Using Supply Chain Analysis To Examine the Costs of Non-Tariff Measures (NTMs) and the Benefits of Trade Facilitation

Michael J. Ferrantino*
U.S. International Trade Commission
January 2012

*The author is with the Office of Economics of the U.S. International Trade Commission. Office of Economics working papers are the result of the ongoing professional research of USITC Staff and are solely meant to represent the opinions and professional research of individual authors. These papers are not meant to represent in any way the views of the U.S. International Trade Commission or any of its individual Commissioners. Working papers are circulated to promote the active exchange of ideas between USITC Staff and recognized experts outside the USITC, and to promote professional development of Office staff by encouraging outside professional critique of staff research.

Address correspondence to:
Office of Economics
U.S. International Trade Commission
Washington, DC 20436 USA
ABSTRACT

It has become increasingly common to produce goods in a number of geographically dispersed stages linked by international trade. This tendency, known by names such as “production fragmentation,” “processing trade,” and “vertical specialization,” has important implications for the analysis of non-tariff measures (NTMs) and trade facilitation. First, different types of NTMs or trade facilitation issues are naturally associated with different stages in the movement of goods. Different price gaps can be assigned to these stages, making it possible to decompose the overall amount of distortion and to prioritize the policies with the largest potential efficiency gains. Second, NTMs may accumulate in long supply chains, implying that their trade-distorting effects are greater for goods produced in a fragmented manner than for goods with simple production processes. There is evidence that trade costs are more important for high technology goods or goods undergoing several stages of processing. Issues with product standards may be particularly important for goods with long supply chains. The link between NTMs and supply chains also has implications for economic development and for the relationship between liberalization in services and goods.

1. Why supply chain analysis of NTMs?

(a) The growing role of supply chains

In recent decades, it has become increasingly common to produce goods in a number of geographically dispersed stages linked by international trade. Such international supply chains have been...
described variously by economists as “production fragmentation” (Arndt and Kierzkowski, 2001), “processing trade” (Görg, 2000), “vertical specialization” (Hummels, Rapoport and Yi 1998), “slicing up the value chain” (Krugman, Cooper and Srinivasan 1995), or “the second unbundling” (Baldwin 2006). The implications of this global change in the organization of industry is that it takes many more export and import transactions to provide a single unit for final demand of complex goods like computers and automobiles than previously. While there are examples of production fragmentation going back to ancient times, the widespread adoption of this method of production and trade has a number of implications for how the world economy works today. These include reallocating the value added by trade among different countries depending on where they fit in the supply chain (Koopman et al. 2010) and, possibly, making international trade flows more sensitive to the business cycle, as demonstrated in the recent Great Recession and Great Trade Collapse of 2008-09 (Baldwin 2009).

(b) The inseparability of price gaps, and the desire to prioritize policy efforts

At the same time, there is an increasing interest among policymakers in addressing barriers to trade other than tariffs, known collectively as “non-tariff measures” or NTMs. As tariffs have declined steadily since the 1940s, both in seven GATT/WTO rounds and in numerous unilateral, bilateral, and plurilateral liberalizations, government interventions to restrict imports have increasingly taken non-tariff forms. These include, but are not limited to, quantitative restrictions, technical barriers to trade, sanitary and phytosanitary measures, price-based measures, and so on. Quantifying the effects of these policies on world trade is challenging; indeed, NTMs have often been held to represent “murky protectionism”

---

2 During the height of the Roman Empire, trade in marble and manufactures of marble linked various provinces in multi-stage production processes (Moore and Lewis 1999, pp. 255-260). The marble production chain was coordinated by the equivalent of today’s multinational companies. Major quarries, often owned by the Caesars, operated in Egypt, Numidia (modern Tunisia and Algeria) and Bithynia, in the northwest of modern Turkey. Blocks and columns of marble were mass-produced on a prefabricated and standardized basis. The production of marble caskets or sarcophagi could involve several stages, with hollowing out taking place in Asia Minor (western Turkey) and finishing in Athens, Alexandria, or Beirut. The resulting products were exported throughout the Empire, differentiated for local tastes and customs.

3 The term “non-tariff measures” is often taken to be synonymous with “non-tariff barriers” (NTBs), the latter term being more common in the earlier literature. For the purposes of economic inquiry, I adopt “non-tariff measures” as relatively value-neutral, while “non-tariff barriers” may convey the connotation of trade policies which violate some negotiated or agreed norms.

4 For a categorization of NTMs see UNCTAD (2010, pp. 121-142).
(Baldwin and Evenett 2009), since, unlike ad valorem tariffs, they are not immediately associated with numerical measures of their impact.

Attempts to assess the trade impacts of NTMs have led to the development of the “price gap” or “tariff equivalent” method, which seeks to estimate the level of ad valorem tariff that would have an equally trade-restricting effect to the NTM in question. If country A is imposing an NTM, and its import price is higher than the “world price”, this can be taken as evidence that the NTM is trade-restrictive. There are a number of issues involved in estimating the “world price” – should the CIF prices of other importers be used, or the FOB prices of exporters? Can transport costs be accounted for? More importantly, what about the effects of differences in quality on prices? These issues are more or less surmountable. Given sufficient data on export and import prices, tariffs, and transport margins,⁵ it is possible to do a reasonably good job of estimating price gaps.

An alternate method of assessing the impact of NTMs is to estimate “quantity gaps” – i.e. are actual trade flows in the presence of the NTM less than expected trade flows, as estimated by a statistical model of trade, such as a gravity model? Quantity gap estimates are essentially the dual of price gap estimates. While price gaps are often preferred on a number of grounds,⁶ quantity gaps are particularly useful when the NTM is absolutely prohibitive, so that no prices are observed, or when the product is highly differentiated, so that unit values are either not observed or not particularly informative.

However, there are problems in moving from estimates of price gaps or quantity gaps to recommendations to policymakers. It is notoriously the case that when there is a preference to restrict imports, multiple NTMs may be in place. Indeed, exporting firms, or governments negotiating on their behalf, may remove one NTM only to see new ones emerge, leading to what is often called the “whack-a-mole” problem.⁷ When there are multiple NTMs in place, it is natural for policymakers to want to know

---

⁵ Transport costs must be subtracted from the price gap in order to identify the residual potentially attributable to NTMs. Since the CIF and FOB prices usually used to calculate the price gap do not include tariffs, an additional adjustment for the tariff is unnecessary or superfluous. See Linkins and Arce (1994), Deardorff and Stern (1998) and Ferrantino (2006) for information on the price gap method.


⁷ “Whack-a-mole” is a children’s arcade game in which the player strikes rodents on the head with a mallet only to have them pop up again repeatedly. See Tilton (1998) for an example of multiple NTMs in practice. See also Ferrantino (2011).
which are more restrictive or more important. Suppose that an imported product is affected simultaneously by non-automatic licensing, a technical standard, and slow customs procedures. The total effect of such procedures is estimated to be represented by a tariff equivalent of 50 percent. Is it possible to decompose the tariff equivalent, so that we can say, e.g. that non-automatic licensing has a 25 percent ad valorem equivalent, the technical standard 15 percent, and customs procedures 10 percent? Such a decomposition would be very useful in prioritizing policy efforts, and targeting interventions to the most severe problems first.

In principle, price gaps cannot be so decomposed. Since there is only one distorted domestic price and one world price (after appropriate adjustments to each), there is only one price gap. No further information is available. If there are multiple policies, their individual and specific impact on the distortion in question is unknown. Indeed, it may be that one or more of the policies are binding constraints – there may be a key policy which, when removed, gets rid of most of the distortion, or it may be necessary to reform the whole set of policies in order for anything observable to happen in the market place. The same objection applies to quantity gaps in the presence of multiple NTMs – there is only one actual quantity observed in the market place, and one estimated quantity, and thus one non-decomposable quantity gap.

(c) The possibility that NTM effects may cumulate in supply chains

In 2006, I proposed that NTMs could be decomposed by the study of goods as they move through supply chains.8 The idea is to follow a typical exported good from its location of production (ex-farm or ex-factory) through multiple steps in the process of shipping and delivery. For example, goods once produced are moved to the export port; handled in the export port; moved internationally by water, air, or road; handled in the import port; cleared through customs, paying any applicable duties; moved in the import market; and subject to wholesaling and retailing. At each stage in the process the price of the good increases, as additional costs are imposed (Figure 1). Moreover, the costs associated with each move through the supply chain can now be separated into their constituent parts. Different policies and practices apply to each part of the supply chain. For example, market distortions in international shipping

---

specifically affect the difference between the FOB and CIF prices; import customs procedures affect the
difference between the CIF price and the landed duty-paid price, and restrictions on the size or hours of
retail operations in the importing country affect the difference between the wholesale and retail price.
Thus, it is possible at least in principle to have a common metric to compare the restrictiveness of
different types of NTMs.
Some costs, such as those associated with land transport to export or port procedures, may not represent NTMs as usually conceived, but may be amenable to trade facilitation interventions. Another advantage of a supply chain framework is that NTMs, which raise prices of traded goods, and trade facilitation efforts, which should lower prices, can be compared using a common metric. Indeed, this reflects the general point that NTMs and trade facilitation can be analyzed with similar tools.\textsuperscript{10} For example, inefficiencies in customs procedures are sometimes thought of as NTMs and sometimes as trade facilitation issues. Since this framework reveals the comparability of NTMs and trade facilitation, it does not matter what one considers them – improving customs procedures reduces a distortion.

The limited available evidence suggests that total markups along the supply chain can be substantial. In one widely-cited exercise, Anderson and Wincoop (2004) estimate that among developed countries, the typical cost increase from the factory in an exporting country to the retailer in the importing country amounts to 170%, consisting of 21% transportation costs, 44% border related trade barriers, and 55% retail and wholesale margins ($1.21 \times 1.44 \times 1.55 = 2.7; 2.7 - 1 = 1.7$, for a markup of 170 percent). The

\textsuperscript{9} For an algebraic representation of Figure 1, see Ferrantino (2006, Annex 1).
\textsuperscript{10} Dee and Ferrantino (2005), Ferrantino (2006), Shepherd (2010).
44% markup may include tariffs, NTMs and “natural” barriers, such as different languages, information costs, and the transaction costs associated with using different currencies. An even higher estimate is given by Feenstra (1998), citing Tempest (1996), which suggests that the mark-up on Barbie dolls produced in China and sold in the United States is on the order of 900 percent.

While some of the costs associated with international trade are unavoidable, others are associated with policy-induced distortions or technological inefficiencies. Thus, it should be possible in principle to compare actual costs at each step of the supply chain with best-practice costs, consisting of necessary marginal costs of processing the goods in the absence of rents, and with efficient use of technology (Figure 1). The differences at each step can be attributed to step-specific NTMs, or to unresolved trade facilitation issues. Some of these will have rents associated with them which accrue to specific actors, while others represent pure inefficiency. With a supply-chain decomposition, it would be in principle possible to identify where the greatest rents and inefficiencies are, and to identify policy priorities which are most likely to expand trade and benefit both producers and consumers of traded goods.

In principle, the price increase at each step should include not only the monetary costs of moving along the supply chain, but the costs associated with the time of waiting. Since Hummels (2001), it is well-established that the delays experienced by traders are perceived as costly, as evidenced by the willingness of traders to pay more for faster air freight as opposed to slower water freight. Thus, they can be expressed as a tariff equivalent. Furthermore, these delays vary from product to product (Hummels et al. 2007). Given measures of the delays associated with exporting and importing, such as those in the World Bank’s “Doing Business” indicators for trading across borders, it is possible to simulate the effects of reducing those delays (Minor and Tsighas 2008). Moreover, time costs vary widely along inefficient transport corridors, often including an unpredictable “long tail” of very slow transit times (Arvis, Raballand and Marteau 2010; Christ and Ferrantino 2011). Thus, the uncertainty of time costs ought to be considered along with the mean transit time.

The discussion that follows will consider the types of costs, both monetary and time costs, associated with traveling through each step of the supply chain, as well as the types of costs and delays attributable to policy. These policies may include NTMs as traditionally conceived, inadequacies in trade
facilitation, or other types of policies insofar as they add to the costs and time associated with an international transaction. Such a framework could be used in a case study following a particular good through various stages of the supply chain. The result of such a study would be to identify those costs of moving goods which are technologically necessary, those which are due to technical inefficiency (such as poor roads), and those which are imposed by policy. The sum of the policy-induced costs along each step of the supply chain would amount to the NTM price gap as traditionally conceived, decomposed by the type of policy involved. Both the policy-induced costs and the technical inefficiencies would potentially be amenable to policy interventions that would reduce the overall price and time gap between exporters and importers and expand trade.

The quantitative illustrations of the effect of NTMs on supply chains in this paper frequently draw on metrics developed in the study of trade facilitation and transport costs. This is for both a general and particular reason. First, the analysis of NTMs and trade facilitation is largely analogous, since one examines factors that make trade more difficult and the other considers policies that make trade easier (Dee and Ferrantino 2005). Second, the transport cost and trade facilitation literature has developed several metrics which allow the amount of trade costs in many countries to be compared, such as the cif/fob margin and the Doing Business Trading Across Borders data. Similar broad comparisons of trade costs associated with SPS, TBT, non-automatic licensing or other NTM policies are not at present available due to challenges in quantification. The examples presented here can be extended to the analysis of specific situations in which measures of the impact of specific NTMs are available.

2. The stages of a supply chain, and policies and technologies corresponding to each

(a) A linear supply chain

Let us consider first the case of a good which is produced in a single location in the exporting country, such as an agricultural good or a carpet, and simply moved from place to place until it reaches the consumer in the importing country, following the steps illustrated in Figure 1. \(^{11}\) One of the insights

\(^{11}\) The number of points in the supply chain at which prices and costs can be evaluated is not limited to those listed in Figure 1. Subramanian, Paludetto and Yee (2007, pp. 61-62) identify 13 points between “ex works”
derivable from the supply chain approach is that there may be market power exerted at various stages along the supply chain, with each transfer point representing a bilateral monopoly or bilateral oligopoly. The difference between the retail price the consumer in the importing country pays and the ex-farm/ex-factory price, minus all necessary average costs for logistics, equals the total amount of rents extractable along the supply chain. Exertion of more market power at any point along the supply chain squeezes rents at other points along the supply chain, as well as increasing the overall “Barbie-doll” markup between the original producer and the ultimate consumer.

(i) *Ex-farm/Ex-factory*

The good leaves the farm or factory at an ex-farm or ex-factory price, which may represent the average cost of production or may itself include a markup if the producer has market power. The costs of production themselves may be influenced by NTMs or other aspects of trade policy. For example, compliance with TBT or SPS measures in the importing country may involve changes in the production process that are costly. The additional costs required to meet a product standard in an importing market may be considered to be part of the NTM price gap embodied in the costs of production. There may also be separate production processes for an export market, to which an NTM applies, and the domestic market, to which it does not. One example of this is a farmer who grows GMO grains in one field, for Western Hemisphere and Asian markets, and non-GMO grains in another field, for the EU market. A similar problem exists for rules of origin; for example, an apparel producer in Mauritius may maintain separate production lines, one with cloth satisfying the AGOA rules of origin for export to the United States, and one with Chinese-origin cloth for other markets. In such cases the difference between the NTM-compliant and non-NTM compliant costs, plus any costs of maintaining separate record-keeping and inventories, may be considered as part of the NTM price gap embedded in the product price.

(labeled in Figure 1 “ex-farm/ex-factory”) and “delivered duty paid” (corresponding approximately to “landed duty paid” in Figure 1) at which the costs and responsibilities between the buyer and seller can be divided differently. See also the discussion of import port procedures, below.

(ii) *Movement to port*

In an undistorted market, with competition among trucking companies, the difference between the ex-farm/ex-factory price and the price at the port gate should equal the average cost of providing trucking services along the route in question. Such costs may be high both for natural reasons, such as landlocked status or difficult terrain, and for artificial reasons such as lack of competition among trucking companies. The monetary costs of trucking alone are on the order of 10-50 percent of the ex-farm price for most exports from the Central African Republic, Chad, and Zambia. The time costs associated with road transport alone are higher, as high as 166.8 percent for cotton exports from Chad (Christ and Ferrantino 2011). Some of these costs are due to geographic difficulties in remote locations. Additional costs may arise due to the cartelization of trucking services (Arvis, Raballand and Marteau 2010), which is widespread in sub-Saharan Africa and which interacts with physical difficulty of transport — roads which flood or break axles reduce the number of trucking companies willing to travel down them, which in turn makes it easier for the companies willing to serve the market to collude. Both the physical difficulty of transport and the cartelization of transport may be addressed by policy — the former by investments in infrastructure, the latter through competition policy and market-based reforms.

(iii) *Export procedures*

Export procedures may take place at a seaport, airport, or land crossing. If we consider the example of a seaport, which is still the most common case, these procedures consist of warehousing, yard procedures such as stacking of containers, loading of ships, and various bureaucratic formalities. Some of these are analogous to import port procedures and will be discussed in more detail under that heading. Port warehousing is often in short supply in developing countries, increasing its cost when it is available or causing time delays which have a tariff equivalent. In many cases, trucks are used as makeshift warehouses with the associated risk of theft of cargo. Thus, improvement of warehousing by port authorities has a tariff (or more precisely, export tax) equivalent in terms of both cost, time, and uncertainty. Bureaucratic procedures associated with exporting are generally less than those associated with importing, because export taxes are less often collected then customs duties. Nonetheless, such procedures can cause a non-trivial burden on exporters. A recent survey of traders in six developing
countries,\textsuperscript{13} conducted by UNCTAD and ITC-Geneva, identified 6,225 cases of NTMs, of which approximately 2.6 percent (about 160) were related to export procedures (Basu, Kuwahara and Dumesnil 2011). The bulk of the complaints concerning export NTMs were about certification requirements imposed by the exporting country (e.g. for SPS purposes), licensing or permit requirements imposed by the exporting country in order to export; and export taxes.\textsuperscript{14}

(iv) \textit{International transport}

Technological improvements in shipping, such as the increasing use of containerization as opposed to bulk shipping with its concomitant economies of scale, regular liner routes as opposed to tramp shipping, and the electronic tracking made possible by bar codes on containers, have induced a secular reduction in costs associated with international transport over time. Similar gains in technical efficiency have influenced air transport, at least in the developed countries. For the four countries which measure both a CIF and an FOB price for the same import transactions – Australia, Brazil, Chile, and the United States – the CIF-FOB margin has declined both on average and for both sea and air modes in all cases, except for air transport in Chile (Table 1). This means that improvements in shipping efficiency have been more than sufficient to outweigh the increasing share of international cargo that moves by air.

\textbf{Table 1.}

\textbf{Average Trade Costs (CIF-FOB margin, expressed as a proportion to the FOB price)}

<table>
<thead>
<tr>
<th>Year</th>
<th>Chile All</th>
<th>Chile Air</th>
<th>Chile Sea</th>
<th>Brazil All</th>
<th>Brazil Air</th>
<th>Brazil Sea</th>
<th>United States All</th>
<th>United States Air</th>
<th>United States Sea</th>
<th>Australia All</th>
<th>Australia Air</th>
<th>Australia Sea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>0.093</td>
<td>0.087</td>
<td>0.096</td>
<td>0.087</td>
<td>0.068</td>
<td>0.100</td>
<td>0.050</td>
<td>0.040</td>
<td>0.053</td>
<td>0.080</td>
<td>0.066</td>
<td>0.086</td>
</tr>
<tr>
<td>2008</td>
<td>0.078</td>
<td>0.087</td>
<td>0.076</td>
<td>0.053</td>
<td>0.065</td>
<td>0.051</td>
<td>0.038</td>
<td>0.026</td>
<td>0.043</td>
<td>0.049</td>
<td>0.036</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Source: Pomfret and Sourdin (2010)

International shipping costs which exceed the technological maximum may be due either to technological inefficiencies (e.g. smaller ships on certain routes) or to market imperfections. Private

\textsuperscript{13} Chile, India, Philippines, Thailand, Tunisia and Uganda.
\textsuperscript{14} Author’s tabulations based on UNCTAD Pilot Project survey data.
anticompetitive practices raise shipping rates both by sea and air. Global shipping alliances, known as “conferences”, dominate containership service, and consolidation has proceeded both by mergers and joint ventures. In some cases, shipping conferences are given exemptions from national antitrust laws. Estimates of the impact of maritime conferences, price-fixing agreements and associated cooperative working agreements (including vessel-sharing) vary. Clark, Dollar and Micco (2004) find that maritime conferences add at most 5 percent to transport costs, while Fink, Mattoo and Neagu (2002) estimate that on U.S. routes, the breakup of cooperative working agreements would decrease transport costs by more than 7 percent, while the breakup of price fixing agreements would cause prices to decline by a further 19 percent. Micco and Servrisky (2004) find that increased air competition, such as that associated with the U.S. Open Skies agreement in the 1990s, could reduce air transport costs by 8 percent. Francois and Wooton (2001), in a theoretical paper illustrated with empirical data, show that the gains from tariff liberalization in the presence of market imperfections in shipping could be limited, accruing primarily as additional rents to shipping firms rather than as gains to exporting producers and importing consumers.\textsuperscript{15}

The costs of shipping services vary across products and countries in ways that remain to be fully documented, though data are improving (OECD 2011). It is certain, though, that the costs associated with market power fall more heavily on developing countries and smaller markets, simply because the number of daily departures available in a port such as Mombasa or Douala is far fewer than those serving Los Angeles, Shanghai, or Singapore. The same market forces that cause U.S. air travelers to pay higher ticket prices traveling to Boise or Cheyenne than to New York or Chicago boost air and shipping rates to developing countries, even if all technical inefficiencies associated with transport were absent.

(v) \textit{Import port procedures}

The costs associated with import port procedures vary widely from location to location, and over time. Blonigen and Wilson (2006) use U.S. trade data on “import charges” to identify the relative efficiency of U.S. and foreign ports. Import charges are defined as “the aggregate cost of all freight, insurance, and other charges (excluding U.S. import duties) incurred in bringing the merchandise from

\textsuperscript{15} See USITC (2005, chapter 5), for further discussion of both technological and market-based impediments to logistic inefficiency.
alongside the carrier at the port of exportation – in the country of exportation – and placing it alongside
the carrier at the first port of entry into the United States.” This amount includes costs associated with
the non-U.S. export port, international transport costs, and costs associated with the U.S. import port, i.e.
the CIF-FAS margin.16 There are substantial differences in costs associated even with U.S. import ports.
For example, import costs associated with the port of Gulfport, Mississippi are about 8 percent lower
than those associated with the reference port of Oakland, California, while those associated with Port
Arthur, Texas, are 26 percent higher than Oakland, and import costs for San Juan and Honolulu appear
to be even higher. Similar differences exist in costs associated with exporting to the United States –
export port costs associated with Bonny, Nigeria are 79 percent higher than those associated with
Rotterdam.17

Port efficiency can be associated with both technical factors (such as crane unloading moves per
hour and efficiency of stacking and unstacking containers) and with management factors potentially
addressable by public policy. For example, the ownership and operation of port assets can be structured
in a number of different ways. Government authorities may own and operate port infrastructure (a
“service port”), allow private firms to supply port and maritime auxiliary services (a “landlord port”), or
also allow private firms to lease and operate port assets (a “tool port”). Privatization of port assets can be
associated with efficiency gains (Fink, Mattoo and Neagu 2002; Londoño-Kent and Kent 2003). Further,
efficiency can be measured for different parts of port procedures, which may be associated with different
public and private actors. Table 1 gives a breakdown of procedures in an import port which can be
further used to decompose inefficiencies and policy-related costs associated with import charges. Note
that some of these procedures (e.g. fumigation) may also be associated with SPS policies.

16 The difference between FAS (free alongside ship) and FOB (free on board) is that the former does not
include the costs of loading the vessel, while the latter does. The CIF-FAS margin is thus slightly higher than the
CIF-FOB margin. FAS is referred to as “customs value” in U.S. trade data, since customs duties are levied on the
FAS value rather than the CIF value as in most other countries.

17 Calculated by the author using an exponential transformation of reported fixed effects in Blonigen and
Table 1.

The Twelve Sub-Stages of Import Port Procedures

1. The process of physically guiding the vessel into the berth, which involves navigation, pilotage, tug assist and line handling charges;
2. Application of berthage or “parking” charges to vessels secured to the berth;
3. Inspection, e.g. for security and drug enforcement;
4. Unloading by crane, the efficiency of which is measured in crane moves per hour, and which incurs charges if the port’s rather the ship’s crane is used;
5. Charges for “wharfage,” the use of the apron and other areas in which cargo is moved around;
6. Inspecting the seal;
7. Dispatching the cargo to and from an assigned spot in the yard;
8. Storage, either in the port or in an alternate storage facility such as a bonded warehouse;
9. Customs clearance per se;
10. Fumigation, if necessary;
11. Possible charges for trucks enter the port from inland; and
12. Gate processing at the point of physical exit from the port.

Each stage involves identifiable costs and/or time.


(vi) Customs

Costs and delays associated with customs procedures are widespread. Customs procedures giving rise to complaints include documentation requirements, direct consignment requirements (goods must be shipped directly from the country of origin without passing through a third country), restrictive regulations on land, sea, and air transportation, and requirements to pass through a specified port of customs, which could slow down the import clearance process. Some requirements to pass through a specific point of entry are associated with SPS and TBT testing. 18 Using 2007 data from the World Bank’s Doing Business Trading Across Borders data, Minor and Tsigas (2008) calculate that time associated with customs procedures ranges from an average of 2.0 days in high-income Europe to 9.1 days in low-income sub-Saharan Africa, with better and worse performances in specific countries. The CoRe NTM database (Martinez, Mora and Signoret 2009), which compiles traders’ complaints gathered indirectly through the EU’s Market Access Database, USTR’s National Trade Estimate, and WTO’s

---

18 See http://ntb.unctad.org/docs/Classification%20of%20NTMs.pdf for the MAST classification of NTMs, particularly A380, B380, and C200.
Trade Policy Reviews, reports concerns about customs procedures in 59 different customs territories, including trading nations large and small.

Some costs and delays associated with customs procedures can be addressed by trade facilitation policies; for example, the replacement of paper recordkeeping with electronic systems (Yasui and Engman 2005), or the use of risk assessment as an alternative to opening every package at the border (Moïsé 2004). There are other iceberg-type costs associated with the fact that a certain fraction of goods are denied entry due to the failure to comply with SPS and TBT standards, or with rules of origin. The costs and losses associated with customs denials can potentially be quantified using databases on customs refusal which are maintained by several countries.19

(vii) Wholesaling and retailing

The example of the Barbie doll, cited above, suggests that a large part of the markup between the ex-factory price and the retail price consists of wholesale and retail markups. Studies based on “tear-down” reports of electronics confirm this. In the often-cited example of the Apple iPod, Linden, Kraemer and Dedrick (2007) assign $75 of the retail value of an iPod selling for $299, about 25 percent of the total, percent to distribution and retail operations in the United States, and another $80, about 27 percent, to Apple as compensation for intellectual property and organization of the supply chain. Looked at differently, the input costs for the iPod amount to $144.40, of which nearly all are produced in the United States. Wholesaling and retailing costs amount to a 52 percent markup over input costs, with the “U.S. markup” amounting to 107 percent if Apple’s shares are included. These calculations do not reflect international transport and distribution, nor do they reflect the relatively small share of value captured by China, but they are illustrative of point that the amount of total markups behind the border in the importing country can be substantial.

Some part of wholesale and retail markups are no doubt due to technical inefficiencies in logistics or bad transport, similar to that discussed previously under the heading of movement to port. Others may be associated with policies regarding the industrial organization of wholesaling and retailing, such as

19 An example is the U.S. Food and Drug Administration’s Import Refusal database, located at http://www.fda.gov/ForIndustry/ImportProgram/ImportRefusals/default.htm. The European Union and Japan maintain similar databases.
limited entry into logistics services, and restrictions on hours and locations of retail operations. These include both domestic barriers to entry and limitations on foreign direct investment in distribution, retailing, and domestic transport, which may be classified as NTMs insofar as they affect the market for internationally traded goods.

(b) A hub-and-spoke supply chain

The above discussion of NTMs and other costs along the supply chain assumes a fairly simple case in which a single good is moved from place to place without being transformed. However, many modern supply chains are more complex than this, involving different stages of production in different countries, and gathering components together from many locations for final assembly. This is particularly true for manufactures with multiple components such as electronics and motor vehicles.

A classic example of this is the production of a computer disk drive as discussed in Hiratsuka (2005) and Baldwin (2006). The disk drive is assembled in Thailand, which acts as the hub of the supply network, using 43 components from 10 other countries and 11 components produced in Thailand. Thus, there are at least 10 international moves of the type described above, and likely more, depending on the extent to which shipments can be bundled. Since the disk drive will be shipped to the location of final computer assembly (e.g. China), at which other major components are gathered, the number of cross-border moves multiplies. China then serves as a larger hub linking the disk drive hub as well as other hubs for major components. To all the cross-border moves in such a network must be added the final move of the finished product to the consumer.

As Hiratsuka (2005) notes, logistics firms have a number of strategies for reducing the number of times material has to be moved, including maintaining hub warehouses and the “milk-run” system, which involves regular truck runs within a country. It will be readily apparent that if there is a high degree of competition in sales of the final product, trade costs of all kinds must be reduced to a minimum in order for a hub-and-spoke supply chain to operate. If the sum total of NTMs, tariffs, transport and logistics costs in a certain region exceeds the maximum that can be borne by the final product price, the
hub-and-spoke supply chain will simply not operate in that region – trade costs will act as a prohibitive barrier to the establishment of supply chains.

(c) The cumulation of trade costs along the supply chain

It will be clear from the above discussion that in a supply chain which requires that semi-finished goods cross international borders more than once, the effects of NTMs and other trade costs are compounded. This implies that the effect of a marginal increase in trade costs everywhere in the supply chain is much larger than would be the case if there were a single international transaction.

This point can be illustrated by a simple example. Suppose that the total value-added necessary to produce a product is equal to 1. The product is produced in stages in n countries, each of which adds \( \frac{1}{n} \) to the total value of the product. After production, the product will be exported to a final destination, so that it is moved n times altogether. Let trade costs for moving the product from one country to another equal t on an *ad valorem* basis. At each stage, the trade cost t is charged on the entire value of the product produced up to that point, including previous trade costs. Let \( c(n) \) be the total cost of the product delivered to the final consumer when it is produced in n stages, so that

\[
  c(1) = (1+t)
\]
\[
  c(2) = \frac{1}{2}(1+t)^2 + \frac{1}{2}(1+t)
\]
\[
  c(3) = \frac{1}{3}(1+t)^3 + \frac{1}{3}(1+t)^2 + \frac{1}{3}(1+t)
\]
and in general,

\[
  c(n) = \sum_{i=1}^{n} \frac{1}{n}(1+t)^i
\]

Suppose that the trade cost at each stage is 10 percent *ad valorem*, so that t = 1 and \( c(1) = 1.1 \). As the value chain is sliced up further, trade costs compound fairly quickly: \( c(5) = 1.343 \) (a tariff equivalent of 34.3 percent) and \( c(10) = 1.753 \) (a tariff equivalent of 75.3 percent). Further, marginal increases in trade costs are compounded; if t increases from 0.1 to 0.2, a doubling at each stage of the supply chain,
trade costs along the supply chain more than double, with more compounding for longer supply chains
\(c(5) = 1.786, c(10) = 3.115.20\)

More formally, Yi (2003) has shown that when the structure of production is endogenous, so that
the degree of vertical specialization depends on tariffs, a small decrease in tariffs can induce a tipping
point at which producers introduce vertical specialization when it had previously not existed, so that there
is a large and non-linear increase in international trade. By the same token, an increase in NTMs or other
trade costs can have the reverse effect, making vertical specialization unprofitable, restricting trade to
more simple production patterns involving fewer countries in the production of a particular good, and
inducing a large and non-linear decrease in international trade. Such a model is more successful in
explaining how the tariff reductions in recent decades could have induced the large observed increase in
international trade than a model in which production takes place in one country alone.

It follows that the secular decrease in transport costs documented above was likely a driver in the
development of global supply chains, and that it benefited trade in goods with long supply chains more
than trade in goods with simple supply chains. Similarly, the reduction in tariffs in electronics associated
with the Information Technology Agreement of 1997 helped to foster the development of supply chains in
Asia (Anderson and Mohs 2011). The ITA reduced bound and applied tariffs for approximately 95
percent of information technology products to zero on a phased basis. Simple average tariffs for ITA
members were an estimated 3.6 percent before the agreement (Bora and Liu 2006). Some ITA members
implemented much larger tariff cuts, including India (from 36.3 percent), China (from 12.7 percent) and
Egypt (from 12.1 percent). Trade for ITA products is estimated to run around $4 trillion annually, and to
have grown at an 11 percent annual rate over 1996-2008 compared to 7 percent for all trade in
manufactures (Anderson and Mohs 2011). While some of this growth is no doubt due to income-elastic
demand for electronics that would have been observed without tariff reductions, the confluence of falling
tariff rates and falling transport costs has no doubt contributed to the further development of electronics
supply chains in Factory Asia.

\(^{20}\) For convenience, the terms of \(c(n)\) can be calculated recursively using
\[c(n) = \frac{1}{n} \times (1+t) \times [(n-1)(c(n-1)+1)].\]
At the other extreme, very high trade costs can inhibit the growth of supply chains altogether. Sub-Saharan Africa is virtually absent from the electronics supply chain, and participates to any great extent only in the supply chain for textiles and apparel, almost exclusively in on the lowest rung of cut, make, and trim assembly (CMT) using imported cloth (Gereffi and Frederick 2010). There is relatively little in the way of more elaborate relationships between producers and suppliers, such as exist in Bangladesh and Indonesia. While many factors inhibit the growth of the textiles and apparel supply chain, such as lack of electricity, high trade costs associated with maritime and land transport impose severe constraints on apparel producers in sub-Saharan Africa. Due primarily to delays in receiving imported inputs, apparel producers in Swaziland require roughly 130 days to produce and ship apparel to the United States, as compared to 30-5 days for apparel originating in India or China (USITC 2009a, 6-29 to 6-34). Cost and time delays associated with land transport alone may suffice to disqualify much of sub-Saharan Africa from the “just-in-time” delivery expectations required to participate in the electronics supply chain (Christ and Ferrantino 2011).

3. The role of standards in supply chains

It has often been noted that in a world of falling transport costs and tariffs, many of the remaining impediments to trade take the form of NTMs. For electronics, where tariffs and transport costs have already fallen substantially, reduction of NTMs offers the promise of further trade growth, particularly harmonization of standards (Portugal-Perez, Reyes and Wilson 2009). The available literature on standards and trade in general yields a mixed picture; use of international standards by either exporters or importers is likely to promote trade, or at least not to harm it, while use of national standards has a more ambiguous effect (Swann 2010). Studies of standards and trade that focus on sectors for which supply chains and intermediate goods are important find either that standardization per se promotes trade, or that international standards are more trade-promoting than national standards (Blind 2001 for medical instruments; Czubala, Shepherd and Wilson 2007 and Shepherd 2007 for textiles and clothing; Moenius 2006 for electrical products). Mutual recognition agreements have positive effects on trade in telecom equipment and medical devices (Baller 2007). All of these results taken together suggest that either
stronger standards, more international standards, or standards cooperation can be trade-promoting for goods involved in supply chains. Whether this makes goods with complex supply chains different from goods with simple supply chains is an open question.

In some cases, expanded trade is actually associated with weaker standards rather than stronger ones. For some primary products and commodities, the center of global demand has shifted from the OECD to large developing countries like China and India. This shift has become more pronounced in the wake of the global growth slowdown in developed countries beginning with the Great Recession of 2008-09. Kaplinsky and Farooki (2010) examine the implications of this demand shift for the relationship between supply chains and product standards. Countries at a middle stage of development, like China, India, and Brazil are more likely to be engaged in labor-intensive stages of intermediate processing. Thus, the poorest countries which used to export some semi-processed goods to the North will increasingly export less-processed goods to the South, in increasing volumes. Since unprocessed goods are less standards-intensive than intermediate goods, the new Southern importers are likely to demand less both in the way of production standards and labor standards from their suppliers than the old Northern importers did. Kaplinsky, Terheggen and Tijaja (2010) provide evidence for these patterns in case studies of Gabon timber exports and Thai cassava exports. In both cases, increased demand from China has pushed these exporters back into earlier stages of the value chain while relaxing the level of standards compliance necessary.
4. **Trade in intermediate goods and trade costs – available evidence**

The literature on standards suggests that there may be an association between NTMs and the volume of trade in supply chains, with stronger or more harmonized standards being associated with more trade in manufacturing supply chains, while increases in agricultural trade may be in some cases associated with weaker standards and movement of intermediate processing from low-income to middle-income countries. At present, there does not appear to be a clean test of the hypothesis that either NTMs or other trade costs make more of a difference for trade in complex supply chains than for trade in simple supply chains, as predicted by theory (Yi 2003), though there are some suggestive findings. This section examines some of the limited literature that bears on this topic, and then presents some simple econometric findings that may address the relationship between trade costs and supply chains more directly.

(a) **Econometric evidence**

Hanson, Mataloni, and Slaughter (2005) use firm-level data on U.S. multinationals to examine vertical production networks. Their data includes information on trade between U.S. parents and their foreign affiliates. Although the commodity content of this trade is unknown specifically, it is probable that much of it is trade in intermediate goods reflecting the operation of international supply chains. Their measure of trade costs is defined as the sum of ad valorem tariffs and freight rates. The elasticity of demand for imported inputs is estimated to be quite high, with a central estimate of -3.28 (i.e., a 1 percent reduction in input prices due to lower trade costs implies a 3.3 percent increase in intrafirm trade). This high value is interpreted as evidence in favor of the Yi hypothesis that trade costs have a magnified effect in vertical production networks. Their results for nontariff barriers are insignificant and of the wrong sign. However, their measure of nontariff barriers is derived from the UNCTAD TRAINS database, which has a number of limitations from the standpoint of empirical work (Ferrantino 2006, pp. 7-9).

database; cost of importing and exporting, number of documents required to import and export; and time required to import and export. All of these variables are negatively correlated with trade flows, as is to be expected. The authors do not explicitly consider trade in intermediate goods. However, they do find that the trade facilitation variables have a larger-than-average impact on trade in high technology goods; such goods are more likely than other goods to involve complex supply chains. The impact of trade facilitation variables when a developed country is the exporter is generally three to five times larger than when a developing country is the exporter. This result may seem counterintuitive, given that developing countries have poorer environments from the standpoint of trade facilitation. However, the coefficient measures the effect of a marginal change in the trade facilitation variable, e.g. one additional delay in exporting or importing. Since developed-country trade is more likely to consist of intermediate goods relative to primary or final goods, and since just-in-time vertical production networks are likely to be denser at higher stages of development, developed-country status may serve as a proxy for intermediate goods trade. The results in Martínez-Zarzoso and Márquez-Ramos (2007) are at least broadly consistent with the Yi hypothesis.

Ma and Assche (2010) take advantage of unique China Customs data associated with China’s processing trade regime. About half of China’s international trade is conducted under two special customs regimes, known jointly as “processing trade,” in which certain imports are identified as intermediate inputs destined to be embodied in exports and certain exports are identified as being produced using processing-trade inputs. The advantage of this is that specific trade flows can be identified directly as being involved in vertical production networks. The foreign content of China’s processing exports is approximately 82 percent, as compared to 11 percent for non-processing exports (Koopman, Wang, and Wei 2008). The tax and tariff treatment of processing trade is different from that of normal trade.

The measure of trade costs used in Ma and Assche (2010) is distance. They motivate their econometric work by a three-region theoretical model (China, East, and West). East and West are both advanced countries (more productive than China), but trade costs between China and East are relatively low compared to either China-West or East-West. Firms come in four types with respect to their position
in the supply chain; purely domestic firms (Type D); domestically headquarterd firms that offshore final goods production to China and sell in their home market (Type O); advanced-country firms that process at home and export to the other advanced country (Type X); and advanced-country firms that process in China and export to the other advanced country (Type T, for triangular trade). The types of firms are endogenous with respect to the parameters of the model, with trade costs being of particular interest.

The model yields several predictions, one of which is that China’s processing exports to East are more sensitive to export distance and less sensitive to import distance than its processing exports to the West. Put differently, the trade costs associated with intermediate goods are more important than the trade costs associated with final delivery. This result arises from the endogeneity of firm types with respect to trade costs; for example, an increase in trade costs from East to China would induce some type-T firms, e.g., Japanese multinationals which export intermediate goods from Japan to China and final goods from China to the United States, to become type-X firms processing in Japan and shipping directly to the United States. This prediction is confirmed by the empirical work. Ma and Assche (2010) also find evidence for the sensitivity of processing exports to oil prices, a component of transport costs. They also demonstrate the so-called “bullwhip” effect (a decline in processing exports leads to a larger decline in processing imports)\(^\text{21}\). The “bullwhip” effect holds for 15 of 20 product categories during the Great Recession period of 2008:Q1 to 2009:Q1.

(b) Case study evidence

NTMs affect parts and components in a number of industries. Since there can be several stages of assembly in complex manufactured goods – parts of parts and components of components – the cumulative imposition of NTMs at several stages of the production process can have a cascading negative effect on international trade analogous to that described by Yi (2003) for tariffs. While this section relies on somewhat dated descriptions of the measures in question, there is good reason to believe that similar problems persist in the market today.

\(^{21}\) See also Escaith, Lindenberg and Miroudot (2010).
Henson et al. (2000) describe costs associated with meeting regulatory requirements in automotive components and terminal telecommunications equipment between the United States, the United Kingdom, Germany, and Japan. For automotive components, mandatory technical requirements differed among all the countries, and were particularly problematic for such components as seat belts and exhaust systems. In the EU, satisfying standards for technical conformity, safety, and emissions requires laboratory testing to obtain a type approval certificate, with re-testing and re-inspection required for relatively small changes. Border inspections are also possible prior to importation. In the United States, there are federal, state, and local standards for automotive products – for example, in California emissions standards are particularly strict. There is also a de facto mandatory quality standard (QS9000) among the “Big Three” (Chrysler, General Motors and Ford). EU exporters state that the distinction between essential safety regulations and optional quality requirements in the U.S. market is unclear, due in part to the role of private providers of assessment and certification.

As of 2000, the Common Technical Requirements (CTRs) for terminal telecommunications equipment in the EU did not cover all types of terminals and components, requiring U.S. manufacturers to undergo the costs of meeting various national standards. Even when common EU standards exist, there are costs of testing to obtain the CE mark. Differences within the EU and between the EU, Japan, and the United States limit the ability of manufacturers to achieve economies of scale in the manufacture of components.

As a result of differing standards, companies can either incorporate into the original design the special features demanded by individual markets, or design for the domestic market only and make adaptations once export markets are identified. The former strategy is feasible for large multinationals, but rarely for smaller firms who may lose export markets due to the costs of adapting products. The costs of complying with foreign product requirements and assessments are difficult to assess ex ante, so that firms face significant uncertainty in estimating compliance costs – for example, firms complain that

---

22 The study also considered dairy products, which raise different sorts of issues from those focused on here.
conformity assessment bodies are inconsistent in their assessment of products. Moreover, non-mandatory or local standards can be at least as problematic as national standards.

It has been argued that procedures of national telephone authorities for approving imported telecommunications are arbitrary, undefined, or unavailable, and designed to limit the importation of foreign telecommunications components. (Linvill et al. 1984). This characterization, dating from the 1980s, takes for granted the existence of national telephone service monopolies. Competitive reforms in telephone service provision may since have liberalized the market for telephone components, but similar problems may persist in less liberalized markets. An analysis of U.S. trade sanctions on Cuba by the U.S. International Trade Commission suggests that even if such sanctions were lifted, U.S. exports of telecommunications equipment and components sold in other international markets would be limited in Cuba because both the landline and cellular networks had evolved on a base of European and Asian equipment, thus rendering much U.S. equipment non-interoperable (USITC 2001, 6-29 ff.)

5. Implications for policy

How does looking at NTMs and trade facilitation from a supply chain point of view provide insights into policy? There are a number of positive statements that can be made about the way NTMs and supply chains affect the global economy, and these have normative implications. Some examples of these are presented below, and there are no doubt others.

(a) There are low-level development traps associated with NTMs and lack of trade facilitation, and these are manifest both on a national and on a regional basis.

Since the effects of NTMs and other trade costs compounds along the supply chain, NTMs can have a discontinuous effect on trade flows. Increased levels of trade costs can lead to a “tipping point” beyond which the operation of a modern supply chain becomes simply infeasible. Since global supply chains seek to minimize transactions costs, they often operate on a regional basis, e.g. East Asia for electronics, North America for motor vehicles. It is necessary for the supply chain to be successful that
both physical and government-induced trade costs be minimized. The near absence of electronics and automotive trade in sub-Saharan Africa can be attributed in part to trade costs associated with NTMs. If each country in a region has lengthy customs delays, non-automatic licensing, and technical inefficiency at border checkpoints, then in effect the region is no longer a region from the standpoint of trade, because of internal distance.

(b) **Lack of connections to developing world**

The examples of China and Mexico as successful exporters of final goods produced by supply chains emphasizes the value of being economically close to the source of final demand. This is accomplished in the case of Mexico by contiguity to the United States, and in the case of China by relatively good shipping connections and efficient export procedures. However, for many countries NTMs take the form of licensing and certification requirements that slow exports, and technical inefficiency in export ports can also inhibit trade. On top of this, there are relatively few sea and air connections to many developing countries, inhibiting competition and raising the cost of trade further. Thus, many countries are further away geographically from final markets than they would otherwise appear to be, and are knocked out of the final-assembly part of supply chains.

(c) **Time barriers are particularly important**

The operation of regional supply chains requires the close coordination of the steps of production taking place in different countries, with a smooth coordination of inventories of intermediate goods, delivery of final goods, and return of defective products. In the old days of large integrated manufacturing, mammoth facilities such as Ford’s River Rouge auto factory in Michigan accomplished the objective of just-in-time manufacturing by physically locating operations side by side. When the “factory” for a single disk drive extends from Japan to Indonesia, any delays at border checkpoints have a magnified effect on technical inefficiency. In many cases it is possible from a technical standpoint to
conduct dispersed manufacturing over a large geographical area, but not possible from an economic standpoint because of government-induced delays at borders.

(d) The role of standards varies by the stage of development. NTMs of the standards type can either promote or inhibit trade, depending on the situation.

As discussed above, the way that product standards interact with international trade is complex. Harmonized standards can promote trade, and also make supply chains more efficient. The cost of non-harmonization can be easily viewed in a laptop power supply, which bears many small symbols printed in white indicating various government and private entities that must test the power supply for such reasons as radio non-interference. When harmonization takes place, it is possible for a producer of intermediate goods to follow its customers into more markets and take place in the supply chain in more locations.

Other standards, whose main effect is to add production costs in order to enhance product quality for the final consumer, act more like traditional NTMs and have a trade-reducing effect which can be measured as a tariff equivalent. For standards like these, the traditional cost-benefit considerations apply – do the social benefits of higher product quality and safety outweigh the costs of imposing the standard?

As the center of gravity of the global economy shifts increasingly to large developing economies such as China and India, the demand for unprocessed and intermediate goods changes. This means not only that poorer suppliers of raw materials are pushed “upstream” in the supply chain, but that their goods are expected to meet lower product standards than they would if they were sold in developed countries. Industrial strategy in developing countries, whether private or public, needs to take this into account. Is it better to sell larger volumes to big developing economies, and save the costs of complying with elaborate product standards, or is it better to bear the costs of standards compliance, sell to developed economies at higher unit values, and possibly retain more steps of the production process at home? Are there sufficient economies of scale at the national level to permit both types of markets to be served simultaneously?
(e) **Regional initiatives can help bring supply chains to new parts of the world**

The gains from improving efficiency of customs procedures, reducing the number of non-automatic licenses required, reducing corruption, improving physical conditions in ports, and similar measures can be multiplied if several countries in a region undertake such reforms together. Just as many regions are on the wrong side of the “tipping point” and do not attract global supply chains at present, the simultaneous reduction of trade costs in several neighboring countries is likely to have benefits over and above the benefits to each individual country, as it becomes feasible to locate several steps of a production process in different locations within a region to achieve stage-specific economies of scale.

Mutual recognition agreements of conformity assessment procedures have had significant effects on the cost of compliance with standards (Henson et al. 2000). Regional trade agreements can also lower the costs of NTMs for intermediate goods. For example, the NAFTA enabled U.S. companies to freely attach their terminal equipment to the Mexican telephone network, and provided that a single laboratory can be recognized to test a telecommunications product in any of the NAFTA countries (Trade Compliance Center).

(f) **NTMs affecting logistics and related services are particularly important**

In almost all cases, the successful operation of supply chains is facilitated by third-party logistics firms (3PLs), which provide coordinated services in supply-chain consulting, transport management, freight transport services, trade finance, express delivery, wholesale trade, packing, product returns, customs brokerage, and other areas (USITC 2005). In many countries national policies create barriers to entry for logistic services, which inhibits the growth of supply chains and thus international trade. This suggests one direct connection between trade policy and supply chains. Measures to liberalize market access in logistic services, whether unilateral, embodied in FTAs, or in the form of GATS commitments, can substantially enhance the feasibility and lower the costs of operating supply chains, with a concomitant growth in international trade.
REFERENCES


