in the market. Increases in competition, associated with increases in the number of competing products and/or lower average prices of those competing products will shift in this residual demand (see graph). On the other hand, increases in the overall size of the market (holding the number of products and their prices fixed) will rotate the demand curve out (see graph). This implies that an increase in market size (holding competition fixed) would lead to the same proportional increase in quantity sold for the firm at any given price.[10]

On the production side, the economies of scale arise from a fixed cost that a firm must initially incur to develop a product and set up its initial production. This cost, associated with firm entry into the market, is sunk once a firm initiates production. The marginal cost of production is assumed constant.

This technology is represented graphically in panel (b) of Figure 2 with a flat marginal cost curve ($MC=c$) and a decreasing average cost curve (AC). The shaded area represents the total production costs for a firm producing q units of output. This cost can be divided into operating costs (the darker shaded area below the MC curve), and the fixed entry cost (the lighter shaded area above the MC curve).[11]

In this model, every firm has market power because it produces its own version (or brand) of a product that is differentiated – hence the downward-sloping firm residual demand curve. However this firm's market power is limited by the responses of other firms in the market (entry and pricing decisions) that affect the firm's residual demand. To keep the analysis as simple as possible, we assume that there are a large number of competing firms, so large that no firm can 'move the market' in the sense of doing something that could alter its residual demand curve. Each individual firm therefore maximizes profits subject to its residual demand curve.

---

[8] The model we introduce in this section is a simplified version of the one developed in Melitz and Ottaviano (2008).

[9] Eckel and Neary (2010), Bernard et al (forthcoming), and Mayer et al (2011) develop related models that further incorporate the firms' production and export decisions across multiple products.

[10] The equation for the demand facing a firm that is used in what follows is $Q = S\left[(1/n) - b(\bar{p} - p\}\right]$, where Q is the quantity of output demanded, S is the total output of the industry, n is the number of firms in the industry, $b > 0$ is a constant term representing the responsiveness of a firm's sales to its price, p is the price charged by the firm itself, and $\bar{p}$ is the average price charged by its competitors. This demand equation may be given the following intuitive justification: If all firms charge the same price, each will have a market share $1/n$. A firm charging more than the average of other firms will have a smaller market share, whereas a firm charging less will have a larger share.

[11] Note that the distance between the AC and MC curves equals the portion of the fixed entry cost f that is paid per unit of output produced.

This leads to the standard outcome whereby each firm chooses an output level q that equalizes marginal cost and marginal revenue, as shown in Figure 3. That graph shows the same cost and demand curves from Figure 2, and adds the marginal revenue curve (*MR*) associated with the demand curve. The output choice q leads to a corresponding price p on the demand curve, and a markup $p - c$ over marginal cost. The two shaded areas represents the firm's operating profit π^o (revenue $p \times q$ minus operating cost $c \times q$). This operating profit does not take into consideration the fixed cost f that is paid upon entry. The firm's net profit is then $\pi = \pi^o - f$, and is represented by the darker shaded area (recall that the lighter shaded area between the *AC* and *MC* curves represents the fixed cost f).

In Figure 3, the cost and demand curves are drawn such that the firm's net profit is positive. If all firms had access to the same technology (same entry cost f and marginal production cost c), then this could not represent a long run equilibrium with entry: a prospective entrant would recognize that they could earn a positive net profit (an operating profit exceeding the entry cost) and would therefore enter. More competition via entry would shift in the residual demand curve as we previously discussed, leading to an equilibrium where all firms earn zero net profits. This is the standard monopolistic competition equilibrium.

Differences in Firm Performance

However, we want to focus on the case where firms do not have access to the same technology. We consider the case where firms have different marginal costs of production c_i. For simplicity, we still assume that demand for all products is symmetric and hence that all firms face the same residual demand curve.[12] Figure 4 shows the consequences of differences in marginal cost for firm performance. In panel (a), we see that firm 1 has a lower marginal cost (c_1) than firm 2 (c_2). This leads firm 1 to choose a higher output level than firm 2 ($q_1 > q_2$), associated with a lower price ($p_1 < p_2$). Note, however, that firm 1 still sets a higher markup than firm 2: $p_1 - c_1 > p_2 - c_2$.[13] This, in turn, entails that firm 1 earns a higher operating profit than firm 2: $\pi_1^o > \pi_2^o$ (represented by the shaded areas in panel (a) of Figure 4).[14] We assume that all firms face the same entry cost f so firm 1 also earns higher net profits: $\pi_1 > \pi_2$. (Differences in fixed costs would not affect firm behavior in terms of price and output.)

We can thus summarize all the relevant performance differences that result from marginal cost differences across firms. Compared to a firm with higher marginal cost, a firm with a lower marginal cost will: (1) Set a lower price, but at a higher markup over marginal cost, (2) produce more output, and (3) earn higher profits.

Panel (b) in Figure 4 shows how firm profit varies with its marginal cost c_i. As we just mentioned, both operating and net profit will be decreasing functions of marginal cost (the difference between the two is just the fixed entry cost f). Going back to panel (a), we see that a firm can earn a positive operating profit so long as its marginal cost is below the intercept of the

[12] Product-quality differences between firms would lead to very similar predictions for firm performance as the ones we now derive for cost differences.

[13] This is a consequence of the marginal revenue curve being steeper than the demand curve.

[14] Operating profit $\pi_i^o = p_i \times q_i - c_i \times q_i$ can be re-written as the product of the markup $p_i - c_i$ times output q_i. Firm 1 sets a higher markup and sells more output, leading to higher operating profits than firm 2.

demand curve on the vertical axis. Let c^* denote this cost cutoff. A firm with a marginal cost c_i above this cutoff is effectively "priced-out" of the market and would earn negative operating profits if it were to produce any output. Such a firm would choose to shut down and not produce (earning zero operating profit but incurring a net profit loss $-f$ due to the fixed cost). Why would such a firm enter in the first place? Clearly, it would not if it knew about its high cost c_i prior to entry and paying the fixed cost *f*.

We assume that entrants face some randomness about their future production cost c_i. This randomness disappears only *after f* is paid and is sunk. Thus, some firms will regret their entry decision as their net profit is negative. This is the case for firm 2 in panel (b).[15] On the other hand, some firms discover that their production cost c_i is very low and earn a high (and positive) net profit. Firms consider all these possible outcomes, captured by the net profit curve in panel (b) of Figure 4 when they make their entry decision. Firms anticipate that there is a range of lower costs where net profits are positive (shaded area to the left above the horizontal axis), and another range of higher costs where net profits are negative (shaded area to the right below the horizontal axis).[16] In the long run equilibrium, firms enter until their *expected* net profit across all potential cost levels c_i is driven to zero. If every cost level c_i from 0 to $c_{\max}$ is equally likely, then this equilibrium is reached when the two shaded areas are equal.[17]

The Effect of Trade: Market Integration

Panel (b) of Figure 4 summarizes the industry equilibrium for a given market size. It tells us which range of firms survive and produce (with cost c_i below c^*), and how their profits will vary with their cost levels c_i. What happens when economies integrate into a single larger market? A larger market can support a larger number of firms than a smaller market, which implies more competition in the larger market. Figure 2 showed how increased competition – absent any increase in market size – leads to an inward shift of each firm's residual demand curve. On the other hand we also saw that, holding competition fixed (the number of firms and the prices they set), a larger market rotates out the residual demand curves for all firms. Putting these two effects together gives us the combined effect of increased market size (including the associated increase in competition) on the firms' residual demand curve. This is depicted in panel (a) of Figure 5 as the shift from demand curve D to D'. Notice how the demand curve shifts in for the smaller firms (lower output levels) that operate on the higher part of the demand curve: here, the effect of tougher competition dominates. However, this demand curve also rotates out for the larger firms that operate on the lower part of the demand curve: here, the effect of the larger market size dominates.

Panel (b) of Figure 5 shows the consequences of this demand change for the operating profits of firms with different cost levels c_i. The decrease in demand for the smaller firms translates into a new lower cost cutoff $c^{*\prime}$: Firms with the highest cost levels (above $c^{*\prime}$) cannot

---

[15] Firm 2's marginal cost c_2 is below the cutoff c^*, so it still earns a positive operating profit and will therefore choose to produce and not exit.

[16] Recall that any firm with marginal cost c_i above the cutoff c^* exits and incurs a net profit loss $-f$. This is a lower bound for net profits.

[17] The region above c^* will be important below when we discuss the exporting decision.

survive the decrease in demand and are forced to exit. On the other hand, the flatter demand curve is advantageous to firms with the lowest cost levels: they can adapt to the increased competition by lowering their markup (and hence their price) and gaining some additional market share.[18] This translates into increased profits (both operating and net) for the best performing firms with the lowest cost levels c_i. Figure 5 illustrates how increased market size generates both winners and losers amongst firms in an industry. The low cost firms thrive and increase their profits and market shares, while the high cost firms contract, and the highest cost firms exit.

Although integration does not directly affect firm productivity, it nevertheless generates an overall increase in aggregate productivity as market shares are reallocated from the low productivity firms (those with high cost) to the high productivity ones (those with low cost). Again, think back to our example of your average grade for a class: if your better performing test scores are given higher weights, and you are allowed to drop your lowest score, than your average for the class will increase.

Trade Costs, Export Decisions, and Trade Liberalization

We have just modeled economic integration as an increase in market size. This implicitly assumes that this integration occurs to such an extent that a single combined market is formed. In reality, integration rarely goes that far: initial trade costs are not so high that there is no trade prior to liberalization; and liberalization does not fully eliminate those costs: it reduces them. What happens to our key prediction regarding the effect of integration on aggregate productivity via reallocations? The main punch-line is that this kind of trade liberalization has a very similar effect to the simpler case of full integration that we previously described. Partial trade liberalization generates both winners and losers: the better performing firms expand while the worse performing ones contract and the worst performing ones exit. This generates the same type of reallocation effect previously described and leads to a rise in aggregate productivity.[19]

Adding trade costs also allows us to analyze an important new implication of the theoretical model regarding the firm export decision. With trade costs, exporting is profitable only for a subset of better performing firms. There are some firms who do not export and only serve domestic consumers (so long as their operating profit from doing so is positive). We now extend our theoretical model to incorporate trade costs and this firm export decision. We can no longer analyze a single integrated (or entirely non-integrated) market: we need to jointly look at the firms' decisions in both the domestic and export markets. For simplicity, we consider a special case where both countries are symmetric, so demand conditions in both the domestic and export markets will be identical.

Assume that a firm must incur an additional trade cost *t* for each unit of output that it sells to customers across the border. This trade cost will induce a firm to set different prices in its

---

[18] Recall that the lower the firm's marginal cost c_i, the higher its markup over marginal cost $p_i - c_i$. High cost firms are already setting low markups, and cannot lower their prices to induce positive demand, as this would mean pricing below their marginal cost of production.

[19] The model developed here is a special case of a more general version analyzed by Melitz and Ottaviano (2008). That paper develops a model with multiple countries of different sizes; and with arbitrary trade costs between any country pair (though the trade costs are proportional to production costs instead of per output unit as in the current version). The paper shows that the effects of multilateral liberalization (all countries proportionally reduce trade costs) are very similar to the case of full economic integration that leads to a single larger market.

export market relative to its domestic market. This will lead to different quantities sold in each market, and ultimately to different profit levels earned in each market.[20]

Consider the case of firms located in Home. Their situation regarding their domestic (Home) market is exactly as was illustrated in Figure 4, except that all the outcomes such as price, output, and profit relate to the domestic market only. Now consider the decisions of firms 1 and 2 (with marginal costs c_1 and c_2) in the export (Foreign) market. They face the same demand curve in Foreign as they do in Home (the two countries are identical). The only difference is that the each firm's marginal cost in the export market is shifted up by the trade cost *t*. Figure 6 shows the situation for the two firms in both markets.

What are the effects of the trade cost on the firms' decisions regarding the export market? We know from our previous analysis that a higher marginal cost induces a firm to raise its price, which leads to a lower output quantity sold and to lower profits. We also know that if marginal cost is raised above a threshold level c^*, then a firm cannot profitably operate in that market. This is what happens to firm 2 in Figure 6. Firm 2 can profitably operate in its domestic market, because its cost there is below the threshold: $c_2 \leq c^*$. However it cannot profitably operate in the export market because its cost in the export market is above the threshold: $c_2 + t > c^*$. Firm 1, on the other hand, has a low enough cost such that it can profitably operate in both the domestic and export markets: $c_1 + t \leq c^*$.

We can extend this prediction to all firms based on their marginal cost c_i. Panel (a) of Figure 7 separates a firm's operating profit into a portion earned from domestic sales, and a portion earned from export sales. (Both portions are functions of a firm's marginal cost c_i as in Figure 4.) Note how the trade cost *t* shifts down the operating profit for the export market relative to the domestic market.[21] The figure shows how firm 1 earns positive operating profits from sales in both the domestic and export markets. However, firm 2 only earns positive operating profits from sales in the domestic market: If firm 2 exported, it would earn negative operating profits from this activity – and thus chooses not to do so. Any firm with cost above $c^* - t$ will be in this same situation and therefore will not export. As we have previously discussed, we know that firms with a cost above c^* would also earn negative operating profits from their operations in their domestic market. Those firms cannot profitably serve either market and exit. Panel (a) of Figure 7 summarizes these firm-level decisions: as before, the worst performing firms exit, and the best performing firms export. There is now a new intermediate range of firms who serve their domestic market but do not export.

Panel (b) of Figure 7 summarizes the effects of trade liberalization (a reduction in the trade cost *t*) for those firm decisions. The figure shows the same two operating profit curves from panel (a) both before and after (dashed curves) trade liberalization. The operating profit for the domestic market shifts down due to the increase in competition (which shifts in the residual demand curve as shown in panel (a) of Figure 2). This induces some additional exit as some of the higher cost firms that used to produce no longer earn a positive operating profit after trade liberalization. On the other hand, the operating profit for the export market shifts up due to the

[20] As each firm's marginal cost is constant (does not vary with production levels), those decisions regarding pricing and quantity sold in each market can be separated: a decision regarding the domestic market will have no impact on the profitability of different decisions for the export market.

[21] Because the only difference between the domestic and export markets is the additional per-unit trade cost *t*, the horizontal distance between the two curves is equal to the trade cost *t*.

direct effect of the lower trade cost.[22] The lower trade cost induces a subset of firms to start exporting. Exporting was not profitable for those firms when the trade cost was high, but becomes profitable once trade liberalization reduces that trade cost.

Putting together the downward shift in domestic operating profits and the upward shift in export operating profits, we see that trade liberalization generates both winners and losers – just as in the case of economic integration. Non-exporters lose because they only incur the losses from the lower domestic profits. Exporters, on the other hand, stand to gain as they can make up for the loss of domestic profits with profits earned from exporting. This is the case for some, but not all of the exporters.

Gains from Inter-Firm Reallocations: The Great FTA Experiment

We have seen how the introduction of performance differences across firms substantially alters predictions about the effects of trade policy (both market integration and trade-cost reductions). In this section we visit these predictions empirically by examining the impact of the Canada-U.S. Free Trade Agreement (FTA) on the distribution of Canadian manufacturing plants.[23]

Performance Differences Across Producers

When marginal costs are low we typically expect productivity to be high. Therefore, the inverse of marginal costs ($1/c$) is always proxied in empirical work by productivity. Here we will use labor productivity i.e., value added per employee, where value added is sales less the costs of intermediate inputs.

Panel (a) of Figure 8 shows that there is a remarkable degree of dispersion in plant productivity. Consider the curve labeled '1996.' It summarizes the productivity distribution of all 35,000 Canadian manufacturing plants in 1996. The horizontal axis is the log of labor productivity. The vertical axis is proportional to the share of plants with the indicated level of productivity. To get a sense of the degree of productivity dispersion, suppose that productivity at plant A is one log point higher than at plant B. This is equivalent to saying that plant A is 3 times more productive than plant B. If A is three log points more productive than B then A is 20 times more productive than B.[24] Productivity heterogeneity is more than an interesting theoretical possibility: it is a pervasive feature of all economies including, for example, the United States (Bernard, Eaton, Jensen, and Kortum 2003) and many European economies (Mayer and Ottaviano 2007).[25]

---

[22] Increased competition in the export market reduces operating profits, but this effect is dominated by the direct effect of the trade cost reduction.

[23] We use plants (a specific production location) as the basic production unit for our empirical analysis; this corresponds to the firms from the theoretical model.

[24] Let φ_A and φ_B be productivities of A and B and suppose the log of productivities are 1 point apart i.e., $\ln(\varphi_A) - \ln(\varphi_B) = 1$. From the property of logs, $\varphi_A / \varphi_B = e^1 = 2.7 \approx 3$. For a difference of 3 log points, $e^3 \approx 20$.

[25] To ensure that dispersion is driven by within-industry rather than between-industry differences in labor productivity, we scale each plant's log productivity by subtracting median log productivity for the plant's 4-digit SIC industry. Thus '0' corresponds to the log productivity of the median plant in the industry. Also, the frequencies are weighted by plant employment; otherwise, the figure is dominated by tiny plants that account for only a tiny fraction of total employment. (The median plant had only 12 employees in 1988.) Thus, for example, the average height of the 1996 curve on the interval (-1,0) is about 7%, which means that plants with log productivity between -1 and 0 account for 7% of manufacturing employment.

Figure 8 displays one significant departure from the theory. In Figure 7, only firms with marginal cost below c^* enter and survive. Correspondingly, only firms with productivity above $1/c^*$ should enter and survive. In practice, there are many other firm characteristics other than observable labor productivity that influence the decision to enter and continue operations. For example, in a dynamic model in which a firm's factory depreciates over time (the factory is the fixed cost *f* of entering), a low-productivity firm may continue operating until it has run its factory into the ground. Thus, we will always see some low-productivity firms operating, but we will not see many of these. This shows up as the steep slopes to the immediate left of 0 in Figure 8; in words, as productivity falls below median productivity, the number of plants drops precipitously.[26]

Market Integration and the Entry/Exit Decision

How does improved market integration affect this distribution? The Canada-U.S. Free Trade Agreement (FTA) provides a great place to look because in many ways it is a perfect natural experiment. For one, the FTA policy experiment is clearly defined: it dealt only with market integration and was not part of a larger package of macroeconomic reforms that typically accompany trade liberalization. For another, it was unanticipated: a general election was fought on the issue one month before the FTA was to be signed into law and pollsters unanimously predicted that the ruling party – along with the FTA – would be defeated. See Brander (1991) and Thompson (1993). Finally, the FTA is the only major agreement between two developed countries that has been extensively studied.

The FTA came into effect on January 1, 1989. Panel (a) of Figure 8 shows the distribution of plants both in 1988 (the last year before implementation of the FTA) and in 1996 (when the FTA had largely been implemented and firm adjustments completed). The obvious feature of the panel is that the distribution of firms shifted rightward. Between 1988 and 1996 the share of low-productivity plants in manufacturing declined and the share of high-productivity plants rose. As we shall now see, this was largely due to the reallocation mechanism across plants that we outlined above.

The first of these mechanisms that we examine is the fall in c^*, or equivalently, the rise in the break-even level of productivity. One can examine this mechanism by looking either at exit rates or at entry rates. A plant may not exit until it has completed the multi-year process of depreciating its fixed capital; in contrast, entry rates adjust quickly to shocks such as the FTA. Thus, entry-rate plots give a slightly clearer picture of the FTA-induced changes in the break-even level of productivity. Panel (b) of Figure 8 displays entry rates for the pre-FTA period 1980-1988 and for the FTA period 1988-96. There was a striking FTA-period decline in the entry rates of low-productivity plants. To use a sports analogy, in the pre-FTA period even the lowest-productivity plants made the cut and joined the team while in the FTA period such low-productivity plants no longer made the cut.[27]

In terms of formal econometric analysis of how the FTA raised the productivity threshold, researchers typically estimate exit probits. (Entry probits are conceptually difficult to estimate.) Baggs (2005) and Baldwin and Gu (2006a) show that the FTA tariff cuts raised exit

[26] Similar comments apply below to our discussion of the Figure 8(c) export threshold.

[27] Bernard and Jensen (1999), Trefler (2004), Lileeva (2008) and Lileeva and Trefler (2010) all point out that one must look not just at pre-FTA l*evels* (as in Figure 8), but also at pre-FTA *trends*. All of the FTA results reported here hold with pre-FTA controls for both levels and trends. For example, variants of some of the panels in Figure 8 with pre-FTA trend controls appear in Lileeva (2008).

rates, as predicted in Figure 7. Further, the impact was large. Lileeva (2008) estimates that the FTA tariff cuts raised exit rates by as much as a whopping 16%, with all of the increase involving exit of non-exporters. Bernard, Jensen and Schott (2006) find similar results for U.S. plants faced with U.S. tariff reductions.

Trade Costs and the Export Decision

A central prediction of the theory is that in the presence of trade costs, only low-cost, high-productivity firms export (panel (a) of Figure 7). Panel (c) of Figure 8 shows the distribution of Canadian plants separately for exporters and non-exporters. The striking feature of the panel is that the distribution for exporters is to the right of that for non-exporters. This means that exporters are typically more productive than non-exporters, precisely as predicted by the model. Figure 8 is inspired by Bernard et al. (2003): their Figure 2 is the most important data display in international trade of the past decade. In the late 1980s, the average difference in productivity between exporters and non-exporters (after controlling for industry fixed effects) was about 40% for the Baldwin and Gu (2003) sample of all Canadian plants and about 10% for the Lileeva and Trefler (2010) sample of larger plants.[28] [29]

A much more demanding prediction of the theory deals with who will *start exporting in response to falling trade costs*. From panel (b) of Figure 7, new exporters are predicted to be more productive than the non-exporting population from which they came. To test this prediction, Lileeva and Trefler (2010) examined a sample of over 5,000 plants that had never exported prior to implementation of the FTA period. A very large percentage of these plants, 40%, started exporting in the FTA period. Lileeva and Trefler (2010) then divide up this sample into quartiles of the 1988 distribution of labor productivity.[30] For each quartile, they calculated the percentage of firms that started exporting. Figure 9 displays these percentages. As predicted, we see that the plants that started exporting after implementation of the FTA were predominantly the more productive of the pre-FTA non-exporters: the probability of exporting among those non-exporters rises significantly with the plant's relative productivity in 1988.[31]

In Figure 9 we assumed that the decision to start exporting was due to the FTA tariff cuts. One of the strengths of the approach in Lileeva and Trefler is that they have an excellent instrument for the decision to export. For each plant, Lileeva and Trefler constructed plant-level measures of the U.S. tariff cut experienced by each Canadian plant. This is done by attaching a tariff cut to each HS6 product produced by a plant and then by averaging over the plant's

[28] Since the seminal work of Bernard and Jensen (1995), a huge body of research covering dozens of countries has found that on average exporters are more productive than non-exporters. This was first confirmed for Canada by Baldwin and Gu (2003) and subsequently by Baggs (2005) and Lileeva (2008).

[29] The more critical reader will wonder why there are so many highly productive non-exporters and whether this contradicts the theory. A simple but prominent example explains why there is no contradiction. Highly productive auto parts plants often cluster around a giant auto assembly plant – Ford, General Motors, and Honda all have major auto assembly plants near Toronto, Canada that are surrounded by parts suppliers. These parts suppliers are highly productive, but do not directly export. This is clearly not a challenge to the theory: these highly productive plants are supplying parts that are built into autos that are ultimately exported to the United States: highly productive parts suppliers are 'indirect' exporters.

[30] In order to net out industry characteristics, the quartiles are defined separately for each industry.

[31] In our model, profits play a key role in all mechanisms. Baggs and Brander (2006) confirm that profits move in the expected directions. In particular, they find that falling Canadian tariffs are associated with declining Canadian profits, especially for import-competing firms, while falling U.S. tariffs are associated with increasing Canadian profits, especially for export-oriented firms.

products to generate an average tariff cut for the plant. This is a first in international trade and creates a very credible instrument, something that had been sorely missing in the literature. Using IV methods (the Local Average Treatment Effect by productivity quartiles), they show that exporting decisions were heavily influenced by the FTA tariff cuts.

To recapitulate the results of this section, we have shown that in the wake of the Canada-U.S. FTA, Canadian manufacturing productivity rose sharply. A significant portion of this productivity gain can be explained by the reallocation of resources across heterogeneous firms. First, it became tougher for low-productivity plants to enter manufacturing and more common for them to exit. Second, exporters were more productive than non-exporters and the FTA tariff cuts enticed the more productive of non-exporters into growing by becoming exporters.

Quantifying the Gains from Trade Due to Reallocation Across Heterogeneous Plants

Overall, Figures 8 and 9 provide strong evidence of the reallocation mechanism that is at the heart of the model presented above. Comparative advantage is alive and well at the plant level. How important were these mechanisms for overall Canadian manufacturing productivity growth? The answer is very important. First, consider the reallocation effect: Trade liberalization induces the bigger more productive exporting plants to expand while simultaneously inducing contractions among the smaller less productive exporters. This raises productivity by shifting market shares away from less productive firms towards more productive ones. From Lileeva and Trefler (2010), we know the extent to which the FTA increased exporting. We can therefore calculate the impact of the FTA on the overall market share shift towards exporting plants. Based on the productivity differential between exporters and non-exporters in Panel (c) of Figure 8, this share-shift led to a 4.1% increase in the average productivity of Canadian manufacturing plants.

While the U.S. tariff cuts promoted exporting by Canadian plants, the corresponding Canadian tariff cuts pressured some other Canadian plants to contract and even exit. This selection effect also generates overall productivity gains since the contracting and exiting plants (overwhelmingly non-exporters) were substantially less productive than the average Canadian manufacturing plant. Trefler (2004) estimated that this selection effect increased overall Canadian manufacturing productivity by 4.3%.[32]

Putting these numbers together, we see that the reallocation and selection effects induced by the FTA generated a productivity increase of 8.4% (= 4.3% + 4.1%) for Canadian manufacturing. This represents a massive productivity increase in just a short time – especially when one considers that this productivity gain did not come from productivity improvements at the plant level: it only came from the shifting of market shares from less- to more-productive plants.

[32] Specifically, Trefler (2004) regressed labor productivity growth in the FTA period relative to the pre-FTA period on U.S. and Canadian tariff cuts mandated by the FTA. He then showed that the *Canadian tariff cuts* raised productivity at the industry level, but not at the plant level. This means that the gains in productivity were coming from selection, rather than from improvements at the plant level. Using this approach, he finds that the FTA raised manufacturing labor productivity by 5.8% ($t = 3.79$) of which 4.3% was due to the exit associated with the Canadian tariff cuts.

Canada is not the only country to have experienced such a substantial productivity boost from reallocations driven by trade liberalization. Pavcnik (2002) studies the response of the Chilean manufacturing sector to a massive trade liberalization episode that took place from 1979-1986. She finds that two thirds of the ensuing 19% increase in productivity (another example of a massive increase in aggregate productivity) was generated by composition changes within industries due to a reallocation of market shares towards more efficient producers. The surveys by Greenaway and Kneller (2007) and Wagner (2007) summarize the connections between trade liberalization and aggregate productivity (including this reallocation effect across heterogeneous firms) for a wide range of studies and countries.

Rising Within-Plant Productivity

Our empirical analysis so far has examined only those changes in average productivity that stem from the reallocation of resources away from less-productive plants and towards more-productive plants (including the effect on exit). This is sometimes referred to as a 'between-plant' effect. However, it is also possible that the FTA raised the productivity of individual plants, which is a 'within-plant' productivity effect. At least as far back as Schmookler (1954), we have known that the larger the market, the more profitable it is for firms to invest in productivity-enhancing activities. This is because firms in large markets have the large sales volumes needed to justify incurring the high fixed costs of innovation. The FTA-mandated U.S. tariff cuts effectively increased the size of the market faced by Canadian firms. It should therefore have promoted both increased exporting *and* increased investments in productivity-enhancing technologies.

If so, then at a minimum we should see that productivity rises when firms start exporting. There is now clear evidence of this. See Baldwin and Gu (2003, 2004) and Lileeva (2008) in a Canadian context, Van Biesenbrock (2005) for a subset of African countries, De Loecker (2007) for Slovenia and Lopez (2005) for a recent survey. More recently, Lileeva and Trefler (2010) use LATE techniques to establish a causal relationship running from exporting to productivity growth, a relationship that is driven by innovation.[33] We describe their results in more detail, but first start with a short extension to the theoretical model that captures how larger markets generate incentives for some firms to innovate.

A Theory of Market Size and Firm Innovation

Consider a process innovation that requires an up-front fixed cost f_I and generates a *reduction* in marginal cost equal to Δc_I.[34] Consider the decision by a firm that produces q units of output to engage in innovation. If this firm innovates, it will save $q \times \Delta c_I$ in lower production costs. This firm will weigh that cost saving against the fixed innovation cost f_I, and innovate if the former exceeds the latter. This implies that a firm will innovate so long as it produces

[33] Bustos (2011), Aw, Roberts and Whinston (2007), and Aw, Roberts and Xu (2011) have also empirically connected exporting with increased innovation at the firm level. See Burstein and Melitz (forthcoming) for a survey of the literature that incorporates exporting and firm-level innovation.

[34] For simplicity, assume that all firms have an initial marginal cost c above Δc_I so that the marginal cost of production after innovation, $c - \Delta c_I$, is positive for all firms.

$q > f_I / \Delta c_I$ units of output. In other words, only the firms with large volumes q (i.e., those with initially lower levels of marginal cost) will find it profitable to innovate.

What happens to this firm-level innovation decision when trade is liberalized? The lower trade cost increases an exporter's sales in the export market, and thus its overall output level q. For some exporters, this increase in output will tip the balance in favor of innovation.

Empirics of Within-Firm Productivity Growth

Lileeva and Trefler (2010) look at the subset of 5,000 plants that did not export prior to 1988 and divide these plants into those who started exporting in the FTA period and those who did not. In the raw data, those who started to export experienced a much higher rate of labor productivity growth in the 1988-1996 period. Specifically, their labor productivity rose 29% more than for non-exporters.

Of course, this 29% does not take into account a potentially serious endogeneity bias: does exporting lead to increased productivity or does increased productivity lead to exporting? The FTA provides a perfect natural experiment for sorting out this causality. Using a LATE estimator with plant-level tariff cuts as instruments, Lileeva and Trefler (2010) establish that the FTA had a causal impact that raised the productivity of new exporters by 15.3%. This raised *overall* Canadian manufacturing productivity by 3.5%.[35]

Lileeva and Trefler (2010) then examine the sources of this productivity gain. It turns out that it can be traced back to investments in productivity. To investigate, the authors look at plants' adoption rates of advanced manufacturing technologies and rates of innovation. They expect to find high adoption and innovation rates for those plants that had FTA-induced productivity gains from exporting. This is exactly what they find.

Table 1 presents the results. Consider the first row, which deals with management techniques associated with Lean Manufacturing. In the period immediately after implementation of the FTA, 16% of new exporters adopted such techniques while only 6% of non-exporters did. The difference of 10 percentage points is mostly due to the FTA. This is shown in column 4 where the instrumented (LATE) estimate attributes 7 percentage points to the effects of increased exporting due to the U.S. tariff cuts. As shown in Table 1, similar results hold for other technologies and for innovation.

Other evidence on innovation and exporting is reported in Baldwin and Gu (2004). They find that relative to non-exporters, exporters invest more in R&D and training, and adopt more advanced manufacturing technologies. Also, Feinberg and Keane (2006, 2009) find that the 1983–1996 increase in trade between U.S. multinationals and their Canadian affiliates was driven largely by improved logistics management, such as adoption of just-in-time production techniques. Such techniques appear in Table 1 as 'Manufacturing Information Systems.'

A New Dimension of Heterogeneity

One of the puzzles in Figure 9 is that so many low-productivity plants started exporting in the FTA period and that so much of this new exporting was directly due to the FTA tariff cuts. There is a second puzzle that we have not yet noted: The plants that gained most from starting to export (both in terms of productivity gains and increased innovation) were primarily plants that initially had low productivity. That is, among plants that started to export, the benefit was greatest for the least-productive plants.

To see why, consider two firms with different initial productivities and suppose that both

[35] These (5,000) plants accounted for 23% of Canadian manufacturing output and 3.5% (= 15.3% × 23%).

are just indifferent between (1) exporting and investing and (2) doing neither. Indifference means that $q = f_I / \Delta c_I$ or $\Delta c_I = f_I / q$, where Δc_I is the reduction in marginal cost or the increase in productivity. The less productive of the two firms has lower sales q so that the firm's indifference means the initially less-productive firm expects a larger productivity gain Δc_I from innovation. Restated: Among the new exporters, the less productive the firm, the higher the gain from jointly exporting and investing. This is precisely what Lileeva and Trefler (2010, Table III) find. Among the new exporters, less productive firms experienced the biggest gains both in productivity and in rates of innovation/technology adoption. Thus, a simple modification of our original model can explain even such subtler features of the data.

Summary of FTA Effects

The reader will not be surprised to discover that there are other gains from trade liberalization associated with heterogeneous firms. Examples can be found in Blum et al. (2010) and Holmes and Schmitz (2010). In the context of the FTA, one last effect that has been examined involves the benefits to Canadian plants of improved access to U.S. intermediate inputs such as parts and U.S. capital goods such as machinery. The fall in the Canadian tariff enabled those plants to buy those inputs more cheaply and this contributed an additional 0.5% to Canadian manufacturing productivity growth. This effect of trade liberalization on within-firm productivity via cheaper intermediate inputs is typically even more important for developing economies. For example, Kasahara and Rodrigue (2008) and Amiti and Konings (2007) measure substantial productivity gains for Chile and Indonesia, respectively, while Topalova and Khandelwal (2011) measure substantial productivity gains for India. Goldberg et al. (2010) also report that India's tariff cuts on intermediate inputs induced Indian firms to source previously unneeded intermediate inputs from abroad, intermediates that are now being used to develop and produce new products.

We summarize the causal effects of the FTA on overall Canadian manufacturing productivity in Table 2. The top section of the table lists the effects from the 'between-plants' reallocation mechanism that we mentioned in the previous section. The bottom section of the table lists the effects of the FTA on productivity growth 'within-plants'. As we just described, there was a substantial innovation response by new exporters that contributed 3.5% to overall productivity. Of course, new exporters were not the only plants to respond to the increased export opportunities. Firms that had a history of exporting even prior to the FTA also responded to improved market access by increasing their innovation and productivity. This contributed to an overall 1.4% productivity growth for Canadian manufacturing (Lileeva and Trefler, 2010).

The last row of Table 2 shows that the grand total of our documented FTA effects is an increase in Canadian manufacturing labor productivity of 13.8%. The idea that a single government policy could raise productivity by such a large amount (and in such a short time-span) is truly remarkable!

Conclusions

Recent research into the welfare gains from increased trade liberalization have focused on three sources of gains: (1) consumer gains from increased variety, (2) productivity gains at the industry level from shifting resources away from low-productivity firms and towards high-productivity firms, and (3) productivity gains at the firm level from improved access to foreign

markets. Each of these 3 mechanisms have proven to be highly important empirically in the context of the exhaustively studied Canada-U.S. Free Trade Agreement, as well as in other less-studied contexts. It is unfortunate that there are no other FTAs that have been so extensively studied as the Canada-U.S. FTA. However, Balisteri et al. (2011) have been able to show that adding firm heterogeneity to standard computable equilibrium models of trade raises the gains from trade liberalization by a very large factor of 4. Empirical confirmation of the gains from trade predicted by models with heterogeneous firms represents one of the truly significant advances in the field of international economics.

In writing this review, we have focused on the gains from trade. Yet, the model we have developed highlights how intra-industry trade will generate both winners and *losers* – just like inter-industry trade. Particularly worrying are the implications for workers in the short-run. In the context of the Canada-U.S. FTA, Gaston and Trefler (1997) and Trefler (2004) show that there are large costs borne by workers in low-productivity firms. 12% of these workers lost their jobs. This result highlights the conflict between those who endure the short-run adjustment costs (displaced workers and struggling plants) and those who garner the long-run gains (consumers and efficient plants). Clearly, this leaves an important role for policies that provide an adequate safety net for those affected workers.[36]

References

Amiti, Mary, and Jozef Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." *The American Economic Review* 97: 1611-1638.

Antras, P., and E. Rossi-Hansberg. 2009. "Organizations and Trade." *Annual Review of Economics* 1(1): 43-64.

Aw, B. Y., M. J. Roberts, and T. Winston. 2007. "Export Market Participation, Investments in R&D and Worker Training, and the Evolution of Firm Productivity." *World Economy* 30(1): 83-104.

Aw, Bee Yan, Mark J Roberts, and Daniel Yi Xu. 2011. "R&D Investment, Exporting, and Productivity Dynamics." *American Economic Review* 101(4): 1312-1344.

Baggs, Jen. 2005. "Firm Survival and Exit in Response to Trade Liberalization." *Canadian Journal of Economics* 38(4): 1364–1383.

Baggs, Jen, and James Brander. 2006. "Trade Liberalization, Profitability, and Financial Leverage." *Journal of International Business Studies* 37(2): 196–211.

Baldwin, John R., and Wulong Gu. 2003. "Export-Market Participation and Productivity Performance in Canadian Manufacturing." *Canadian Journal of Economics* 36(3): 634–657.

[36] The blow to those workers could also be cushioned by policies that impede the reallocation process across firms. However, such policies (as opposed to policies that provide some form of direct assistance to the affected workers) would also entail a substantial long run cost. It is precisely this reallocation process that generates some of the long run gains that we have described. In addition, policies that impede the reallocation process (by making firm contractions and expansions costlier) would also reduce the potential gains to firm innovation and hence lead to less innovation and further depress the potential long run gains from trade.

———. 2006a. “Plant Turnover and Productivity Growth in Canadian Manufacturing.” *Industrial and Corporate Change* 15(3): 417–465.

———. 2006b. *The Impact of Trade on Plant Scale, Production-Run Length and Diversification*. Statistics Canada,. Working Paper.

———. 2004. “Trade Liberalization: Export-Market Participation, Productivity Growth and Innovation.” *Oxford Review of Economic Policy* 20(3): 372–392.

Baldwin, John R., Desmond Beckstead, and Richard Caves. 2002. *Changes in the Diversification of Canadian Manufacturing Firms (1973-1997): A Move to Specialization*. Statistics Canada,. Analytical Studies Branch Research Paper Series.

Baldwin, John, Richard E. Caves, and Wulong Gu. 2005. “Responses to Trade Liberalization: Changes in Product Diversification in Foreign and Domestic Controlled Plants.” In *Governance, Multinationals and Growth*, eds. Lorraine Eden and Wendy Dobson. Cheltenham, UK: Edward Elgar Publishing.

Beamish, Paul W., and Nikhil Celly. 2003. *Vincor and the New World of Wine*. Harvard Business School,. Case Study.

Bernard, A. B., and J. B. Jensen. 1999. “Exceptional Exporter Performance: Cause, Effect, or Both?” *Journal of International Economics* 47(1): 1-25.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum. 2003. “Plants and Productivity in International Trade.” *American Economic Review* 93(4): 1268-1290.

Bernard, A. B., J. B. Jensen, and P. K. Schott. 2006. “Trade Costs, Firms and Productivity.” *Journal of Monetary Economics* 53(5): 917-937.

Bernard, Andrew B., and J. Bradford Jensen. 1995. “Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987.” *Brookings Papers on Economic Activity* Microeconomics: 67–112.

———. 2004. “Why Some Firms Export.” *Review of Economics and Statistics* 86(2): 561–569.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. forthcoming. “Multi-Product Firms and Trade Liberalization.” *Quarterly Journal of Economics*.

Van Biesebroeck, J. 2005. “Exporting Raises Productivity in Sub-Saharan African Manufacturing Firms.” *Journal of International Economics* 67: 373-391.

Bloom, Nicholas, Mirko Draca, and John Van Reenen. 2011. “Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity.” *NBER Working Paper*.

Brander, James A. 1991. “Election Polls, Free Trade, and the Stock Market: Evidence from the 1988 Canadian General Election.” *Canadian Journal of Economics* 24(4): 827–843.

Broda, C., and D. E. Weinstein. 2006. “Globalization and the Gains from Variety.” *Quarterly Journal of Economics* 121(2): 541-585.

Brülhart, Marius. 2009. “An Account of Global Intra-industry Trade, 1962–2006.” *The World Economy* 32(3): 401-459.

Burstein, A., and M. Melitz. forthcoming. "Trade Liberalization and Firm Dynamics." *Advances in Economics and Econometrics: Theory and Applications, Econometric Society Monographs*.

Bustos, Paula. 2011. "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms." *American Economic Review* 101(1): 304-340.

Eckel, Carsten, and J. Peter Neary. 2010. "Multi-Product Firms and Flexible Manufacturing in the Global Economy." *The Review of Economic Studies* 77(1): 188 -217.

Feenstra, Robert C. 2010. *Product Variety and the Gains from International Trade*. Cambridge MA: MIT Press.

Feinberg, Susan E., and Michael P. Keane. 2006. "Accounting for the Growth of MNC-Based Trade Using a Structural Model of U.S. MNCs." *American Economic Review* 96(5): 1515–1558.

———. 2009. "Tariff Effects on MNC Decisions to Engage in Intra-Firm and Arms-Length Trade." *Canadian Journal of Economics* 42(3): 900-929.

Gaston, Noel, and Daniel Trefler. 1997. "The Labour Market Consequences of the Canada-U.S. Free Trade Agreement." *Canadian Journal of Economics* 30(1): 18–41.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *The Quarterly Journal of Economics* 125(4): 1727 -1767.

Greenaway, David, and Richard Kneller. 2007. "Firm Heterogeneity, Exporting and Foreign Direct Investment." *The Economic Journal* 117(517): 134-161.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, Mass. and London: MIT Press.

Helpman, E., and P. R. Krugman. 1985. *Market structure and foreign trade*. MIT Press.

Helpman, Elhanan. 2011. *Understanding global trade*. Cambridge, MA: Harvard University Press.

Holmes, Thomas J., and Jr James A. Schmitz. 2010. "Competition and Productivity: A Review of the Evidence." *Annual Review of Economics* 2: 619–642.

Kasahara, Hiroyuki, and Joel Rodrigue. 2008. "Does the Use of Imported Intermediates Increase Productivity? Plant-level Evidence." *Journal of Development Economics* 87(1): 106-118.

Krugman, P. R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9(4): 469-479.

———. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70(5): 950-959.

Lileeva, Alla. 2008. "Trade Evidence and Productivity Dynamics: Evidence from Canada." *Canadian Journal of Economics* 41(2): 360-390.

Lileeva, Alla, and Daniel Trefler. 2010. "Improved Access to Foreign Markets Raises Plant-level Productivity…For Some Plants." *The Quarterly Journal of Economics* 125(3): 1051 - 1099.

De Loecker, J. 2007. "Do Exports Generate Higher Productivity? Evidence from Slovenia." *Journal of International Economics* 73: 69-98.

Lopez, R. A. 2005. "Trade and Growth: Reconciling the Macroeconomic and Microeconomic Evidence." *Journal of Economic Surveys* 19(4): 623-648.

Mayer, T., and G. I. P. Ottaviano. 2008. "The Happy Few: The Internationalisation of European Firms: New Facts Based on Firm-Level Evidence." *Intereconomics/Review of European Economic Policy* 43(3): 135-148.

Mayer, T., M. J. Melitz, and G. I. P. Ottaviano. 2011. "Market size, Competition, and the Product Mix of Exporters." *NBER Working Paper*.

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

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

Nunn, N., and D. Trefler. 2008. "The boundaries of the multinational firm: An empirical analysis." In *The Organization of Firms in a Global Economy*, eds. E. Helpman, T. Verdier, and D. Marin. , p. 55–83.

Pavcnik, N. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean plants." *Review of Economic Studies* 69(1): 245-276.

Schmookler, Jacob. 1954. "The Level of Inventive Activity." *Review of Economics and Statistics* 36(2): 183–190.

Thompson, Aileen J. 1993. "The Anticipated Sectoral Adjustment to the Canada - United States Free Trade Agreement: An Event Study Analysis." *Canadian Journal of Economics* 26(2): 253–271.

Topalova, Petia, and Amit Khandelwal. 2011. "Trade Liberalization and Firm Productivity: The Case of India." *Review of Economics and Statistics* 93(3): 995-1009.

Trefler, Daniel. 2004. "The Long and Short of the Canada-U. S. Free Trade Agreement." *The American Economic Review* 94: 870-895.

Wagner, Joachim. 2007. "Exports and Productivity: A Survey of the Evidence from Firm-level Data." *World Economy* 30(1): 60-82.

Yeaple, S. R. 2005. "Firm Heterogeneity, International Trade, and Wages." *Journal of International Economics* 65(1): 1-20.

Figures

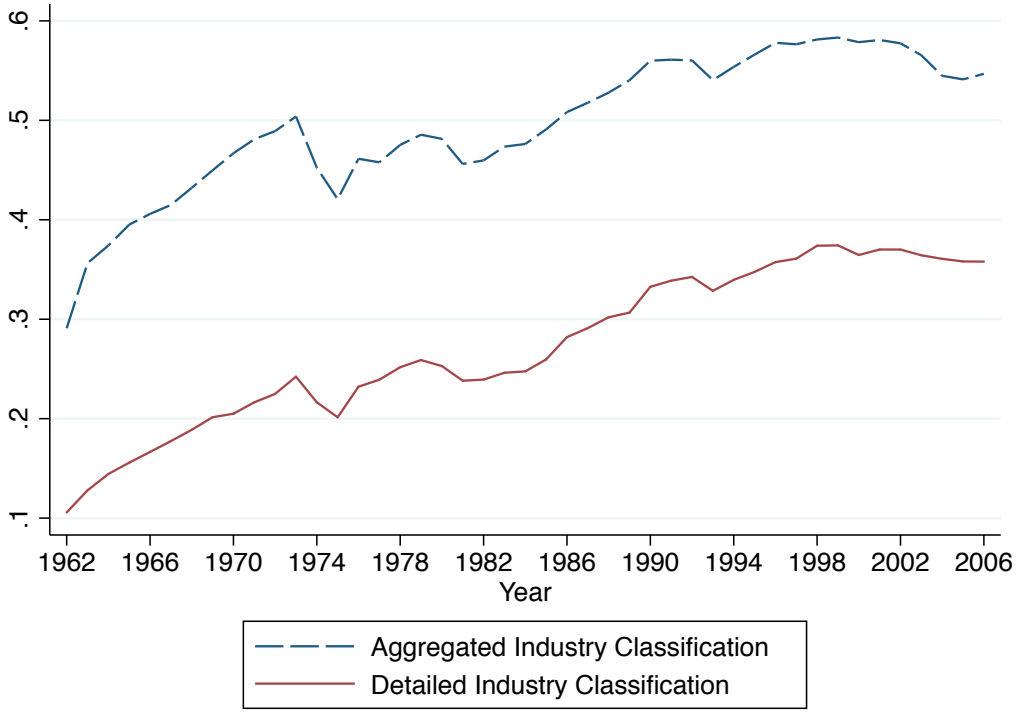

Figure 1: World Share of Intra-Industry Trade 1962-2006

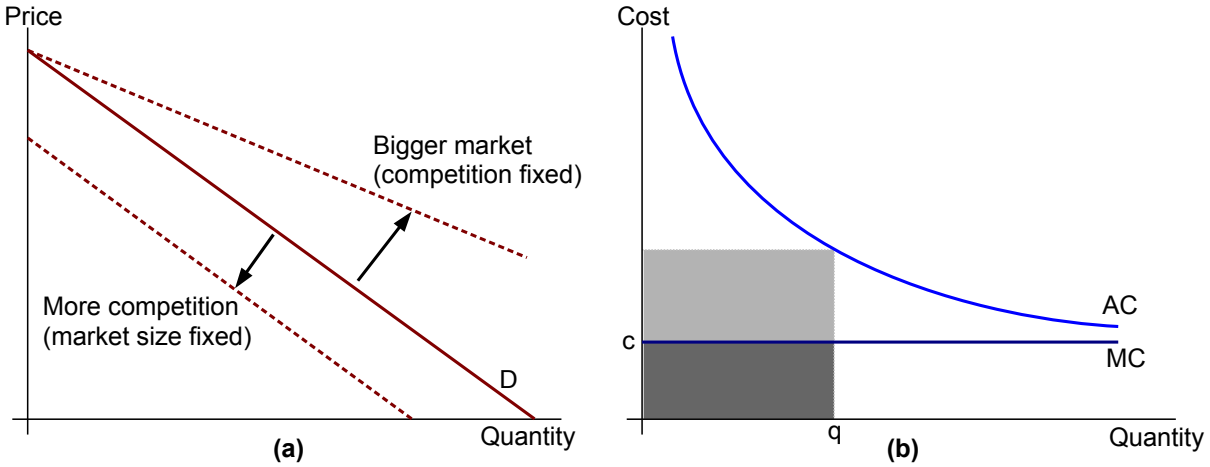

Figure 2: Demand and Cost

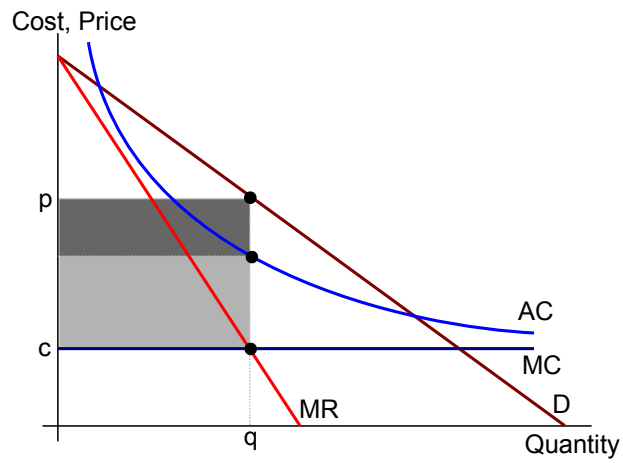

Figure 3: Profit Maximization

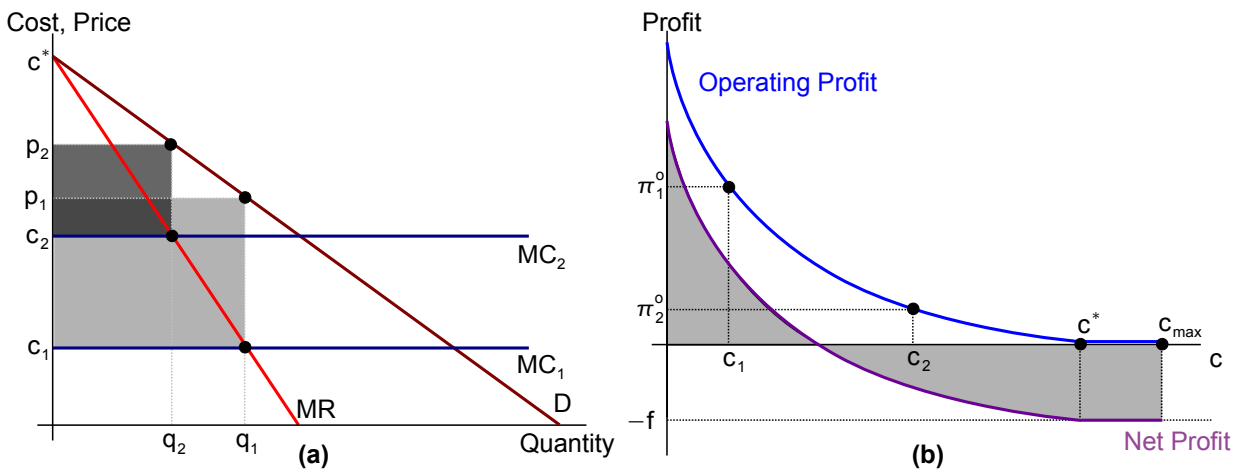

Figure 4: Performance Differences Across Firms

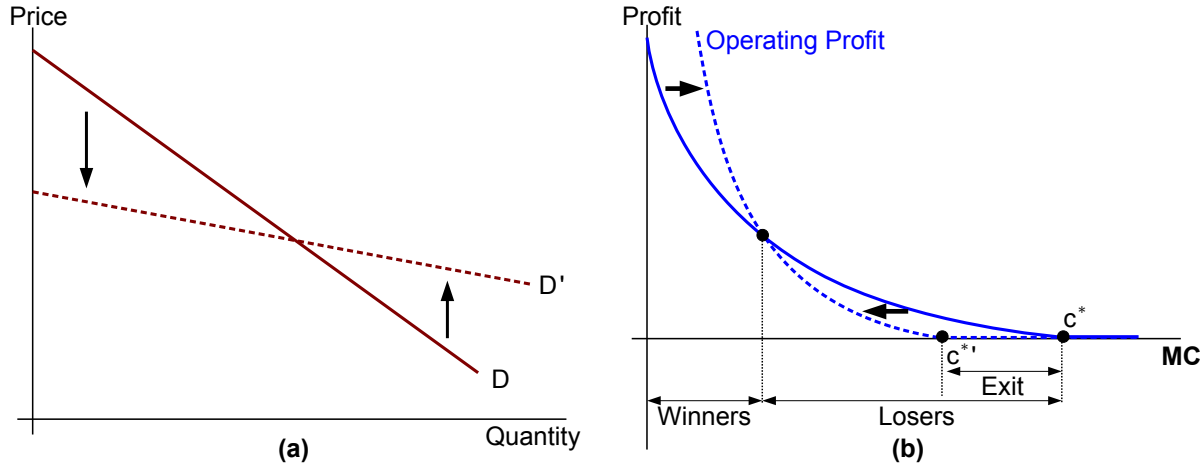

Figure 5: Winners and Losers from Market Integration

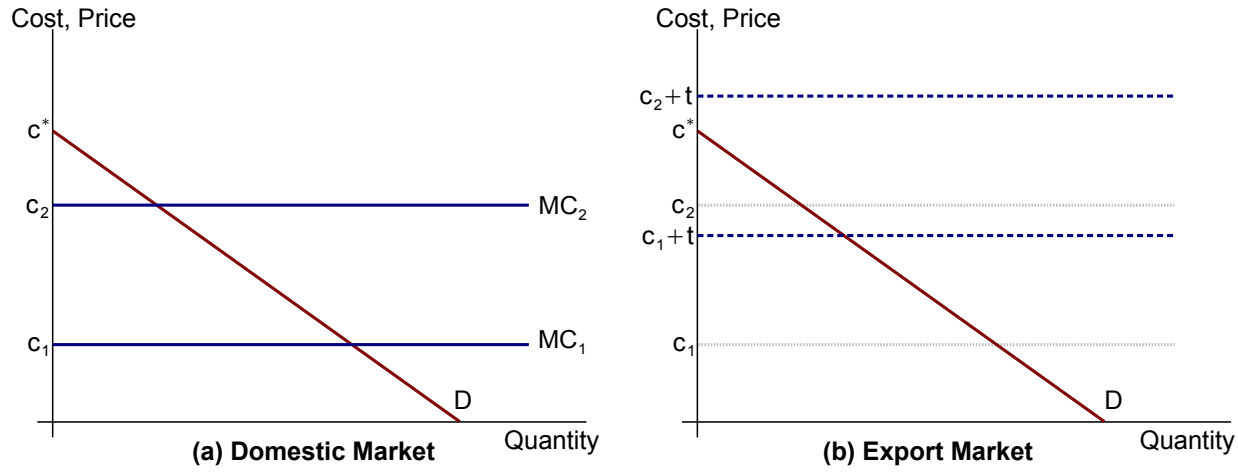

Figure 6: Trade Costs

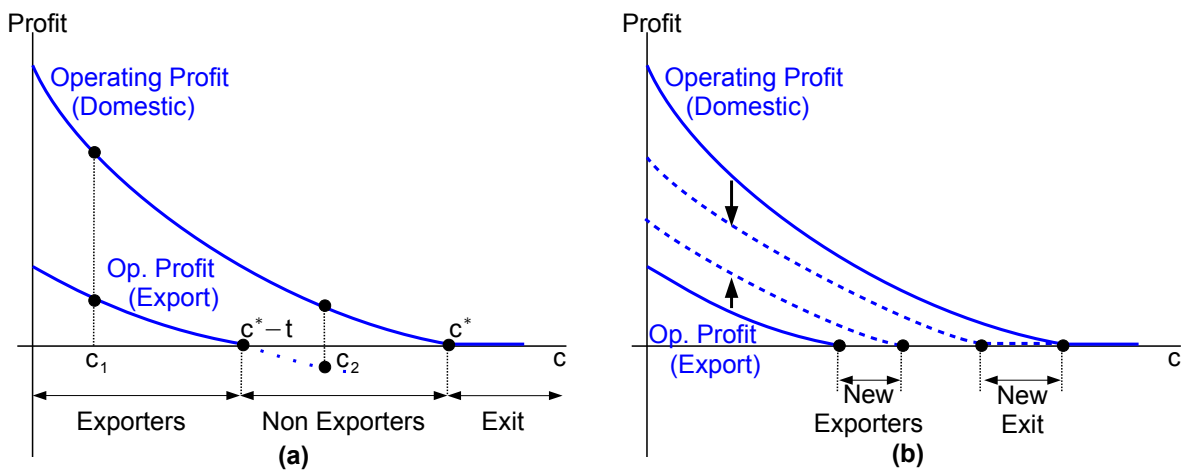

Figure 7: Export Decision and Trade Liberalization

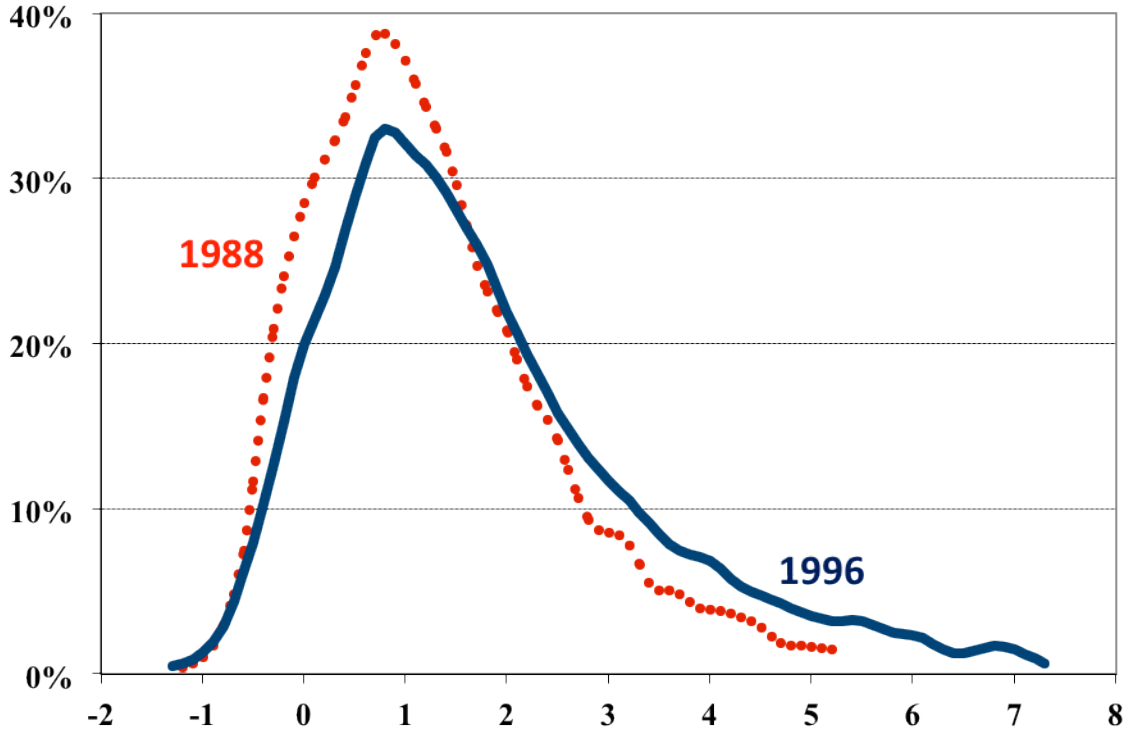

Figure 8 – Panel (a) Labor Productivity Distribution of *All* Canadian Manufacturing Plants 1988 and 1996 (employment weighted)

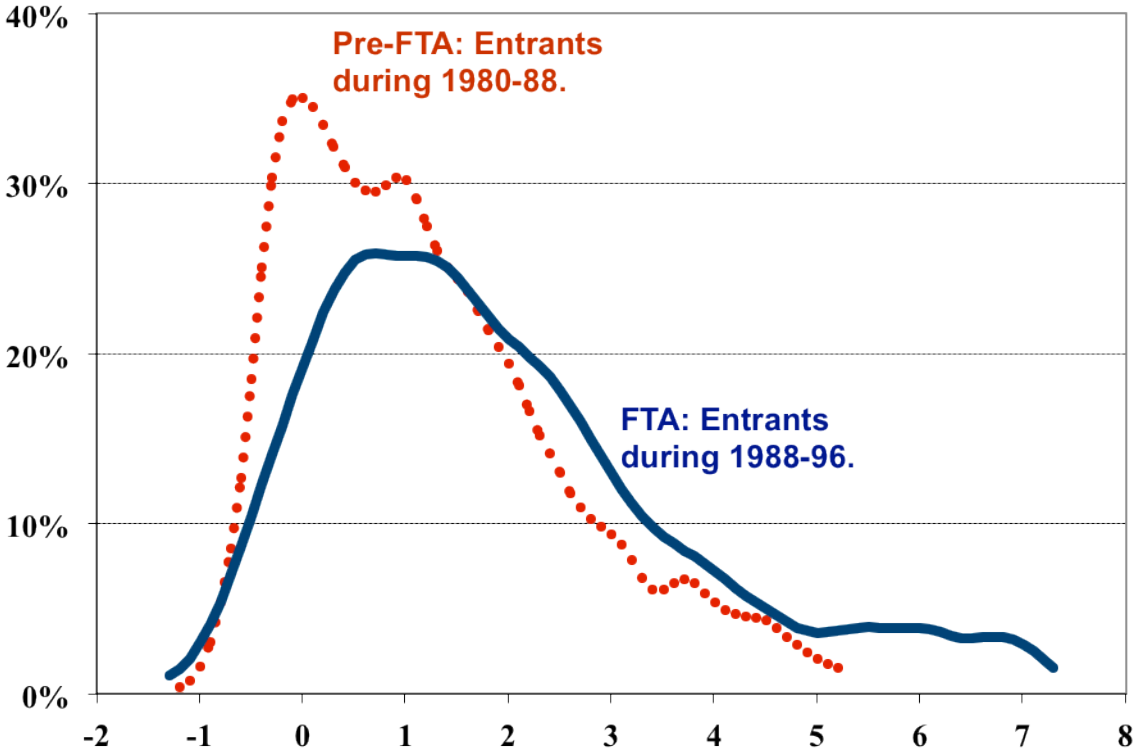

Figure 8 – Panel (b) Labor Productivity Distribution of *Entering* Canadian Manufacturing Plants 1980-1988 and 1988-1996 (employment weighted)

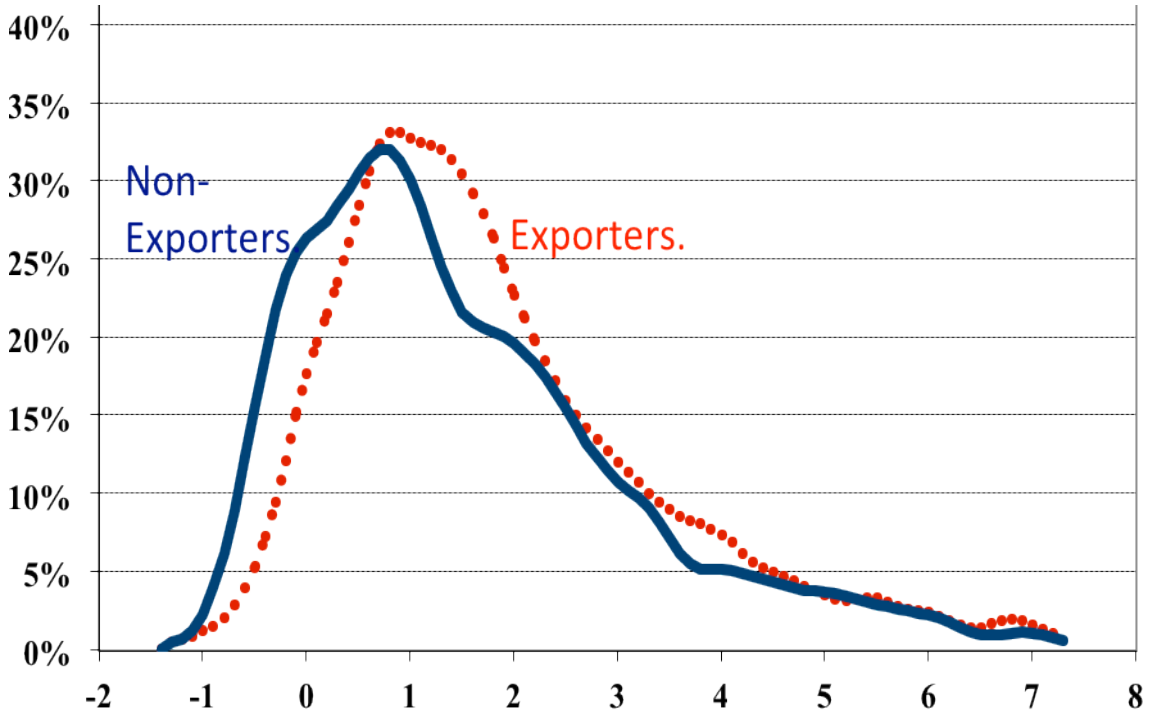

Figure 8 – Panel (c) Labor Productivity Distribution of Exporters and Non-Exporters, 1996 (employment weighted)

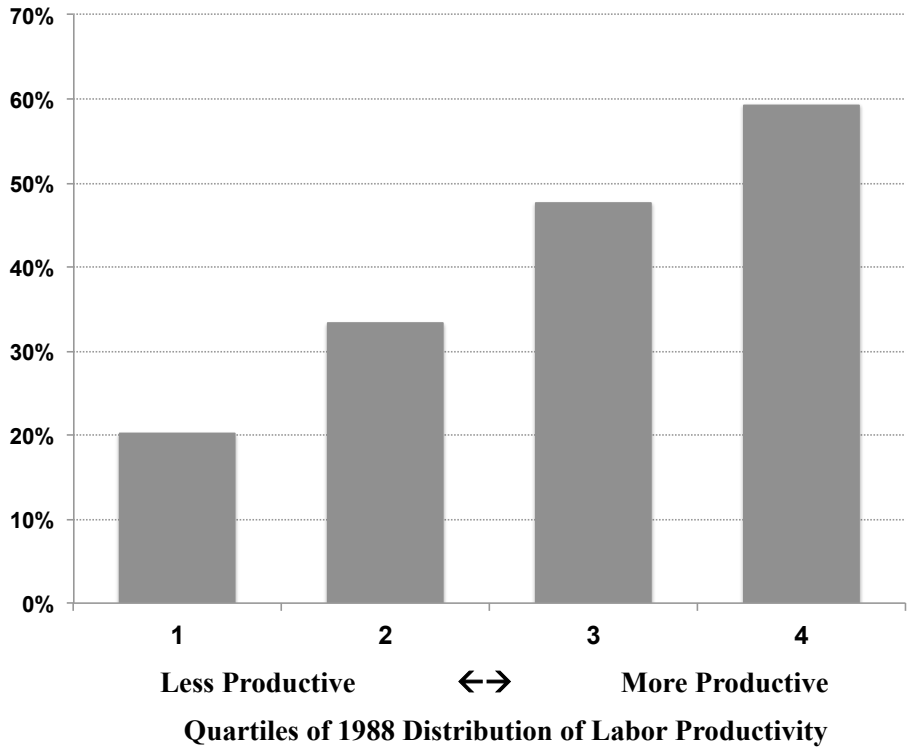

Figure 9: Share of Plants that Start to Export, by Initial Labor Productivity

Tables

	Raw Adoption and Innovation Rates			LATE
	New Exporters	Non-exporters	Difference	Difference
	(1)	(2)	(3)	(4)
Manufacturing Information Systems	16%	6%	10%	7%
Inspection and Communications	18%	10%	8%	8%
Any Product or Process Innovation	30%	20%	10%	8%
Any Product Innovation	26%	14%	12%	7%

Table 1: Innovation Response to FTA by New Exporters

Selection/Reallocation (Between Plants)	
Growth of exporters (most-productive plants)	4.1%
Contraction and exit of least-productive plants	4.3%
Within-Plant Growth	
New exporters invest in raising productivity	3.5%
Existing exporters invest in raising productivity	1.4%
Improved access to U.S. intermediate inputs	0.5%
Total	
Total	13.8%

Table 2: Overall Effect of FTA on Canadian Manufacturing Productivity