Alternative Sources of the Gains from International Trade:

Variety, Creative Destruction, and Markups

Robert C. Feenstra
February 21, 2018

Abstract

The modern theory of international trade identifies several new sources of the gains from international trade, in addition to the gains from traditional comparative advantage. These are the gains from importing new product varieties; the gains from “creative destruction” as the relatively most productive firms expand their output by exporting, while the less productive firms exit; and the gains from competition between firms in different countries, which can lead to reduced markups. Estimates of these various gains are provided for the United States and other countries.

The author thanks Eric Bartelsman and Gordon Hanson for helpful comments.

Robert C. Feenstra is the C. Bryan Cameron Distinguished Chair in International Economics, University of California-Davis, Davis, California, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is rcfeenstra@ucdavis.edu.
The modern theory of international trade allows for several sources of the gains from trade in addition to that of traditional comparative advantage. We discuss these sources and provide estimates of the gains for the United States and other countries. It turns out that the formula used to measure these new gains can be used to measure the gains from traditional comparative advantage, too, as we shall explain.

The first alternative source of gains from trade is not that new, and are the gains from increased variety of products. These gains were recognized by David Ricardo (1817, ch. 7) when he wrote: “Foreign trade, then, . . . [is] highly beneficial to a country, as it increases the amount and variety of the objects on which revenue may be expended.” In Ricardo’s time, these new varieties included cacao from Africa, spices from Southeast Asia, sugar from the Caribbean, and tea from India, all shipped to Europe and America. The production of these goods relies heavily on climate and soil, so they might be thought of as reflecting fundamental comparative advantage. The modern theory of international trade, however, allows countries to trade product varieties that do not have such fundamental differences, as with different types of cheese from France and Holland or different types of cars from Germany and Japan. Indeed, the modern theory often assumes that all countries and industries are producing differentiated varieties, which means that firms are operating under monopolistic competition rather than perfect competition. That is, firms retain some limited monopoly power in their unique product varieties so that prices are above marginal cost, but there is free entry of firms in the long run, so that markups just cover the fixed costs of entry and industry profits are zero.¹

¹ Research in international trade is nearly always done in a long run, general equilibrium setting, since it is felt short run or partial equilibrium settings do not lend themselves to theories of the pattern of trade. Likewise, the gains from trade are usually evaluated in a long run, general equilibrium model, where allowing free entry in all industries eliminates the need to keep track of the distribution of industry profits throughout the economy.
A second additional source of gains from trade also comes from the monopolistic competition model, but in contrast with the first-generation models that had homogeneous firms (Krugman, 1979, 1980, 1981; Helpman and Krugman, 1985), the second-generation model of monopolistic competition allows firms to be heterogeneous in their productivity levels (Melitz, 2003; Chaney, 2008). In this setting, international trade allows the most productive firms to expand their sales through exports, while the least productive firms are forced to exit because of competition with imports. This process that Joseph Schumpeter (1942, pt. 2, ch. 7) had in mind when he wrote: “The opening up of new markets, foreign or domestic, … revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism.” Through this process, average productivity in the industry rises due to increased sales of the most productive firms. These gains are analogous to those from traditional comparative advantage in a Ricardian model, when the relatively most productive industries expand their share of output through exporting. Indeed, as we shall see, the Ricardian model with stochastic technologies due to Eaton and Kortum (2002) has a very similar formula for the gains from trade as in the heterogeneous firm model.

A third alternative source of the gains from trade occurs when competition between firms in different countries leads them to reduce the markups that they charge. If there is only a single monopolistic domestic firm, then the reduction in markups leads to consumer gains and a reduction in the monopoly distortion, but the consumer gains would be substantially offset by the firm losses. In a situation of monopolistic competition, however, the entire reduction in consumer prices potentially becomes a social gain. That is because the free entry of firms under monopolistic competition drives industry profits to zero, and so the consumer
gain from a reduction in markups is not offset by any fall in profits (which remain at zero). There is still a potential offset to the consumer gain, however, because with some firms exiting there will be reduced product variety, and the net effect on welfare must take all these effects into account.

To illustrate the potentially ambiguous effect of competition between markets on social welfare, consider the comment from the popular food author Michael Pollan, who has said: “America ships tons of sugar cookies to Denmark and Denmark ships tons of sugar cookies to America. Wouldn't it be more efficient just to swap recipes?” Mr. Pollan is referring to the social cost of shipping goods between markets, which he views as a waste, but he is not taking into account the beneficial effect of competition between American and Danish firms. If having the cookies sold between the two countries leads to lower prices due to reduced markups, then there can be social gains despite the shipping costs. Whether there are social gains or not will depend on whether there is free entry into the sugar cookie industry, as we shall discuss.

The gains from trade are often computed as compared to autarky (i.e. no trade at all) and we will report some of these estimates. But it is important to also consider the gains from free trade as compared to actual restrictions on trade such as tariffs and quotas. In other words, we are interested in the cost of such trade restrictions as they are used in practice. These restrictions are often more complex than the simplified version used in modern models, and therefore have some unexpected costs that we shall describe.

**Product Variety in Trade**

Measuring gains from new varieties is difficult because there is no observed price for a product before it is available. The solution given many years ago by Hicks (1940) is that the
relevant price of a product before it is available is the reservation price for consumers – namely, a price so high that demand is zero – can be used to measure the consumer gains from the appearance of that new good. This idea of Hicks has been applied to new products by Hausman (1997, 1999). But when we try to apply this idea to the appearance of new products varieties from many countries due to international trade, we run into a problem: if each exporting country is providing a different variety, then we potentially have hundreds or thousands of new product varieties through trade, and it is impractical to estimate the reservation price for each.

A common way to address this problem is in this literature to adopt a constant-elasticity-of-substitution utility function. Provided that there are many goods so that the share of income spent on any good is small, then the elasticity of demand is approximately equal to the elasticity of substitution, which we denote by $\sigma > 1$. The constant-elasticity demand curve approaches the vertical axis asymptotically as the price goes to infinity, but it does not touch this axis. So the reservation price is infinity, but the area under the demand curve measuring the consumer surplus gain from a new product is bounded and well-behaved: the consumer gains from the new variety shrinks as the elasticity of substitution is higher, indicating that the new good is a closer substitute for an existing good.

With the assumption of a constant-elasticity-of-substitution (CES) utility function, Feenstra (1994) shows how new varieties affect an exact cost-of-living index (that is, the cost of purchasing goods that provide a constant level of utility). After making some assumptions—for example, that profits are zero due to free entry so that consumer gains from trade are equal the social gains—one can invert the change in the cost-of-living index to obtain an expression for the welfare gain from the new products.\(^2\) Suppose that we compare two different equilibria of the

---

\(^2\) Some other relevant assumptions are that labor is the only factor of production and that marginal labor costs are fixed. Normalizing the wage at unity and with constant markups due to the constant elasticity of demand, it follows
home economy with different amounts of imports. Let the variable \( \lambda \) denote the *share of home spending on its domestic products* in the first equilibrium, and let \( \lambda' \) denote this share in the second equilibrium. If we start in autarky where imports are zero and then move to free trade, it follows that \( \lambda = 1 \) initially and \( \lambda' < 1 \) under free trade. Then the social gains from trade can be expressed as:

\[
\text{Gains from trade} = \left( \frac{\lambda'}{\lambda} \right)^{-1/(\sigma-1)}.
\]

For example, suppose that \( \sigma = 3 \) so that \(-1/(\sigma - 1) = -1/2\), and that 90 percent of expenditure with trade goes to domestic goods, with 10 percent going to imports. Then, the formula equals \((0.9)^{-1/2} \approx 1.05\), which suggests that the availability of the new imported varieties, with the specified elasticity and spending level, raises welfare by about 5 percent.

Broda and Weinstein (2006) use this approach, extended to many industries, to estimate the gains from trade for the United States due to the expansion of import varieties. As in Armington (1969), varieties of a good from different countries (like cars from the United States, Germany, and Japan) are treated as imperfect substitutes, and the elasticity of substitution between countries is estimated using the techniques of Feenstra (1994). So importing a “new variety” really means that the United States starts importing a good from a country that did not export that good to the U.S. before. Broda and Weinstein find gains due to import variety that grow by 1.2 percentage points per year over 1972-2001, to a total of 28 percent of import expenditure in 2001, or 2.6 percent of GDP. While this number may seem modest, remember that prices are also fixed. Feenstra (1994) shows that the cost of living index for the home consumer between these two equilibria is \((\lambda'/\lambda)^{1/(\sigma-1)}\). The formula in the text for the social gains is the inverse of this expression.
that it is an estimate of the *incremental* gains from growing import variety from 1972-2001, not the total gains from trade as compared to autarky.

Gains from trade as compared to autarky for many countries are computed by Ossa (2015), using a similar approach. For the United States, he finds gains from trade equal to 13.5 percent of GDP, which is at the low end as compared to other countries. The median gain from trade in varieties across all countries is 55.9 percent of GDP. A number of other small European countries like the Netherlands, Norway, and Sweden are close to those median gains. The country with the highest gains is Belgium at 505.2 percent of GDP. Larger economies tend to have lower gains, because their share of spending on domestic firms is higher. Thus, gains from trade in varieties as compared to autarky are 30.8 percent of GDP for China, 35.3 percent for France, and 21.4 percent for Japan. Ossa also takes into account non-traded goods (which lowers the gains from trade) and inter-industry flows of intermediate inputs (which raises the gains).

A histogram of Ossa’s estimates of the gains across all countries is provided in Figure 1, where the horizontal axis measures the gains from trade (in percent), and the vertical axis is the frequency (or density) of countries with those gains. We distinguish several different cases: "naïve" gains, which use an *average elasticity* for all industries, rather than allowing it to vary across industries; "unadjusted" gains, which do not take into account non-traded goods or intermediate inputs; and "adjusted" gains, which take all these factors into account. The naïve gains are much smaller because they do not allow for low estimates of the elasticity in some industries, which would increase the gains from trade. The unadjusted gains are less than the adjusted gains for most countries, because the extra gains from intermediate input linkages usually exceed the reduced gains from having non-traded goods. The importance of such input-output linkages is stressed by Chaney and Ossa (2012) and Melitz and Redding (2014).
The median gains across all 50 countries considered by Ossa (2015) are 16.5 percent of GDP for the naïve gains, 48.6 percent for the unadjusted gain, and 55.9 percent for the adjusted gains (as noted above). These estimates may seem high, for two reasons. First, since individuals often consume just one variety of a differentiated product, we can ask whether the increase in variety due to international trade really translates into the gains computed from a CES utility function. This question is answered in the affirmative by Anderson, De Palma and Thisse (1992). They show that even when each individual consumes their preferred variety, social welfare can quite possibly be measured by a CES utility function.3

A different concern with the gains from imported variety are that they might be offset by reduced domestic variety as home firms leave the market due to import competition. Such exit of home firms does not occur in the simplest monopolistic competition model due to Krugman (1980) which assumes CES preferences and homogeneous firms, i.e. all firms in each country have the same marginal and fixed costs. If there are zero transportation costs, then it turns out that home firms faced with import competition can still sell enough abroad that they do not exit as trade is opened. But that result is highly stylized, and in reality we expect to observe the exit of some home firms, as we will find in a model with heterogeneous firms discussed in the next section. The heterogeneous firm model will allow for additional gains from trade due to creative destruction, as the least efficient firms exit. Even though that exit results in less product variety, the gains from trade may very well be larger than measured by Ossa (2015) using the formula for product variety described in this section.

---

3 The assumptions needed to obtain a CES utility function is that individuals have utility that includes a “random” component reflecting their own tastes for each variety, and that this random component is distributed across varieties in the same way as for the logit demand system. In order to obtain the CES rather than the logit demand system, individuals have to be able to consume multiple units of their preferred variety rather than just a single unit.
Productivity Differences and Gains from Trade

International trade can also lead to gains through the creative destruction process under which firms of higher productivity expand their output through exporting, while firms of lower productivity exit the market through bankruptcy. This source of gains from trade is emphasized in the model of monopolistic competition with heterogeneous firms due to Melitz (2003; also exposited in this journal in Melitz and Trefler 2012).

A good example of an industry where firms differ in their productivities and where the most productive firms benefit from exports is the production of solar (i.e. photovoltaic) cells. There are many such firms in the United States and worldwide, with rapidly growing capacity in China. Indeed, import competition from China led two U.S. manufacturers, Suniva and SolarWorld, to file a petition with the U.S. international Trade Commission in 2017 asking for the application of import tariffs. A tariff of 30 percent on solar cells and modules was approved by President Trump in January 2018 (which will fall by 5 percent per year over the next four years). The tariff was opposed by U.S. industry groups such as the Solar Energy Industries Association, which represents companies that install solar panels. Interestingly, the import tariff will not protect another U.S. firm, First Solar, which both produces cells with a more advanced technology (to which the tariffs do not apply) and is also a leading U.S. installer of solar systems worldwide. First Solar’s stock price fell on the announcement of the tariff. So it appears that the tariffs applied by the United States will enable less-efficient firms, like Suniva and SolarWorld, to survive while not benefitting the most efficient, exporting firms.

There are many other industries where plants differ in their productivities, as has been well documented in the industrial organization literature. Syverson (2004a,b; 2011) finds that the productivity distribution across plants is related to the extent of product substitutability:
when substitution between products is greater, then dispersion is smaller, because the lower-productivity plants find it harder to survive. In the macro literature, too, imperfect substitution across products is key when examining the potential productivity gains due to creative destruction. For example, in Garcia-Macia, Hsieh and Klenow (2016) and Aghion et al. (2017), an approach similar to the one described here for looking at gains from variety is used to measure productivity gains when one plant is replaced by another.

With international trade, the U.S. tariff on solar cells shows us that barriers to trade between countries allow for a dispersion of productivities that is higher than would otherwise occur. Such trade barriers can take many forms: tariffs, shipping costs, difficulties of communication or making contracts between countries, and others. In the initial formulation of the Melitz model, the trade costs have been modeled as “iceberg” costs of trade—that is, some fraction of the good “melts” in route to the destination country. While Samuelson (1956) proposed this way of modeling transport costs as a theoretical simplification, it has been used so often since then, including in the Melitz model, that it is sometimes not realized how sensitive certain results are to that assumption. We will continue with the simplification of “iceberg” costs of trade throughout this section, but then broaden the discussion to incorporate actual tariffs (as in the solar cell example) later in the paper.

When the “iceberg” costs of trade are reduced, the most productive firms expand their output by exporting more and the least productive firms exit. As a result, average industry productivity grows and consumer prices fall. This productivity-enhancing effect of creative destruction has been demonstrated for Canada by Trefler (2004) under the Canada-U.S. Free Trade Agreement, and for a broader sample of countries and free-trade agreements by Badinger (2007a, 2008; for evidence on Chile and Mexico, see also Tybout et al. 1991,
1995). As an example of this literature, Trefler (2004, p. 870) finds that:

The Canada-U. S. Free Trade Agreement provides a unique window onto the effects of a reciprocal trade agreement on an industrialized economy (Canada). For industries that experienced the deepest Canadian tariff cuts, the contraction of low-productivity plants reduced employment by 12 percent while raising industry level labor productivity by 15 percent. For industries that experienced the largest U. S. tariff cuts, plant-level labor productivity soared by 14 percent. These results highlight the conflict between those who bore the short-run adjustment costs (displaced workers and struggling plants) and those who are garnering the long-run gains (consumers and efficient plants).

This comment alerts us that the heterogeneous firm model has within it the potential for losses by less-productive firms and their workers. In the Melitz (2003) model, such losses are assumed away with a single type of labor that is perfectly mobile between industries. An alternative model due to Yeaple (2005) has workers varying by ability, with heterogeneous technologies available in one sector. Higher-ability workers match with the more advanced technology, workers with mid-level abilities use the inferior technology, and the lowest-ability workers are in a traditional sector, which is not traded in equilibrium. In this setting, opening trade between countries generates gains for the highest-ability workers without any losses to the lowest-ability workers, but the mid-level workers using the inferior technology suffer a loss in wages. As noted by Helpman (2018), this pattern of wage changes accords with the “hollowing out” of the U.S. labor market, whereby workers at the upper and lower-ends of the skills distribution have seen real gains, but not workers in the middle of the distribution (Autor, 2014). Helpman (2018) comprehensively discusses the potential losses to workers in recent models of international trade with matching between heterogeneous workers and firms.
Our goal here is to develop a general formula for the long-run gains due to creative destruction, which in turn requires thinking about the distribution of productivity across firms and how it would be altered by trade. To that end, it is convenient to simplify the Melitz model by adopting a Pareto distribution of productivities across firms, as proposed by Chaney (2008) in what is called the Melitz-Chaney model. In this setting, one parameter governs the variance of productivities across firms, which we will denote by $\theta$. If $\theta = 1.16$, for example, then 20 percent of the firms account for 80 percent of the output. The higher is $\theta$, the less spread-out are the firms in their productivities, and as $\theta$ approaches infinity then we are back in the case of homogeneous firms, all with the same productivity.

Once we know the parameter $\theta$, then the only other information needed to compute the gains from trade in the one-sector Melitz-Chaney model is the share of home spending on its domestic products, which we again denote by the variable $\lambda$ in an initial equilibrium. Let $\lambda'$ denote this share in a second equilibrium. As discussed earlier, if we start in autarky and then move to trade, then $\lambda = 1$ initially and $\lambda' < 1$. With these assumptions, Arkolakis, Costinot, and Rodríguez-Clare (2012) show that the home gains from trade between these two equilibria are:

$$Gains\ from\ trade = \left( \frac{\lambda'}{\lambda} \right)^{-1/\theta}.$$  

For example, compare the United States in autarky (so $\lambda = 1$) with the imports of 15 percent of GDP, so that the share of home spending on its domestic products equals 0.85. Suppose we choose the Pareto parameter $\theta = 4$. Then the U.S. gains as compared to autarky equal $(0.85)^{-1/4} \approx 1.04$, which represents gains of 4 percent of GDP. This example gives gains that are unrealistically small, however, because the above formula applies to a one-sector economy, with trade in final goods only. When we take into account many industries, some
of which have lower values of $\theta$ (so the firms with higher productivities are more spread out), and also include realistic input-output flows between industries, then the gains from trade become much larger, as we illustrated earlier in Figure 1.

Notice that the formula above is much the same as the formula in the previous section, except that they differ by the exponent used in each case. It is fair to ask what has happened to the elasticity of substitution $\sigma$, which appeared in the formula of the previous section, and measures the preference for different product varieties. This question is answered by Feenstra (2010; 2015, chapter 6), who argues that the gains from new import varieties in the Melitz-Chaney model cancel out with the losses due to reduced domestic varieties, as some domestic firms faced with import competition will exit the market. In other words, there are zero net gains from product variety in the one-sector Melitz-Chaney model; all of the gains come from rising average productivity in the industry.

In place of gains from trade due to product variety, the gains from trade in the one-sector Melitz-Chaney model are entirely due to creative destruction, i.e. lower-productivity firms exiting and high-productivity firms expanding their output through exporting. Despite having no gains from product variety, it is still quite possible that the Melitz-Chaney model (with heterogeneous firms) has gains that are greater than would occur due to importing new varieties (in the model with homogeneous firms). This result holds for two reasons. First, Melitz and Redding (2015) show that the heterogeneous firm model allows for the greatest expansion of output by the high-productivity firms when trade is opened, leading to more trade and therefore a lower value of $\lambda'$ than in the homogeneous firm model. With a lower value of $\lambda'$ in the formula from gains to trade just above, and a negative exponent, the gains from trade are higher.
Second, even if we simply take the domestic share \( \lambda' \), used in the formula just above and in the previous section, as given from the data, we can still find that the gains are higher in the heterogeneous firm model due to the differing exponents used in each case. Simonovska and Waugh (2014) present a method to estimate the parameters of the gains from trade formula: either the elasticity of substitution \( \sigma \) from the homogeneous firm model of the previous section, or the Pareto parameter \( \theta \) from a heterogeneous firm model. They show that when estimating each of these models on the same data, there is a systematic relationship between the estimated parameters: namely, that \( \hat{\theta} < (\hat{\sigma} - 1) \). So even when using the same value of \( \lambda' \), we will still get greater gains from the process of creative destruction.

Calculations of the gains from trade with homogenous firms, reflecting product variety, versus with heterogeneous firms, reflecting creative destruction, are provided by Balistreri, Hillberry and Rutherford (2011) and Costinot and Rodríguez-Clare (2014). Balistreri et al. find gains that are four times higher with heterogeneous firms than the gains from product variety. Smaller differences are found by Costinot and Rodríguez-Clare: moving from free trade to a hypothetical 40 percent worldwide tariff, they find average country losses of 7.0 percent with heterogeneous firms as compared to 5.3 percent with homogeneous firms; but in a model without intermediate inputs, the loss due to the tariff with heterogeneous firms is slightly smaller than with homogeneous firms.

---

4 As already explained from the work of Syverson (2004a,b, 2011), the dispersion of productivity across plants is limited by the elasticity of substitution. That restriction takes the form of the inequality \( \theta > (\sigma - 1) \), which is required for firms of differing productivities to occur in equilibrium. The estimates \( \hat{\theta} < (\hat{\sigma} - 1) \) do not contradict that theoretical condition, because the estimates are obtained from different models applied to the same data.

5 Balistreri et al. (2011) consider a simplified version of the gains from product variety, where the differences in product varieties across countries are exogenously given in what is called the “Armington model”. This model does not include any increasing returns to scale or monopolistic competition.

6 As noted by Costinot and Rodríguez-Clare (2014, p. 233) and discussed later in the paper, tariffs can lead to changes in product variety in a multi-sector Melitz-Chaney model. So the comparison of the heterogeneous and homogeneous firm models is not a pure comparison of creative destruction versus product variety gains.
We conclude that while there is a *presumption* that the gains from creative destruction exceeds the gains from new varieties, this result does not always hold in calibrated models. The same is true when we consider results from actual changes in trade barriers. For the Canada-U.S. Free Trade Agreement, the above quote from Trefler (2004) shows that productivity in Canadian industries was indeed boosted, and by a considerable amount. But Hsieh, Li, Ossa and Yang (2018) have found that exit of Canadian firms led to such a fall in product variety that it overpowered the entry of new U.S. exporters, such that overall consumer variety in Canada fell in some industries. That empirical result could not occur in the Melitz-Chaney model (where overall variety is constant due to the Pareto assumption). Canadian welfare still rose from the free trade agreement, but not by as much as in the formula we have described in this section.

Finally, it is important to point out that the formula for the gains from trade in this section *also applies* as a measure of the gains due to traditional comparative advantage, in the perfectly competitive model due to Eaton and Kortum (2002). These authors consider a continuum of products that have constant-elasticity-of-substitution preferences between them. The technologies for producing each product are independently drawn in each country and follow a Fréchet distribution, with the parameter $\theta$. This distribution has the convenient property that the minimum value in a sample of $N$ independent draws from a Fréchet distribution is also distributed as Fréchet, but with a reduced variance. There is a natural application of this property to international trade, since for each product, the buyer will be choosing the minimum-cost supplier over all $N$ potential supplying countries.

The Eaton-Kortum (2002) model has many of the same reduced-form properties as the Melitz-Chaney model, but it assumes perfect competition rather than monopolistic
competition. One of these properties is the formula for the gains from trade, which in the Eaton-Kortum model is identical to the equation presented earlier in this section, but with $\theta$ now reflecting the parameter of the Fréchet distribution. In the Eaton-Kortum model, these gains are due to comparative advantage—that is, from the ability to import from countries with the lowest relative costs. The close similarity of the formula used to measure the gains from trade in different models – as pointed out by Arkolakis, Costinot, and Rodríguez-Clare (2012) – is what makes this formula so important for international trade research.

**Pro-Competitive Effects of Trade**

The insight that import competition might create gains from trade by reducing the markups charged by firms was emphasized in the early monopolistic competition literature by Krugman (1979), but the broad idea that trade can be a source of pro-competitive gains pre-dates that literature. In Bhagwati’s (1965) well-known analysis of import tariffs versus quotas, he emphasized that import quotas would inhibit the competitive pressure from imports, because the quota fixes the import quantity regardless of the price charged by domestic firms. As a result, a home monopoly would raise its price more in the presence of a quota than for an “equivalent” tariff (that is, a tariff that results in the same quantity of imports). Harris (1985) and Krishna (1989) show a similar result, but in an oligopoly rather than a monopoly setting. The idea that trade barriers – especially quotas – can lead domestic firms to exercise their market power has been shown empirically for many countries: see Levinsohn (1993) for Turkey; Harrison (1994) for the Ivory coast; Krishna and Mitra (1998) and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for India; Kim (2000) for Korea; Bottasso and Sembenelli (2001) for Italy; Konings, Cayseele and Warzynski (2005) for Bulgaria and Romania; and Badinger (2007b) for other European countries.
To think about how trade can affect markups, we return to the discussion of trade in sugar cookies between the United States and Denmark (in the spirit of Michael Pollen’s comments earlier). Suppose we take the extreme assumption that there is no difference between American and Danish sugar cookies, so that there are no gains whatsoever from product variety. We also ignore differences in productivity between firms in the two countries. In this setting, the only potential source of gain from trade is the pro-competitive reduction in prices as the firms from each country enter the other market. Against these gains, we must count the resources spent in transporting the cookies between the countries.

Brander and Krugman (1983) call this a situation of intra-industry trade in homogeneous products. They argue that with a fixed number of firms, moving from autarky to free trade in this context has an ambiguous effect on global welfare. To grasp the intuition behind this result, suppose that there are two countries with a single firm in each that have identical production costs. Under autarky, the firms sell at the (same) monopoly price in their respective markets. Provided that shipping costs are less than the autarky markup, then with trade the firms will “cross-haul” into each other’s market, because for the first unit sold abroad the marginal revenue equals the price, which exceeds the marginal cost inclusive of shipping costs. That trade will reduce the price charged in each market, with a resulting consumer gain. But from a social welfare perspective, against that gain we must count the reduction in the profits for each firm due to competition from abroad, and also due to the shipping costs. That loss in profits can be greater or less than the consumer gain, so it follows that the social gain can be positive or negative.

With free entry of identical firms and a homogenous product in both countries, however, Brander and Krugman (1983) argue that global welfare necessarily improves due to free trade. In this case, industry profits are zero both before and after trade, so any reduction in consumer
prices becomes a social gain: the source of that gain is the exit of firms and the resulting savings of their fixed costs, which balances against the loss of resources in shipping. Welfare rises going from autarky to free trade if and only if the market price falls, which occurs whenever the firms cross-haul into the foreign market. This is an example of pro-competitive gain that can occur despite the fact that trade is intrinsically wasteful, due to trading a homogeneous product in the presence of shipping costs.

This social gain from trade depends, however, on the assumption that the transportation charges are priced at their true, social value. If gasoline and other fuels are priced too low, without fully reflecting their environmental costs, then the resources devoted to transportation could lead to a social loss. The social costs of international and intranational transportation have been analyzed empirically by Shapiro (2016), who compares the social cost of CO₂ emissions from international trade with the accompanying gains from trade. Shapiro measures the gains from product variety, so these are comparable to the gains we reported in Figure 1. For the world, he computes gains from trade of $5.5 trillion, or 10 percent of world GDP. By his computation, the added CO₂ emissions due to international air and sea shipments represent 5 percent of global CO₂ emissions. He adopts a social external cost of $29 per ton to these CO₂ emissions, in which case the global cost of the international transportation equals $34 billion, or 0.06 percent of world GDP. In other words, the gains from trade are more than 100 times greater than the social cost of these emissions, and the aggregate social gains from international trade vastly exceed the environmental externality from the international shipping.

Shapiro (2016) also analyses the impact of a global tax on emissions of carbon dioxide from air and sea shipping. He finds that implementing such a tax would raise global welfare, but in fact, would raise country welfare only for wealthy countries and decrease it
for poor countries. That result occurs because poor countries depend more on exporting products with low value/weight ratios, such as unprocessed resources. Thus, despite the fact the poorer countries are harmed disproportionately from global climate change – because they depend on agriculture and tend to be located in the equatorial region – they would also be harmed by policies to tax the carbon used in international transportation.

Let us return to the theoretical discussion of international trade and markups. The model of Brander and Krugman (1983) assumes Cournot competition between firms, i.e. each firm treats the other firms’ quantities as given. This paper is an example of the short-lived literature on “strategic trade policy” (e.g. Brander and Spencer, 1984, 1985). Trade researchers quickly moved away from oligopoly, however, because it was felt that the results of that literature were too sensitive to the form of conduct between firms (i.e. Bertrand versus Cournot competition) to be reliable for economic policy (Eaton and Grossman, 1986).

So trade economists have returned to the monopolistic competition framework, but without assuming a constant elasticity, in order to analyze markups that can change. One approach is to use the demand curve that arises from a translog expenditure function for the consumer. To build some intuition about this demand curve, consider a linear demand curve and a constant-elasticity demand curve, which are tangent to each other at the point of consumption. Now consider a family of demand curves all of which are convex (they lie above the linear demand curve) but have finite reservation prices (they lie below the constant-elasticity demand curve), and which all are tangent at this same consumption point. One of these intermediate demand curves will be based on the translog expenditure function. More specifically, with a single good it can be shown that the area under the translog demand curve is one half as large as under the constant-elasticity demand curve (Feenstra and
Shiells, 1997, footnote 27).

Translog preferences allow markups to vary systematically with the elasticity of demand. Specifically, markups fall when the market shares of domestic firms are reduced due to import competition. Industry profits remain at zero, however, due to the exit of some firms, which leads to savings in fixed costs. It follows that the entire reduction in consumer prices is a social gain for the economy. Feenstra and Weinstein (2017) adopt a translog expenditure function to measure the gains from new import varieties for the United States over 1992-2005, while also incorporating the losses from reduced domestic varieties along with changes in mark-ups. They find average growth of gains from trade of 0.85 percent of GDP during 1992-2005, with about half of that amount (0.44 percentage points) due to product variety, and the rest being the pro-competitive gains due to reduced markups. *Adding together* the gains from variety and the pro-competitive gains from trade for the United States, they find about the same *total* gains in the translog case as the pure variety gains under CES as measured by Broda and Weinstein (2006).\(^7\)

Feenstra and Weinstein (2017) do not attempt to measure the productivity gains coming from creative destruction, but that is undertaken by Feenstra (2018) who uses translog preferences and a *truncated* Pareto distribution for firm productivities, where the truncation means that the most productive firm has a productivity that is bounded above. These assumptions allow all three sources of gains from trade to operate: variety gains, pro-competitive gains, and productivity gains. In this case, the *total* gains from trade are larger than when only variety and pro-competitive gains are included: the total annual average gains for the United States are about

---

\(^7\) To be specific, Broda and Weinstein find U.S. gains from increased variety of 2.6 percent of GDP, which are the cumulative gains from new import varieties in the final year of their 1972-2001 period. If we instead consider *average annual gains* relative to GDP, they are 0.8 percent during 1990-2001, which is close to the 0.85 percent average annual gains over 1992-2005 obtained by Feenstra and Weinstein.
1.1 percent of GDP, with roughly one-quarter from productivity gains (as an upper-bound), with the remaining three-quarters evenly divided between variety and pro-competitive gains.

Those total annual average gains of 1.1 percent of GDP for the United States are less than the gains that would be expected in theory from a model that assumed an untruncated Pareto distribution for firms’ productivities. But in that case, there is no upper bound to the highest productivity firms, and these extremely productive firms contribute the most to the gains from trade. Indeed, without any upper bound to firms’ productivities, it turns out that the pro-competitive (and product variety) gains no longer operate. This surprising result comes about because even with the markups of domestic firms falling as trade costs fall, the markups of foreign exporting firms increase, so the overall distribution of markups (and product variety) is not affected. So with the untruncated Pareto distribution, the entire gains from trade come from creative destruction. Essentially, the earlier formula showing social gains from the creative destruction process becomes an upper bound to the total gains from trade, which applies even when all three sources of gain operate (as with truncated Pareto and translog preferences).8

**Beyond Iceberg Trade Costs**

The discussion in the previous sections has mainly considered “iceberg” costs of trade, but in reality, trade costs take many forms: tariffs; quotas; tariff-rate quotas that apply the tariff only to imports exceeding a certain quota level; the threat of tariffs as in dumping cases; export taxes and subsidies; and other policy instruments. These policies differ most obviously from iceberg trade costs in that iceberg costs are paid in terms of a firm’s own output, which hypothetically “melts” along the way to the destination country. An increase in

---

8 Feenstra (2018) obtains this upper bound theoretically for a wide class of homothetic preferences that includes the translog as a special case. A similar upper bound also occurs in Bertoletti, Etro and Simonovska (2016) who examine certain non-homothetic preferences, and in Arkolakis, Costinot, Donaldson, and Rodriguez-Clare (2017) for their estimated non-homothetic preferences.
iceberg costs leads to an increase in the cost inclusive of shipping and therefore in the price charged abroad. An increase in this price reduces the quantity demanded and lowers the profits earned by the exporting firm. But unlike a tariff, iceberg costs also generate demand for the firm’s own product, since those costs are literally used for shipping, and so they do not lower its profits as much as would an import tariff of the same magnitude. That distinction is enough to make a difference between iceberg trade costs and tariffs.\(^9\) In addition, tariff revenue might be redistributed back to consumers, rather than melting away in transit, which is another difference between iceberg costs and tariffs.

Thus, we conclude that it is important to keep track of the tariff revenue (or its equivalent) in the economy: who is paying it, who receives it, and so on. This idea is not new. It has received substantial prior attention in the analysis of import quotas, where instead of generating tariff revenue an import quota will generate rents as the difference between the selling price in the quota-controlled market and under free trade. In the “voluntary export restraints” used to restrict U.S. imports of Japanese automobiles in the 1980s, for example, the rents exceeded $1,000 per imported car or $2 billion annually in 1983-84 (Feenstra, 1988). These rents were earned by the Japanese exporting firms, leading to a substantial increase in the stock market value of these firms when the quota was announced (Ries, 1993). In addition, European automakers reacted to the voluntary export restraints by increasing their prices in the US market, as we would expect under oligopoly pricing, which cost US consumers an additional $3.4 billion (Dinopoulos and Kreinin, 1988). These welfare costs to the United States are far in excess of the deadweight losses associated with import

---

\(^9\) We are assuming that the tariff is charged on imports valued at their price inclusive of the markup over marginal costs. In the alternative case where the tariff is charged only on the variable cost of the import, as analyzed by Costinot and Rodriguez-Clare (2014), then it has much the same effect as iceberg costs; in that case, the two instruments differ only by the redistribution of tariff revenue.
quotas in automobiles, and are outside the scope of the formula for social welfare gains from trade presented earlier.

A similar lesson applies from more recent policy actions, such as the tariff applied to U.S. imports of tires from China by President Obama during 2009-2012. Because this tariff applied only to imports from China, other Asian countries (and also Mexico) exporting to the United States were able to earn rents due to the higher price of tires in the United States. These rents captured by foreign producers amounted to $0.8 billion annually (Hufbauer and Lawry 2012), which far exceeds the wages earned in the additional jobs in the United States. This situation also applies to the tariff on U.S. imports of solar cells, where some emerging economies that export are exempted from the tariff and can therefore collect rents from higher prices in the U.S. Interestingly, one commissioner of the U.S. International Trade Commission recommended instead that the U.S. adopt a quota on imports at the 2016 level (increasing each year thereafter), with quota licenses auctioned by the U.S., which would have allowed the rents to be collected as auction revenue.10 Despite the benefit of collecting the revenue, however, a quota of this type would have allowed U.S. firms to exercise their market power and raise their prices, much as occurred with the voluntary export restraints on U.S. imports of cars from Japan in the 1980s.

A different example of how certain trade policies can generate costs above and beyond those in the earlier formula for social welfare gains comes from the application of antidumping policy. Every year, dozens of antidumping cases are filed by firms in the United States and in other countries, alleging that foreign firms are selling at “less than fair market value,” which can lead to the imposition of antidumping tariffs. To avoid such tariffs, foreign firms may raise their

---

10 See the statements of the commissioners at: https://www.usitc.gov/press_room/documents/solar201_remedy_commissionerstatements.pdf.
prices so that they are less likely to be found guilty of selling at “less than fair market value.” That price increase leads to a welfare cost in the importing country, even when the antidumping tariff is not actually applied. Because foreign firms earn these higher prices, and there is no offsetting tariff revenue, these antidumping actions have a substantial welfare cost.

Staiger and Wolak (1994) find evidence of such price increases in the absence of antidumping duties being applied. Ruhl (2014) quantifies the magnitude of the resulting welfare loss for the United States in 1992. In a calibrated model, he finds that antidumping policy reduces U.S. consumption (and therefore welfare) by 3.2 percent. If we apply Ruhl’s estimate of a 3.2 percent loss in consumption to U.S. imports of $533 billion in 1992, then we obtain a loss of $17 billion. This total is four times greater than the welfare cost of US antidumping and countervailing duties from Gallaway, Blonigen, and Flynn (1999), estimated at $4 billion in 1993—but this estimate considers only the actual antidumping and countervailing duties in effect in 1993, whereas Ruhl allows firms to increase their price to lower the probability of tariffs being applied. Ruhl also allows for firm heterogeneity, which is important because antidumping duties are higher for the most productive (and hence lowest price) firms.

Turning from these specific examples of trade policies back to the theory, what is the impact of import tariffs is in the Melitz-Chaney model? To add a dose of realism to that framework, suppose that in addition to the differentiated-goods sector, we add a perfectly competitive sector that is non-traded. Assume that trade is balanced. Then applying an import tariff to the differentiated goods sector can be expected to have the same impact as applying an export tax to that sector: both will drive down levels of trade, and with the assumption of balanced trade, both will result in reduced imports and exports. This equivalence of import tariffs and export taxes in general equilibrium is called “Lerner symmetry.”
In this general framework, Caliendo, Feenstra, Romalis and Taylor (2017) find that starting from a zero-tariff equilibrium, an import tariff leads to reduced entry into the differentiated goods sector at home. That is, an import tariff or an export tax equivalently leads to the exit of home firms, which creates a social cost that is above and beyond the cost of reducing trade based on the earlier formula on gains from trade. Because of this distortionary effect that tariffs have on entry, it even turns out that the optimal tariff for countries can be negative instead of positive.\textsuperscript{11} That outcome can arise in two theoretical cases: i) countries with very strong international input-output linkages; ii) countries that are very remote and have little trade as a result. In the quantitative model, it turns out the fully one-quarter of the countries in the world have negative optimal tariffs, with most of these falling into the second category (for example, small island economies), but with some examples of strong international linkages (for example, France).

These results of Caliendo, Feenstra, Romalis and Taylor (2017) imply that ongoing efforts to liberalize trade under the World Trade Organization are important, because it is quite possible that countries have first-order gains from reducing tariffs even when tariffs are small (indeed, that is the case when the optimal tariff is negative). This is in contrast to a competitive model, where the worldwide gains from reducing tariffs that are small will also be very small. The idea that the gains from tariff reductions are not necessarily small, especially for highly-linked and for very remote countries, argues for renewed attention to multi-regional and multilateral negotiations to liberalize trade.

\textsuperscript{11} However, if the entry distortion is offset using some domestic policies, then the optimal tariff in the Melitz-Chaney model would remain positive, as shown by Costinot, Rodriguez-Clare and Werning (2016).
Conclusions

Recent theories of international trade have allowed for sources of the gains from trade that go well beyond conventional comparative advantage, including gains from increased variety, a shift toward firms with higher productivity, and lower mark-ups. In the context of the prominent models of these effects, the social gains can be presented in a straightforward formula. For a given sector, the implied gains depend on the share of home spending on its domestic goods, which together with a parameter reflecting the taste for variety or the Pareto productivity distribution of firms, are “sufficient statistics” for the gains from trade, to use the term popularized by Arkolakis, Costinot, and Rodríguez-Clare (2012). This formula tells us that we can expect welfare gains in proportion to the rise in trade shares (or more precisely, to the fall in home shares), regardless of the underlying model of trade.

The gains from trade that we have discussed are “overall” gains in the sense that they reflect increased purchasing power for the economy overall, without consideration of how those gains are distributed across individuals, some of whom can be expected to lose. This point was already mentioned in our discussion of creative destruction, where the mid-ability workers matched to firms with inferior technologies can lose due to trade. We conclude our paper with a further discussion of these distributional concerns (see also the papers by Fort, Pierce and Schott and by Rodrik in this volume).

Krugman (1981) was optimistic that trade in product varieties had the potential to offset certain losses from trade. He started with that the logic of the Stolper-Samuelson theorem, where under perfect competition the factor of production used intensively in export production will gain but the factor of production used intensively in import production must lose. Krugman argued that this result could be overturned under monopolistic competition, since in that case, the gains from product variety could potentially benefit all factors of
production. For example, think of workers benefiting from the low prices and product variety offered at Walmart, even if their nominal wages suffer due to import competition. Is there any evidence to support the idea that they might gain in real terms?

Fajgelbaum and Khandelwal (2016) begin to address this question by examining how consumers of differing incomes benefit from import variety. They find that poor consumers tend to gain the most from imported goods, because they concentrate their spending in sectors that have more trade. That conclusion holds especially for higher-income countries that export high income-elastic goods (luxuries) and import low-income elastic goods (necessities). As a result, the poor in high-income countries have the greatest consumption gains from trade as a share of income. Fajgelbaum and Khandelwal do not incorporate the differing wages earned by various individuals, however. That step is taken in the recent work by Borusyak and Jaravel (2018) for the United States, who distinguish the expenditures and earnings of different educational groups. They find that a reduction of all import and export trade barriers generates a modest increase in inequality between education groups, which is primary due to the earnings (i.e. wages) channel. Those with or without a college degree both gain on average from such a reduction in trade barriers, like in Krugman’s original claim, but the college-educated gain by more.

This optimistic outcome due to the product variety gains from trade may well apply to the gains from reduced markups, too, which can be expected to benefit many consumers. In contrast, creative destruction will most likely negatively impact some workers much more than others. We have already mentioned the sorting of workers across firms, whereby workers with mid-level skills are matched to firms with inferior technologies, and these workers and firms face the greatest import competition and experience losses. Indeed, we
have suggested that the U.S. tariff on solar cells is intended to protect such firms and workers, but at the cost of disrupting the industry engaged in the installation of solar panels. More work is needed to determine how the alternative sources of gains from trade discussed here interact with their distributional consequences, so that we know whether these gains are widely shared, and if there are policies that can help to achieve that goal.
References


Costinot, Arnaud and Andrés Rodríguez-Clare. 2018. Article in this volume


Fort, Teresa C., Justin R. Pierce and Peter K. Schott. 2018. article in this volume


Rodrik, Dani. 2018. Article in this volume.


Source: Gains from trade are from Ossa (2015), graphed with a bandwidth of 0.1.

Figure 1: Gains from Trade (percent of GDP)