I. Introduction

This chapter will review the basic economics of trade in services. Much of the logic behind trade and protection in services the same as for goods trade. However, there are also a number of issues unique to services which will be highlighted.

Service trade differs from goods trade in two major ways. First, while the majority of goods trade involves shipping goods from one country to another; cross border trade accounts for less than half of trade in services. For services which require personal contact between customers and clients, trade is possible only via sales through a foreign affiliate or if either the customer or producer travels across borders. While foreign investment and labour mobility are also issues affecting goods trade, they are fundamental aspects of trade for some services.

Second, services tend to be highly regulated. Many types of services are publicly provided or are produced by regulated monopolies. In contrast to goods, relatively few services are subject to simple discriminatory taxes on trade. Instead barriers to trade in services arise from domestic regulations that often serve the dual purpose of responding to market failures (such as ensuring quality standards for medical practitioners) and protecting local suppliers from foreign competition. This means that identifying and measuring trade barriers in the service sector is very complex. It also means that simple rules for trade liberalization that worked for goods trade (such as reducing all tariffs by 30%) are not available as an option for service trade liberalization. Instead service trade liberalization is organized around the notion of non-discrimination and is often linked with domestic regulatory reform.
This chapter is organized as follows. Section II briefly provides an overview of how services are traded and reviews the four modes of supply: cross-border trade, consumption abroad, foreign direct investment, and labour movement. In section III we review the reasons for trade in services and discuss the sources of potential welfare as well as the effects of liberalization on income distribution. These issues are very similar to the standard analysis of trade in goods and factors. In section IV, we discuss interactions between the different modes of supply. This is important because restrictions on some modes (such as labour mobility) may either render some services non-tradable, or may force service providers to use another (possibly less efficient) mode. Section V analyzes the effects of some of the most common barriers to trade in services. We point out via some examples how the effects of trade liberalization cannot be analyzed independently of the domestic regulatory system. Section VI concludes with a discussion of the relative merits of trade agreements versus unilateral liberalization of service trade.

II. What are services and how are they traded?

A. What are services?

It may be useful to begin by asking: how are services different from goods? Services are often seen as intangible, invisible and perishable, requiring simultaneous production and consumption. Goods, in contrast, are tangible, visible and storable – and hence do not require direct interaction between producers and consumers. However, there are exceptions to each of these characteristics of services: a software program on a diskette or an architect’s design on paper are both tangible and storable, many artistic performances are visible, and automated cash-dispensing machines make face-to-face contact between producers and consumers unnecessary. These exceptions do not however detract from the usefulness of the general definition of services presented above.
Instead of worrying about a precise definition of what a service is, it may be more useful to consider examples of various types of services. Below, we list the range of services covered by the General Agreement on Trade in Services (GATS):

1. Business services  
2. Communication services  
3. Construction services  
4. Distribution services  
5. Educational services  
6. Environmental services  
7. Financial services  
8. Health-related and social services  
9. Tourism and travel-related services  
10. Recreational, cultural and sporting services  
11. Transport services  
12. Other services not elsewhere included

**B. How are services traded?**

We can distinguish between services that necessarily require physical proximity between the user and the provider and those that do not. For many services – whose number is growing with the development of electronic means of delivery – proximity is not necessary, though it may enhance the quality of the service. A variety of financial, entertainment, information and communication services can be produced in one country and delivered, either electronically or stored in some medium (paper, disk, cassette), to consumers in another country. Trade in these services is not much different from trade in goods.

A number of services, however, require proximity between the consumer and the producer, one of whom must move to make an international transaction possible. Such services include construction services, where the supplier moves to the location of the consumer; tourism, where the consumer moves to the location of the supplier; and hair cuts or surgical operations, where either the supplier or the consumer moves. The movement of the supplier could involve the flow of capital, i.e. foreign direct investment (FDI), labor, or both.
Thus, there are four ways in which international service transactions take place (four modes of delivery) which can be categorized as follows:

**Cross-border:** services supplied from the territory of one country into the territory of another. Examples include financial transactions conducted over the phone, and software services supplied by a supplier in one country through mail or electronic means to consumers in another country.

**Consumption abroad:** services supplied in the territory of one country to the consumers of another. Examples are where the consumer moves, e.g. to consume tourism, education or medical services in another country. Also covered are activities such as ship repair abroad, where only the property of the consumer moves.

**Commercial presence (Foreign direct investment):** services supplied through any type of business or professional establishment of one country in the territory of another. An example is an insurance company owned by citizens of one country establishing a branch in another country.

**Presence of natural persons (labour movement):** services supplied by nationals of one country in the territory of another. This mode includes both independent service suppliers, and employees of the services supplier of another country. Examples are a doctor of one country supplying through his physical presence services in another country, or the foreign employees of a foreign bank.

Goods production also frequently involves the movement of producers. After all, many multinationals produce cars and shoes in foreign locations for foreign consumers. However, in the case of goods, such movement is a *substitute* for cross-border trade: the cars and shoes could have been produced at home and shipped across frontiers. The same
is true for some services: a legal service can be supplied over the phone or through the movement of the lawyer. But the important difference is that in the case of many services, cross-border trade is either not possible at all, or only possible if complemented by some form of local presence. A nanny must move if there is to be trade in child-care services; some types of software or insurance company may be able to transmit services long distance provided their representatives are in face-to-face contact with the consumer.
Box 1. Some stylized facts about services

- Services account for a large and growing share of both production and employment in most countries. In developing countries, the average share of services in GDP increased from around 40 per cent in 1965 to around 50 per cent in 1999, while in the OECD countries, the average share increased over the same period from 54 per cent to over 60 per cent. Many of the fastest growing sectors in many countries are services like telecommunications, software, and finance.

- The share of services in world trade and investment has also been increasing. They have been among the fastest growing components of world trade, growing by over 14 per cent per annum over the last twenty years (see Table 1). Services trade, as estimated from balance of payments statistics, was around $1350 billion last year, representing over one-fifth of world trade in goods and services. This value is certainly understated, because much "trade" in services takes place through an established presence, i.e via FDI, and hence generates local activity and value added that do not appear as exports in balance of payments statistics.

<table>
<thead>
<tr>
<th>Table 1. World Exports of Commercial Services, 1980-1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>(in billions of US dollars)</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>World</td>
</tr>
<tr>
<td>North America</td>
</tr>
<tr>
<td>Latin America</td>
</tr>
<tr>
<td>Western Europe</td>
</tr>
<tr>
<td>Africa</td>
</tr>
<tr>
<td>Asia</td>
</tr>
</tbody>
</table>

Source: WTO, 2000

1/ Annual average growth rate between 1980-99

- Today, more than half of annual world FDI flows are in services, and the value of sales abroad by foreign affiliates of US service firms is estimated to be 3.5 times greater than their cross-border exports.

There are two main reasons why trade and investment in services are growing:

- Technological progress, especially in telecommunications and information technology, has greatly enhanced the scope for trade in conventional services, like education and finance, and also created a host of new tradable services, such as software development and internet access.

- There has been a strong trend towards liberalization and regulatory reform in key service industries, creating for the first time scope for private and foreign provision. This trend began with the Thatcherite revolution in the UK in late 1970s, and pursued also in the US, in areas like air transport, rail and telecommunications. Now most countries around the world are allowing private and foreign provision in increasingly competitive markets for services like telecommunications, transport, and finance.
III. What are the causes of trade in services?

There are two major explanations for trade between countries: comparative advantage, and gains from specialization arising from increasing returns to scale or agglomeration effects. Both of these explanations apply to service trade as well as to goods trade. Moreover, both explanations apply not only to cross-border trade, but also to other modes of trade, including commercial presence and movement of natural persons. The first explanation relies on fundamental differences between countries to generate trade. The second approach can explain trade between similar countries: differences may emerge because of trade, but the differences need not have been present at the outset to generate trade. We will discuss each of these in turn.

A. Comparative Advantage

Consider the following examples of service trade: call centers in India provide customer contact services for US firms; nannies from the Philippines move to Canada temporarily to provide childcare services; and Europeans travel to Peru for a week in the jungle as part of an eco-tourism package. In each of these examples, trade takes place via a different mode: the call center services are sent across border using the telephone; the nannies travel to a foreign country to provide their services (movement of natural persons); and the eco-tourism consumers engage in consumption abroad. However, each of these examples has something in common: trade is driven by differences between countries. The trade in both child care and call center services are driven by differences in labor costs across countries; and the Amazon has unique attributes that are not available at home to the European tourists.

Differences in technology, natural resources, land/labor ratios, government policies, institutions, and other factors can all lead to differences in the prices of both inputs and outputs in the absence of trade. These price differences create incentives to trade.
One of the major sources of differences between countries is in factor supplies; that is, differences in availability of arable land, skilled labor, capital, etc. A country with an abundance of forested land will likely export forest products, a country with an abundance of highly skilled workers will export goods and services which are intensive in their use of skilled labor, and a country with an abundance of labor relative to capital and land will likely have relatively low wages and so export labor-intensive goods and services.

Differences between countries can also arise from a variety of other sources. Technological differences affect trade in both goods and services. For example, countries with access to sophisticated medical technology will export medical services. Differences in institutions and legal systems can also be important. Services such as insurance, for example, require that the client trust that contracts will be honored a claim is made. Differences in legal and regulatory systems affect the degree of confidence that a foreign client has in a firm, and this can affect its export success. Differences in regulatory systems can also affect the speed and flexibility with which service providers can respond to customers' needs, and this too will generate differences between countries that affect trade patterns.

Some differences, such as skill levels and technological knowledge, are not innate characteristics of countries but evolve over time in response to economic decisions made by policy-makers and individuals within a country. It is therefore useful to distinguish between short-run differences between countries and those which persist in the long run. Knowledge of new technology will diffuse across countries. Hence specific technologically-based differences may only provide short-run explanations of trade in specific goods. However, countries which have institutions which encourage and reward innovation should on average be expected to export innovation-intensive products or processes even in the long run.
Gains from trade

Trade driven by differences between countries generates two types of potential benefits. Producers gain from access to a larger market and higher prices, and consumers gain because they get access to both a wider variety of goods and services and to lower priced imported goods and services.

Standard trade theory predicts that if markets are perfectly competitive, then a country will always gain from trade, in the sense that the country as a whole can consume more goods and services after trading than before. The logic for this result relies on the simple premise that if markets are perfectly competitive, then profit maximizing firms will end up maximizing the value of national income.

While the gains from trade via export opportunities may seem self-evident, the gains from increased imports may seem less so. As an example, consider a country that imports software programming services and exports clothing, and for clarity, suppose these are the only products. Suppose it costs $100 to import a software package that would have cost $120 to produce at home. If markets are competitive, the $120 represents the opportunity cost to the economy of producing the software program itself. That is, the workers who produced software could produce clothing instead. And the value of the foregone clothing production must be $120 if markets are competitive. So this means that if the economy produces the software package itself, it has to give up $120 worth of clothing; but if it imports it, then it only has to give up $100 worth of clothing. Consequently, importing the software generates real gains to the economy, in the sense that it can consume more of both software and clothing with the same resources.

One of the major differences between service trade and goods trade is that much service trade must take place via movements of factors - such as movement of labour or foreign investment. If trade must take place through the movement of factors, are the basic propositions of trade theory – based on the notion of cross border trade – put into question? One problem does arise from the point of view of positive theory: if different
modes of supply are close substitutes, it is not easy to predict whether comparative advantage will manifest itself as a trade flow, investment flow or labour flow. However, from the point of view of normative theory there is no obvious problem: a country gains from the import of services, irrespective of the choice of mode, if the terms at which international transactions take place are more favourable than those available on the domestic market.

*Figure 1: The Welfare Economics of Trade in Services*

We illustrate this with an analysis of factor mobility (adapting Bhagwati, ). Assume a world where there are two countries, “home” and “abroad” and two factors of production: capital and lawyers. Capital and lawyers together produce legal services which can only be traded by the movement of lawyers to the location of the consumers. Capital is assumed to be immobile but the same analysis applies to capital mobility.
In the diagram, the width of the box is equal to the total lawyer endowment of the two countries. Every point along the horizontal axis represents an allocation of lawyers between home and abroad. The number of lawyers at home is measured from the origin 0, and the number of lawyers abroad is measured from the origin 0*. Suppose initially there are 0L1 lawyers are at home, and the remaining 0*L1 lawyers are abroad. The two lines, MPL and MPL*, represent the marginal products of lawyers in each country, holding the amount of other factors and know-how constant. MPL and MPL* are assumed to be declining in the number of lawyers for the usual reason of diminishing returns.

Now assuming that factors are paid their marginal product, lawyers will be paid L1C at home and L1B abroad. By adding up the marginal product of each lawyer in a particular country, we can find out the total product (that is, output or GDP) of that country. This equals the area under the MP curve for that country, up to the number of lawyers it has. For example, the output of home is everything under MPL to the left of point L1, while the output abroad is everything under the MPL* curve, to the right of point L1.

Note that with this initial allocation, the return to lawyers is lower at home than abroad. If lawyers can move, some will leave home and go abroad to take advantage of these higher returns. This flow from home to abroad will stop when the return to lawyers is equal in both countries, which will occur where the MP curves cross and L1L2 lawyers have moved. At this point, home will have OL2 lawyers, and abroad will have 0*L2 units. The equalized return will be L2A.

Now consider the impact of the movement of lawyers on aggregate economic welfare and on income distribution, both internationally and nationally.
Global welfare

Total world output increases by the area ABD and ADC. Allowing factors to move to their most productive locations has clearly increased global welfare.

Welfare Economics A: Lawyers that Move Counted in GDP of Abroad

<table>
<thead>
<tr>
<th>Capital (Home)</th>
<th>Lawyers Staying at Home</th>
<th>Output (Home)</th>
<th>Capital (Abroad)</th>
<th>Lawyers: Old and New (Abroad)</th>
<th>Output (Abroad)</th>
<th>World Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Trade</td>
<td>FCM</td>
<td>CMOL₁</td>
<td>FOL₁C</td>
<td>HBP</td>
<td>BPO* L₁</td>
<td>HO* L₁</td>
</tr>
<tr>
<td>After Trade</td>
<td>FAN</td>
<td>ANO L²</td>
<td>FOL₂A</td>
<td>HQA</td>
<td>AQQ* L²</td>
<td>HO* L²A</td>
</tr>
<tr>
<td>Change</td>
<td>- ACMN</td>
<td>+AEMN₁</td>
<td>-AC L₁L²</td>
<td>+ PBAQ</td>
<td>+AD L₁L²</td>
<td>+ABD +ADC</td>
</tr>
</tbody>
</table>

Welfare Economics B: Lawyers that Move Counted in GDP of Home

<table>
<thead>
<tr>
<th>Capital (Home)</th>
<th>Home Lawyers</th>
<th>Output (Home)</th>
<th>Capital (Abroad)</th>
<th>Lawyers: Old (Abroad)</th>
<th>Output (Abroad)</th>
<th>World Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Trade</td>
<td>FCM</td>
<td>CMOL₁</td>
<td>FOL₁C</td>
<td>HBP</td>
<td>BPO* L₁</td>
<td>HO* L₁</td>
</tr>
<tr>
<td>After Trade</td>
<td>FAN</td>
<td>+ADL₁L²</td>
<td>+ADL₁L²</td>
<td>HQA</td>
<td>DQQ* L²</td>
<td>HO* L²A-AD</td>
</tr>
<tr>
<td>Change</td>
<td>- ACMN</td>
<td>+ADCE</td>
<td>+ADC</td>
<td>+ PBAQ</td>
<td>-BPQD</td>
<td>+ABD +ADC</td>
</tr>
</tbody>
</table>

International income distribution

World welfare improves. But whether welfare improves for home, the country of emigration, depends on whether the welfare of moving lawyers is included. If such movement is permanent migration, and excluded from the calculation of national welfare, then welfare declines at home by the amount AC L₁L² (first table above) but increases abroad by the amount AB L₁L². However, if such movement is temporary, or if migrants
make substantial remittances, then welfare can increase at home as well. In the limiting case, if all the labor earned abroad is remitted home or is considered part of the welfare of the country from where the lawyers left, total welfare increases by the area ADC. In this case, welfare of abroad also increases but by a smaller amount than before—ABD.

National income distribution

The lawyers who originally worked in home and remained at home \( (\text{OL}^2) \) receive higher returns equal to the amount AEMN: their marginal product has increased because each lawyer has more capital to work with than before. By the same token capital at home loses by an amount ACMN because its marginal product declines. Thus there is a redistribution of income at home. Similarly, those who originally worked \( \text{(O* L}^1) \) abroad receive lower wages and lose by an amount equal to BPQD. Capital abroad benefits from the increased supply of lawyers by an amount PBAQ. Thus, in both countries there is a significant redistribution of income, that plague liberalization with political economy problems.

More generally, trade theory predicts that factors that have strong ties to import-competing industries will tend to lose from trade, while factors with strong ties to export industries will tend to gain. This means that there is unlikely to ever be unanimity regarding the merits of trade liberalization. It also means that it can be quite rational for individuals and interest groups to oppose trade liberalization. In principle, this opposition can be overcome by taxing the gainers to compensate the losers, but this is rarely done and likely to be too complex to be practical. It may also delay or hamper the reallocation of factors of production. Consequently, when assessing the impact of trade liberalization, it is important to examine the impacts on income distribution, and to consider what mitigating policies can be put in place to soften the blows to the losers. These effects can be particularly critical for a country that lacks social insurance mechanisms to help those who lose their jobs or whose incomes suffer major reductions during the adjustment to freer trade.
To keep things in perspective, however, it is important to note that trade is only one of many factors that affects the distribution of income. For example, a number of studies on the evolution of income distribution over the past 30 years in the United States have concluded that technological change has been the main driver of increased inequality, dwarfing the effects of increased trade. Although improved technology generates gains to the economy as a whole, the benefits are not evenly spread, and some people lose. A dynamic and growing economy will constantly be subject to innovations, shocks, and entry and exit of firms. Trade is one of many sources of such change.

*Trade in inputs*

The argument for gains from trade applies both to goods and services destined directly for final consumers, as well as to those which are used as an input into production. At one level the issues are very much the same - if goods can be imported at a lower opportunity cost than if they are produced locally, then the country's real consumption possibilities increase. However, there is an important added twist. Imports of producer services can lower costs of firms in the export sector and thereby stimulate exports. An example would be a country with untapped oil reserves. Access to foreign engineering services could lower the cost of developing oil extraction facilities which could then lead to increased exports of oil. That is, the benefits of allowing increased imports of producer services can be potentially magnified via their effects on improving productivity in other sectors of the economy.

*B. Increasing returns to scale*

While differences between countries is one of the major explanations for trade, particularly between countries with very different income levels, it cannot account for all trade. Much of the world's trade occurs between high income countries, suggesting that
similarities between countries need not deter trade. In fact, there is some evidence that the more similar are countries, the greater is the volume of trade between them. Moreover, much of the trade between similar countries is in similar products. For example, Canadian engineers work on projects in the US, and US engineers work on projects in Canada.

There are several theories which explain why trade occurs between similar countries. Although these explanations were stimulated by a need to explain the large amount of trade between similar countries, these same motives for trade will also be present when countries are different. Hence these additional explanations for trade complement and interact with the standard comparative advantage approach that we have already discussed.

One way to think about how trade can emerge between similar countries is to consider a labor market example. Think of two students starting university who are equally bright and talented. At this point their productive capacities may seem indistinguishable. However, suppose one chooses to study medicine and the other chooses to study engineering. If we revisit these same students 10 years later, they will have very different skills and they can trade with each other via the labor market - the doctor selling medical services, and the engineer selling engineering services.

This example highlights three key processes. First, there are gains from specialization. However, specialization did not initially occur because of comparative advantage. The students were assumed to be initially identical so there was no comparative advantage at the outset. Comparative advantage evolved over time because of the opportunity to trade. That is, each student made investments in education to develop his or her productive capacity in a particular field. This brings us to the second key aspect of this example, which is fixed costs. Instead of trading with each other, the two students could have each decided to spend part of their time being an engineer and part of their time being a doctor. However, because there are large up-front investments in education and experience required to be good in each of these fields, the students can avoid paying these fixed costs twice if they each specialize in different fields. Finally,
because the fixed cost is an up-front investment, there is an element of lock-in to this example. At the beginning of university education, each student could have gone down a different path. But once the investments in education and experience have been made, it is very costly to switch fields. History matters, and decisions made in the past (including past government policies) have a large influence on current patterns of trade and apparent comparative advantage.

Let us now consider how increasing returns to scale can explain trade between countries. There are at least four different ways that scale effects can generate trade: market niche effect, the development of firm-specific intangible assets, agglomeration, and networks.

**Market niche effect**

Let us come back to our engineering example. Why would engineering services be both exported and imported by the same country? First, if product variety is valued either by producers purchasing services as inputs or by final consumers, then firms have an incentive to carve out their market niche and produce a specialized variety of a good or service. If there are fixed costs to establishing a market niche or developing a new variety, then larger markets will have more product variety.

Examples of product variety in the service sector include entertainment (movies, television programs, music), tourism (consumers gain a wider choice of destinations), restaurants (cities which are open to a lot of foreign visitors can support a wider variety of restaurants), architecture and engineering services (a larger market allows firms or individual producers in these sectors to specialize in different types of projects), and many others. In fact since many types of services are tailored to individual customer needs, product variety is a major aspect of service trade.

The product variety or market niche motives for trade is important for producer services as well as for final consumer services. A larger market can support a wider variety of specialized producer services, which can lead to increases in productivity for
producers of both final goods and services. As noted above, imports of producer services can lead to increased exports in other sectors.

Gains from trade and effects on income distribution

Trade driven by the market niche effect will potentially generate three types of gains. First, the total variety of products available to consumers in any given market will increase because consumers gain access to both domestic and foreign varieties of the products. Second, each individual producer will have access to a larger market, and this can allow them to expand their output and reduce costs due to scale economies. Third, a specialized service that might not be economically viable in a small country might become viable as a result of the market-expansion effects of trade. That is, a larger market will allow the development of new goods and services that might not have otherwise been available.

The income distributional effects of trade driven by the market niche effect are likely to be less significant than when trade is driven by differences between countries. Because product variety is valued, and because trade can increase both imports and exports within the same sector, it provides both increased opportunities at the same time as it increases competition in any given sector. As we will see later, it is possible that trade may squeeze out some specialized local services; we will discuss this possibility when we consider trade policy. It should also be noted that the ability to create market niche services may itself be a source of comparative advantage. That is, the market niche motive for trade will interact with country differences, in which case our earlier discussion of the distributional effects of trade liberalization will be relevant here as well.

Firm-specific intangible assets

Consider yet another example of service trade. Large European and American financial institutions and insurance companies set up foreign affiliates to provide services
to local customers in foreign countries. Foreign-owned firms may also operate local electric power plants or retail outlets. Some types of direct foreign investment can be explained either by differences between countries (for example in the costs of capital) or by the market niche effect (such as MacDonald's fast food franchises). However, the market niche affect alone cannot explain why a foreign firm might be better able to more profitably operate a retail store or a power plant.

Many firms exist because they have developed specialized firm-specific assets, such as specialized knowledge of organizational and production processes, distribution and supply networks, and reputations for quality and reliability. This can explain the success of large firms in many different industries, such as financial institutions, construction firms, and courier companies.

Successful large firms can potentially provide services to foreign markets via each of the four modes. However the notion of intangible firm-specific assets is particularly helpful in explaining foreign direct investment. It is useful to think of two types of fixed costs: firm-specific fixed costs, and plant-level fixed costs. Once the investment in firm-specific assets is made, the firm's knowledge and reputation can then be exploited in each of the plants that a firm sets up (see for example, Markusen and Venables 2000). This can create gains from foreign direct investment. By setting up branch plants in various countries, the firm-level fixed costs need only be paid once, and each plant can be set up by paying only plant-level fixed costs. This creates gains to the host country as consumers gain access to a wider variety of specialized services at lower prices than otherwise would be available.

Agglomeration effects

Increasing returns to scale can also lead to agglomeration. Agglomeration can take two forms. A particular industry can concentrate in one area - such as the concentration of financial services in cities like New York and London, and the concentration of the computer industry in Silicon Valley. Or there can be a general
concentration of a wider variety of economic activity in cities, regions, or countries. That is, trade can lead to the emergence of "cores" and "peripheries".

Agglomeration can be explained in a couple of ways. One explanation is that there are spillover effects across firms. When a large number of firms locate close to each other, there can be positive externalities. There can be knowledge spillovers, as firms learn from each other. Firms can benefit from access to a common pool of specialized labor. And infrastructure can be set up to address the needs of a specialized industry.

Another explanation is that agglomeration arises from the interaction between scale economies and transportation costs or trade barriers. Suppose producers require intermediate goods and services, and that production costs fall on average when there is access to a wider variety of intermediate goods and services. Then if it is costly to trade intermediate goods and services either because of transport costs or trade barriers, market forces will tend to concentrate production in one place. Why? First, there is an advantage to being in a larger market because it can support a wider variety of specialized producers of goods and services. Second, trading costs segment markets. If it is costly to trade, then there is a cost advantage to being close to all the other producers of intermediate goods and services. But the more producers choose to locate in one market, the larger the demand for intermediate goods and services, and so the larger the effective market size is. Agglomeration can be self-reinforcing.

Gains and distributional effects of agglomeration-driven trade

The welfare effects of trade in the presence of agglomeration are complicated. Those who live in core gain from trade because the concentration of economic activity in their area leads to benefits from both scale economies and access to a wide variety of goods and services. However, those who live in the periphery may lose from trade if specialized goods and service production migrate to the core.
Whether or not there are gains or losses from trade depend on the magnitude of trading costs. To investigate this issue, Krugman and Venables (1995) consider a simple agglomeration model with two identical countries. There are two sectors, manufacturing and agriculture. Agriculture has constant returns to scale, while manufacturing productivity rises if firms have access to specialized intermediate goods and services. If there is no trade, then both countries are equally well off. As trading costs fall, then initially both countries gain because of access to increased variety of goods and services. However, at some point, the fall in trading costs results in one of the two countries becoming the core. This occurs when trading costs are low enough to make export of final goods cost-effective, but still high enough to create incentives for intermediate goods and services producers to locate close to final good producers. Once the core / periphery production pattern emerges, the periphery becomes worse off because the demand for labor drops, pushing down real wages.

However, as trading costs continue to fall, eventually, the low wages in the periphery lure back some of the manufacturing production, and real income in the periphery starts to rise. Finally, if all trading costs are completely eliminated, distance and location do not matter any more, and the distinction between the core and periphery becomes irrelevant.

This example suggests that while completely free trade could potentially benefit all countries, partial trade liberalization could hurt the periphery. It also suggests that the core does not have an incentive to eliminate all barriers to trade, because then the advantage of being in the core is reduced.

*Networks*

Another scale-related motive for trade arises from access to networks. This motive for trade is at root driven by economies of scale, but the scale economies are not specific to an individual firm or even a given country. In many sectors, such as telecommunications, shipping, financial services, transportation, the efficiency, quality
and benefits to consumers of the services provided depend on access to networks of other consumers and producers. For example, a phone system is only useful if many other people have phones; email became a standard form of communication after a critical mass of people had access to it; and the value of having a debit or credit card increases with the number of places that have access to a network that accepts the card. In such cases, there are gains from trade which arise from increased market size. If two countries of similar size each had their own internet network then establishing a connection between the two networks would roughly double the number of sites that can be accessed, and double the number of people who can be contact by email. This scale effect generates gains from trade.

In the case of shipping and transportation, networks develop within countries but regulatory barriers can limit the connection points between domestic and foreign networks. Allowing smooth connection and integration of systems can also generate gains from trade. For example, if three neighboring countries have customers who want to send packages to Peru, each country could have a airplane that carries packages to Peru, or they could all send their packages to a hub in one of the three countries, from which a single airplane would carry the cargo to Peru. Integration of the shipping network generates economies of scale which in turn can generate gains from trade.

The analysis of the benefits of liberalizing trade rules to allow smoother access to international networks is complicated by two important factors. First, because of the economies of scale that arise from having a large integrated network, some networks can come to be dominated by a small number of large firms. There can be therefore a trade-off between the advantages of economies of scale and the costs of the concentration of power. Second, in part because of the issues of concentration of power, many networks are highly regulated. Networks are publicly owned in some countries, run by monopolies on other countries, and subject to various entry and regulatory constraints in other countries. These differences in regulations across countries lead to barriers to trade which can be difficult to circumvent. These types of policy issues will be discussed later.
Box 1: Dynamic benefits of services trade liberalization

It is not easy to model dynamic gains from trade formally, but there are strong intuitive reasons to believe that well functioning service industries contribute to growth in different ways. An efficient financial sector leads to an efficient transformation of savings to investment, ensuring that resources are deployed where they have the highest returns; increased product variety associated with an efficient financial sector also leads to improved quality of consumer services and better risk-sharing in the economy. In the case of telecommunications, improved efficiency generates economy-wide benefits as telecommunications are a vital intermediate input and are also crucial to the dissemination and diffusion of knowledge—the spread of the internet and the dynamism that that has lent to economies around the world is telling testimony to the importance of telecommunications services. Transport services facilitate both the efficient distribution of goods within a country and a country’s participation in global trade, and hence generate benefits in terms of learning and spillovers in knowledge. Although these are the more prominent services, others are also crucial—business services such as accounting and legal services are important in reducing transaction costs. According to Lawrence Summers (1999), the single most important innovation in the history of the American capital markets was the idea of generally accepted accounting principles. Software development is the foundation of the modern knowledge-based economy. Education and health services are necessary in building up the stock of human capital, a key ingredient in long run growth performance.

The growth effects of services liberalization also arise from allowing movement of factors of production. A country that liberalizes its services sector is likely to augment its stock of capital (through increased FDI) and crucially the stock of human capital and technology that is embodied in or associated with such FDI. The impact of this on long run growth is unambiguously positive. Furthermore, there is evidence that the presence of foreign factors can help enhance the productivity of domestic resources. This is as true for developing country capital importers as for developed country importers of skilled labor services. The contribution of imported skilled labor to the high-technology sectors in the US is now widely recognized.

IV. Modes of Supply

We have reviewed the major different explanations for the pattern of trade. These explanations provide a motive for trade. But they do not always explain why firms choose particular modes to supply services to foreign customers.

The possibility that services may be supplied by more than one mode raises a number of issues. First, if trade in all modes was unconstrained, how would firms choose to supply services to their foreign customers? Second, are different modes substitutes or
complements? Third, what is the effect of allowing trade via some modes but not others? The answers to these questions have important implications both for predicting and assessing the implications of trade liberalization in services. Unfortunately, this issue has not received a great deal of study in the academic literature.

\textit{A. Asymmetric costs across modes of supply}

For many types of services, the costs of provision vary substantially across the different modes of supply, and for some services, supply is essentially not feasible via some modes.

Tourism, for example, cannot take place unless consumers free to travel and spend their money in foreign countries. Fast food restaurant services cannot be provided without a commercial presence in the country where the food is served; hence this requires foreign direct investment or provision for franchising agreements. Construction services requires a physical presence as well, so construction workers have to be able to move across borders if their services are to be tradable.

This heterogeneity across services is particularly important because countries have comparative advantages in different types of services. Rules that allow some modes but not others will favor some countries over others. A set of rules that allows direct foreign investment, direct cross-border export of services, and movement of customers between countries, but which does not allow labor movement across countries will preclude certain types of services from being traded at all. The services that cannot be traded in such a regime are labor-intensive services that require physical contact between the customer and service provider. Hence countries which have a comparative advantage in such services (labor abundant countries) will be seriously constrained in their ability to export these types of services. Although there are many potential gains from importing services (especially via increased productivity in goods production), the gains from trade will generally be larger when foreign barriers to exporting services are removed as well. Consequently, the rules affecting different modes of supply can be critically important.
both in determining which services will be traded, and in determining the distribution of
the gains from trade across countries.

B. Modes of supply as substitutes

For some types of services, different modes of supply are substitutes, in the sense
that if one mode is not available, firms will use a different mode to supply their service to
foreign customers. For example, consider an insurance firm. A foreign customer could
buy an insurance contract by mail or electronically. That is, the insurance services could
be directly exported. Alternatively, the insurance company could set up a branch office
in the foreign country to serve its foreign customers. In that case, the foreign client
would deal with the local office of the foreign insurance company. When it comes time
to settle claims, the insurance company could send agents from the home office to assess
the claim; that is, it could rely on temporary movements of personnel for assessments of
claims. Or, if it set up a branch office, then foreign personnel from the foreign office
could assess the claims. Another example is medical services. A specialized surgical
team could come from a foreign country to perform surgery in the home country; or the
patient could go to the foreign country to receive treatment.

If the different modes of supply are substitutes, then in some cases virtually all of
the gains from trade can be realized by opening up just one mode of trade. This is the
basis of Robert Mundell's (1957) observation that under some conditions, trade in goods
and services and trade in factors are perfect substitutes. That is, a country with an
abundance of skilled workers can either export goods and services directly produced by
these workers; or the workers themselves can move to produce in other countries. In both
cases, gains from trade will be realized.

However, even when modes of supply are substitutes, restrictions on which
modes are available to firms can have important implications. First, modes are typically
not perfect substitutes. In the example of the surgeon above, hospital facilities will differ
across countries and so the success rate and cost of the surgery will differ depending on
where it is performed. Travel costs for patients and surgeons will differ. The health of a seriously ill patient may be jeopardized by travel; and the opportunity costs of travel for a surgeon may be high because of the needs of other patients. In the case of the insurance company, the reliance on a foreign office may increase managerial overhead expenses; but the presence of a local office might allow the firm to better deal with the needs of local clients. If firms are free to choose their modes of supply, one would expect that they would choose the most cost-effective mix. Restrictions on access to foreign markets via some modes but not others can therefore lead to increased costs of provision of the services and therefore potential reduce the gains from trade.

Second, different models of supply will have different effects on income distribution. For example, consider the insurance example again. Suppose a country (North) has a comparative advantage in the provision of insurance services. Suppose also that direct foreign investment in insurance services is liberalized but insurance agents are not allowed temporary access to foreign countries to assess claims and serve the local needs of their clients. In this case, insurance firms from North are likely to set up branch offices in foreign countries to serve foreign clients. This will increase the demand for labor in the insurance sector in foreign countries and may push up wages. On the other hand if direct foreign investment is not allowed, and Northern insurance personnel are allowed temporary access to foreign countries, this can reduce the demand for labor in the insurance sector in foreign countries and may push down wages.

The effects on income distribution can be more subtle and complicated than this example indicates. Complementarities may also occur across similar types of factors. For example, if a large foreign engineering firm has a contract to build a bridge, then allowing temporary access for foreign engineers can increase the productivity of local engineers who also work on the project. That is, allowing temporary access for foreign engineers may increase the incomes of local engineers. As another example, allowing foreign movie stars into a country to make a film can increase the demand for local actors who will work on the film as well.
More generally, allowing direct foreign investment will have different effects on wages and employment than will policies which allow labor to move across countries. Even if the efficiency effects of the two different modes of supply is not very different, the effects on income distribution may be important in determining political support for the different modes.

In models with agglomeration effects, the choice between direct foreign investment and direct exports of services can affect the incentives to agglomerate. If direct trade is the only mode permitted, this tends to favor the country with a larger market because its larger size allows it to take advantage of the agglomeration effects. On the other hand, if direct trade is restricted, and governments require a local presence as a condition for local provision of the service, then the smaller market is favored, because such a policy will encourage direct foreign investment, which can increase the demand for local producer services, and thereby transfer some of the agglomeration benefits from the large market to the smaller one.¹

C. Modes of supply as complements

In many cases the different modes of supply will be complementary. For example, if a firm chooses to have a physical presence in a foreign market, then the effectiveness of their operation may be enhanced if personnel are allowed to move between the home and foreign establishments. As well, there will likely be direct intra-firm trade in services, with for example, research services concentrated in one location, accounting services in others, etc. A policy which restricted any of these three modes would affect either the cost and quality of the service provision. In cases where there are strong complementarities across different modes of supply, fully effective liberalization of service trade requires that all modes be opened up.

¹ See Markusen and Venables (2000) who show how allowing multinationals to set up in smaller countries can erode some of the larger country's advantages.
Services may also be complementary with goods trade. Export of goods requires transportation and insurance services, but it may also require the establishment of distribution networks, facilities to deal with repairs or with training of customers in the use of products, etc. Hence in some cases the potential gains from goods trade cannot be fully realized without liberalization of service trade.

**Box 4: Impact of restrictions and substitutability between modes**

How much a restriction on a particular mode matters depends on the scope for substitution between modes. At one extreme, if a service (e.g. construction) can only be supplied through one mode (movement of builders), then a prohibitive trade restriction on that mode shuts out foreign supply completely. At the other extreme, if modes (say cross border and movement of individuals) are substitutes for the supply of a service (e.g. standardized software development), then the impact of a restriction on one mode is diluted.

What determines substitutability between modes?

*Technology.* A few decades ago back-office services had to be performed quite literally in the back of the office. Today, the development of information and communications technology has made it possible for Swiss Air to have its book keeping services performed in Bombay. Even though a surgical operation still requires proximity between the patient and the surgeon, it is conceivable that in a few decades, with the development of remote controlled robots, it will not.

*Consumer preferences.* It is possible to conduct almost the entire range of financial services electronically. But while most consumers are happy to borrow long-distance, when it comes to depositing money, many still derive reassurance from the concreteness of local banks. Similarly, while an increasingly large number of consumers buy property insurance long distance, life insurance remains for the most part a face-to-face business.

*Regulatory environment.* Sometimes market failures, such as those arising from asymmetric information, may be more difficult to remedy when the supplier is located abroad. For instance, it may be easier to establish that a doctor or lawyer established locally is adequately trained than one located in another country. Some financial services, such as borrowing from abroad, necessarily require capital mobility and governments are sometimes reluctant to allow this for macroeconomic reasons. In the case of regulatory restrictions, however, it is important to distinguish between those which are the best instruments to fulfil legitimate objectives and those which serve a protectionist purpose. This is an issue which we consider more fully later.
V. Trade Policy and the Service Sector: The instruments of protection

A. Overview

In this section, we analyze government policies that affect international trade in services. Although the analysis of barriers to service trade has much in common with the analysis of barriers to goods trade, trade policy in the service sector is much more complex because of the scope of and types of government regulations that inhibit trade in services.

For goods trade, analysts typically distinguish between tariff and non-tariff barriers to trade. Tariffs are discriminatory taxes on trade. An import tax is a tax levied on foreign goods but not domestic goods. Tariffs tend to be easy to measure and are very transparent. It is therefore quite straightforward to design an agreement to liberalize trade via tariff reduction (provided governments have the will to open up their economies to trade). Successive rounds of GATT negotiations were successful in achieving broad based reductions in trade barriers via across-the-board reductions in trade taxes. Another feature of tariffs is that it is fairly clear what is meant by free trade: zero tariffs. This means there is a clear focal point for negotiations, and this has been exploited in numerous regional free trade agreements such as NAFTA, which have eliminated substantially all taxes on trade flows.

Non-tariff barriers, on the other hand, are much more complex. A non-tariff barrier is any government policy that has the effect of favoring local producers over foreign producers or which restricts or raises the cost of access to domestic markets by foreigners. These can include delays at the border, quantitative restrictions on foreign products, government purchasing policies that give preference to local suppliers, subsidies, quality and certification requirements that favour local suppliers, etc. Non-tariff barriers are more difficult to measure and are less transparent than tariff barriers. Moreover, there is often not an obvious focal point for negotiations because the trade-
distorting effects of many policies are intertwined with other government policy objectives, such as protecting health and safety, encouraging regional development, etc.

Most trade barriers in the service sector are non-tariff barriers. Tariffs (discriminatory taxes imposed on foreign service providers) are relatively uncommon. There are several reasons for this. First, cross-border trade in services is often in intangible form, and this makes it difficult to monitor and tax. Second, the modes of supply for many services are different than for goods. Many services are exclusively delivered via commercial presence or via temporary labour movement. Barriers to foreign direct investment and international labour mobility can therefore result in serious restrictions on service trade and obviate the need for additional discriminatory taxes. Finally, many services are highly regulated or are provided by the public sector. Regulations that either intentionally or unintentionally make it relatively more costly for foreign suppliers to operate are major sources of trade barriers in services.

The pervasiveness of non-tariff barriers in the service sector means that trade liberalization in this sector is complex. Moreover, a major reason for the pervasiveness of non-tariff barriers is because of market imperfection in service sectors. Many trade barriers in the service sector are a side effect of domestic regulations that have legitimate purposes. For example, because of issues in asymmetric information, doctors must be certified to protect patients, engineers need certification to ensure that bridges they build do not collapse, and insurance companies have to be regulated to ensure their solvency. However, these same rules can be manipulated to protect local suppliers. For example, a rule requiring that an engineer graduate from a domestic university might ensure that quality standards are met, but would prevent consumers from having access to the services of highly qualified foreign engineers. The regulatory apparatus may therefore serve the dual purpose of responding to market failures and protecting local suppliers at the expense of consumers. A challenge for trade policy analysis is to isolate the protective effect of regulatory policy from the beneficial effects, and to suggest rules for liberalization that provide the benefits of increased trade while ensuring that other
legitimate policy objectives are achieved. In many cases, trade liberalization may not be possible or viable unless it is accompanied by domestic regulatory reform.

In what follows, we analyze a few of the common types of trade barriers that affect the service sector. However, because trade protection is often inextricably linked with domestic regulation, our analysis is not exhaustive. The trade-related implications of regulations will depend on the special characteristics of the service industry in question, and the types of market failures the regulations are designed to correct. Some of the later modules will focus on trade barriers in specific service sectors. Our objective here is to highlight some general principles. Perhaps the most important of these is that the analysis of regulation-induced trade barriers must take place in a framework where the market failures that led to regulations in the first place must be explicitly considered.

We will focus on the following types of trade barriers:

**Tariffs** as noted above are relatively uncommon in the service sector. However, we begin by reviewing the standard analysis of the costs of tariff protection because it serves as a useful benchmark to compare with other forms of protection.

**Discriminatory regulations** which add to the cost of trading services, but which do not yield any direct benefits to local consumers, are a very common form of trade barrier. Examples include delays in crossing the border, country-specific standards for trucks (such as differing weight and trailer length regulations) which add to the cost of cross-border transport services, preferential government procurement policies, and lack of transparency of domestic regulations.

**Licensing and certification requirements** inhibit trade in professional services. Doctors, engineers, architects, lawyers, accountants, and other service providers typically need to satisfy local regulations for certification. In some cases, compliance may be very costly (a domestic residence or graduating from a domestic educational institution may be required). These types of regulations can be justified by the need to protect consumers by
ensuring quality and safety standards are met. However, they can also protect local service providers from foreign competition, which can lead to higher prices and reduced choice for local consumers.

**Quotas** are pervasive. On cross-border trade, they are common in the transport sectors. Foreign providers are either completely shut out (i.e. a zero quota) of certain segments, such as transport within a country; or only provided limited access, as in international transport. On consumption abroad, quotas are sometimes implemented through foreign exchange restrictions; e.g. the ability of citizens to consume services, such as tourism and education, abroad is limited by limits on foreign exchange entitlements. On commercial presence, quotas are imposed on the number of foreign suppliers who are allowed to establish in sectors like telecommunications and banking. Quotas on foreign participation also take the form of restrictions on foreign equity ownership in individual enterprises. Finally, quotas are perhaps most stringent in the case of movement of service-providing personnel, and affect trade not only in professional services, but also in a variety of labor-intensive services.

**B. Tariffs**

We begin by reviewing the analysis of a tariff to use as a the benchmark. For those familiar with the analysis of trade barriers on goods, the analysis of the welfare effect of a tariff on services is the same as that for services.

Figure 2 illustrates the welfare effects of a tariff for a small country facing a fixed world price $p$ for the imported service. The domestic demand for the service is denoted by $D$, and domestic supply is $S$. In free trade, domestic output is $x_o$ and domestic consumption is $c_o$. A tariff $t$ on imports raises the domestic price to $p^d = p + t$, which raises output to $x_1$, but lowers consumption to $c_1$. 
Figure 2.
The tariff redistributes income within the economy and also generates inefficiencies. Producers gain from higher prices, consumers lose, and the government collects tax revenue. Using a cost benefit approach, we can measure the effects on each group by referring to areas in the diagram:

- Rise in Producer Surplus: $A$
- Loss in Consumer Surplus: $A + B + C + D$
- Gain in Tariff Revenue: $C$
- Net social Loss: $B + D$

The social loss is a measure of the inefficiency induced by the tariff; it is the cost of foregone trade. To understand why there is a social loss, first note that part of the loss suffered by consumers (due to higher prices) is offset by the gains to producers (area A) and the government's revenue (area C). However areas B and D do not accrue to anyone in the economy and hence represent a real loss to the economy.

Area B represents the production distortion induced by the trade barrier. The tariff increases output to $x_1$, but at this point the opportunity cost of $X$ production is $p + t$ (the supply curve measures marginal production costs). If instead $X$ is imported, the opportunity cost to the economy of acquiring a unit of $X$ is $p$, which is lower than the marginal production cost. The trade barrier therefore induces excessive production of $X$, and the area B represents this cost.

Similarly, area D represents the consumption distortion induced by the trade barrier. The demand curve represents the marginal benefit to the economy of consuming $X$. And at the tariff-induced consumption level, this marginal benefit is above the opportunity cost of acquiring the good from foreigners ($p$). This means that consumption is too low, and the cost of under-consumption is represented by area D.
C. Restrictions that do not generate revenue

A wide variety of trade barriers are restriction on foreigners that raise the costs of serving the domestic market but which do not generate any revenue for the government. This can involve bureaucratic measures that generate delays, restrictions on travel, requirements, inspections, etc.

These types of policies can also be assessed with the aid of figure 2. Suppose that restrictive measures increase foreign costs by $t$ per unit but do not generate revenue. Then domestic price rises to $p^d$, output rises to $x_1$, and consumption falls to $c_1$. So far the policy has effects that look like a tariff. But there is no tariff revenue and so Area C is a deadweight loss in the sense that it imposes costs on the foreign supplier (and which contribute to a domestic price increase) without generating revenue for the importing country. Hence the total social cost of the policy is $B+C+D$. So the loss in welfare is much greater than for a tariff.

D. Quotas

Many types of quantitative restrictions apply to service trade. For example, many countries, such as Canada and Australia, have restrictions that require that foreign content in broadcasting (such as popular music or television programs) not exceed some limit. There are also many types of licensing restrictions that put explicit quantitative limits on the number of foreign suppliers who are permitted to compete in a local market.

To analyze the effects of a quota, consider a quota that requires that imports do not exceed $M_1$. We can represent this in Figure 3. The curve RD is the residual demand. It is the amount of domestic demand for $X$ left over after the import quota has been filled. Graphically, it is obtained by shifting the demand curve D to the left by the amount of the quota $M_1$. The domestic price is determined by the intersection of the residual demand curve and the domestic supply curve, which yields a price $p^d$. This is the same price that
Figure 3. Effects of Import Quota
was obtained above when a tariff was imposed. As well, output rises to $x_1$, and consumption falls to $c_1$.

A tariff and quota therefore have similar effects. If markets are competitive, as they are in the example so far, producer surplus will rise by the same amount in both cases, and consumer surplus will fall by the same amount. But there are some differences.

The first difference is obviously that a tariff generates tax revenue, while a quota does not. The area $C+D$ in Figure 3 is called the *quota rents*. Since $X$ can be produced in foreign markets at a price $p$, and sold domestically at the higher price $p^d$, the gap in prices generates a surplus, which are referred to as the quota rents. The quota rents accrue to the individuals who have the right to import services into the domestic economy.

If the government auctioned off the rights for foreigners to provide $M_1$ units of services to the economy, then the revenue generated would be area $C+D$ (assuming an efficient auction) and the government would collect rents equal to the revenue it would have collected from tariffs. However, such auctions typically do not occur.

In the case of quotas on goods imports, the quota rents typically accrue to domestic agents who have import licenses. In this case area $C+D$ represents a redistribution of consumer surplus to importers. If domestic importers acquire the import licenses costlessly, then a cost / benefit approach treats area $C+D$ as a benefit and so the social costs of the quota are areas $B+E$, which is the same as a tariff.

In practice, however, quota rents are often dissipated by rent seeking activities. That is, because those who obtain import licenses can earn quota rents, there is an incentive to spend resources on lobbying and other activities (some possibly involving corruption) to try to acquire import licenses. Since these activities consume real resources and are unproductive, they are a cost, and tend to push the social costs of quotas above tariffs.

There is an even more compelling reason to expect that the quota rents will be lost. Because services are provided directly by foreign providers, they are typically not imported by middlemen, as in the case of goods. That is, those foreign service providers
who are allowed to operate in the domestic economy will sell their services directly to domestic consumers at a price $p^d$. Hence the foreign service providers will typically collect the quota rents. In this case, (unless the government auctions off quota licenses for the right to provide services), area C represents a loss in consumer surplus that accrues to foreigners and so counts as a loss to the domestic economy.

Consequently, in the service sector, the social cost of imposing a quota is likely to be area B+C+D+E, which is higher than for tariffs.

Quotas fare even worse when markets are not competitive, as was pointed out by Bhagwati (1965). To see this, suppose there is a single domestic firm. In the absence of international trade, the domestic firm would act as a monopolist and unless it is regulated, would charge a price above marginal cost. However, if there is free trade, then the domestic firm cannot charge a price above the import price $p$. That is, free trade eliminates the domestic monopolist's power.

The key difference between a tariff and quota is that in a tariff regime, the domestic monopolist faces potential foreign competition on every unit of X that it sells. But under a quota, there is competition only for the first $M_1$ units imported. Once the quota is filled, the monopolist faces no more competition. That is, an import quota gives the monopolist power that it does not have in a tariff regime. This means that quotas are significantly inferior to tariffs in markets where there are not very many local firms.

The other major benefit of tariffs over quotas is the added transparency of tariffs. Consumers may not be aware of the extent to which quotas are pushing up domestic costs. However, if they have to pay a tax on foreign services, then the magnitude of the tax is a useful index of the stringency of protection. Moreover tariffs can simplify negotiations over trade liberalization, in part because of the added transparency. An agreement to reduce tariffs by 30% is easier to implement that one that requires the import quotas be relaxed.

Are there any cases where a quota might be better than a tariff? If the government wants to completely exclude foreigners then an import ban (a prohibitive quota) and a prohibitively high tariff will have the same effect, except that the ban will be more certain
in its effects. Also, producers often prefer quotas. Quotas help to insulate the domestic market from price changes occurring in foreign markets. This can reduce uncertainty for domestic producers. However, as discussed above, this added benefit for producers can come at a high cost to consumers.

**Box 3: Fiscal Instruments: Is the Difficulty of Substitution Between Policy Instruments Overstated?**

The difficulty of switching to fiscal instruments of protection in services has probably been exaggerated. As far as cross-border trade is concerned, the imposition of duties is probably most difficult - perhaps impossible, given the current state of technology – when a service is delivered electronically. But in this case, imposing other barriers to trade are also likely to be infeasible. Where quotas are feasible and maintained, as on cross-border trade in transport services, it is easy to conceive of tariff-type instruments: e.g. a tax per passenger or unit of cargo carried by a foreign company. Moreover, the auction of a quota is analogous in economic effect to the imposition of a tariff.

In the case of commercial presence, a number of fiscal instruments are possible, including entry taxes (or auctions of entry licenses), output taxes and profit taxes. An output tax on foreign suppliers increases their marginal cost of providing a service and is similar in effect to a specific tariff. An entry tax increases the fixed costs of firms and their willingness to enter the market: the market structure is therefore likely to be less competitive than in the absence of such a tax. A profit tax is least likely to affect the economic decisions of firms. But if there are any fixed costs of entry which must be covered by future profits, then a profit tax would reduce the number of firms that could recover their costs of entry. One or more of these fiscal instruments could help achieve outcomes superior to quotas from a social welfare point of view.

Ironically, the legal systems of many countries allow discrimination against foreigners through outright bans and entry quotas but make it difficult to impose discriminatory taxes. For instance, in the European Union, a locally established foreign firm is treated in all respects like a European firm and cannot be subject to any form of discrimination.

**E. Restrictions on foreign direct investment in competitive markets**

If markets are competitive, the benefits of foreign direct investment are similar to the standard gains from trade. Figure 4 illustrates a case where a service (X) can be provided only via commercial presence. Domestic supply is \( S^D \) and demand is D. In the absence of foreign investment, output is \( x_o \) and price is \( p_o \). If direct foreign investment is allowed, the supply of service providers will shift out. We denote the expanded supply as the sum of supply from domestic and foreign firms: \( S^D + S^F \). With the foreign presence in the local market, price falls to \( p_1 \) and output increases to \( x_1 \). There is a gain in consumer surplus (A+B+C), and a loss in producer surplus (A), which yields a net social gain of
Figure 4. Foreign direct investment: competitive market
B+C. Foreign direct investment increases welfare by increasing competition, lowering prices and increasing consumer choice. Conversely, restricting foreign investment would shift in the foreign supply curve and lead to social costs as the area B+C is eroded.

There are also many other potential benefits from direct foreign investment - through joint ventures, local firms may gain access to improved technology and financing. For some services such as insurance and finance, there can be increased risk pooling. There may also be spillover effects: local workers employed in foreign owned firms may receive knowledge, experience and training that might otherwise be unattainable. Taking these effects into account would increase the measure of social costs of restricting foreign direct investment.

**F. Trade restrictions vs. subsidies**

The relative merits of different instruments of protection depends in part on why the trade barrier is there in the first place. In many cases, the purpose of the trade barrier is to increase local output in certain sectors. It may be seen as desirable to have a more domestic engineers, or to have a domestic airline, or a domestic banking sector. Restrictions on foreign service providers give domestic providers an advantage and so their output expands.

However, a domestic production subsidy will also increase the output of local producers, and will do so without increasing consumption distortions. To see this, refer to Figure 2 again. Suppose the government's objective is to increase the output of local service providers to $x_1$, which is above the free trade level of $x_0$. We saw above that a tariff or quota will do this. But suppose instead the government provides a subsidy $s$ to producers for each unit that they sell. Moreover, suppose the subsidy $s$ is set equal to $t$ in the figure. Then producers get $p+s = p+t$ for each unit that they sell and so output rises to $x_1$ as desired. Consumers still get to buy at world prices, and so consumption stays at $c_0$. 
The subsidy has to be financed. Total subsidy payments are represented by the amount A+B in the diagram. This will have to come out of tax revenue.

Let us now do a cost / benefit analysis of the subsidy.

Rise in Producer Surplus: A
Loss in Consumer Surplus: none
Subsidy financing cost: A+B
Net social *Loss*: B

The subsidy creates a production distortion measured by area B (which is unavoidable because the purpose of the policy is to raise output), but does not generate a consumption distortion. Consequently, the subsidy is more socially efficient than either a tariff or an import quota. Intuitively, the problem with trade barriers as a means to raise local output is that they also lower local consumption. A subsidy can raise output without distorting consumption and so generates lower social costs, even when we take into account the fact that the subsidy has to be financed.

A couple of factors can weaken the case for subsidies. First, a subsidy needs to be financed, which means that taxes must be raised. The implementation of taxes can create other distortions in the economy which have to be weighed against the distortions caused by tariffs and quotas. However, it should be noted that a tariff on X is essentially a production subsidy for X producers financed by a consumption tax on X consumers. It is highly unlikely that a tax on X consumers is the optimal way to raise revenue.

Second, a more significant problem with subsidies is that money has to be paid out to domestic producers, which can increase administrative costs and open up opportunities for corruption. If subsidies get diverted away from their intended purpose, the social costs of the policy would be higher and could possibly exceed the costs of tariffs.
VI. Interaction between trade policy and domestic regulation

Our analysis of trade policy so far has been based on the premise that markets work well and that trade barriers are in place only to protect local producers. However, many service industries are in imperfect markets. For example financial services, electric power, telecommunications, and air transport are all sectors where there are large firms which have market power. This market power often arises from barriers to entry arising from domestic regulations. Other service industries, especially in the professional service sector, are plagued with problems of asymmetric information - the service provider knows more than the consumer about the quality of service dispensed and the appropriate level of care needed.

In this section, we consider several examples to illustrate the interaction between trade policy reform and domestic regulation.

A. Barriers to entry and the impact of FDI restrictions on welfare

Restrictions on foreign direct investment assume particular significance in the case of services where cross-border delivery is not possible, so that consumer prices depend completely on the domestic market structure. In many service sectors, such as in communications and financial services, there are restrictions both on entry and on foreign ownership. A basic conclusion from the literature on privatization is that larger welfare gains arise from an increase in competition than from simply a change in ownership from public to private hands. What are the implications of alternative policies vis-à-vis FDI for welfare?

If FDI comes simply because the returns to investment are artificially raised by restrictions on competition, the net returns to the host country may be negative (returns to the investor may exceed the true social productivity of the investment). To some extent the rent appropriation may be prevented by profit taxation or by holding competitive
auctions of licenses or equity, but the static and dynamic inefficiencies from lack of competition would still exist.

This situation is depicted in Figure 5. Consider a country in which the domestic market is initially monopolized. Domestic demand is denoted by HD, and MR is the corresponding marginal revenue curve. The constant marginal costs of the monopolist are denoted by MC. The monopoly produces $Q_M$ and charges a price $P_M$. National welfare is equal to the area HAEJ, i.e. the sum of consumers’ surplus, HAP_M, and producer’s surplus, P_M AEJ.

Now suppose one foreign firm is allowed to enter the market, converting it from a monopoly to a duopoly. For simplicity we assume that the foreign firm’s marginal costs are identical to those of the national firm. It is also reasonable to assume that the duopolists neither collude nor indulge in cut-throat competition, so that the outcome is between a monopoly and competitive one – and results in aggregate output $Q_D$ and price $P_D$. Since marginal costs are identical, each firm produces half the aggregate output and makes half the aggregate profits. Aggregate national welfare is equal to the area HCGDJ, as before the sum of the consumers’ surplus HCP_D, and the producer’s surplus, P_D GDJ. Note the former has increased while the latter has declined: consumers benefit from foreign competition but the national firm loses. An examination of national welfare in the two situations reveals that national welfare could decline with entry if area GBDE were greater than area ACB. This is an empirical question, but is clearly possible: the national firm could lose more than consumers gain. However, it is important to remember that the main reason for a decline in welfare would be the appropriation of rents by a foreign firm. If there were no restrictions on entry, and a perfectly competitive outcome resulted, then national welfare would necessarily increase.

To summarize, full liberalization in terms of removal of all barriers to entry leads to an increase in social welfare. Partial liberalization, in terms of limited entry or only change of ownership, could lead to a decline in national welfare.
B. Foreign investment with regulated pricing

In many cases, services (such as electricity, water, telecommunications) are provided by regulated monopolies. In this case, allowing foreign owned firms to provide the service can generate benefits provided that the monopoly is efficiently regulated. However, if the service provider extracts some rents via the regulation process, then the benefits of foreign provision are not so clear.
Consider Figure 6. Let AC denote the average cost of the domestic firm, and D denote domestic demand. Then if the regulator enforces average cost pricing, output would be $x_o$ and price would be $p_o$. Now consider a foreign-owned firm which could provide the service locally with an average cost $AC^* < AC$. With average cost pricing, output increases to $x_1$ and price falls to $p_1$, yielding social benefits equal to area A (reflecting an increasing in consumer surplus). Allowing foreign provision leads to increased productivity, benefiting consumers.

Regulation is not always perfect, however. Suppose that the regulator is not able to extract all average costs, so that the price charged is average cost plus a markup $\pi$. The markup could result either from asymmetric information about the cost function, or from corruption. This is illustrated in Figure 7. With domestic production (with average cost AC), output is $x_o$ and price is $p_o$. Now compare this with the effects of allowing a foreign firm to provide the service. The foreign average cost $AC^*$ is lower and so potentially there are gains from trade. If the foreign firm is able to achieve the same markup $\pi^* = \pi$, then the new price is $p_1$ and output rises to $x_1$. Domestic consumers benefit from the higher productivity of the foreign service provider.

However, it is not obvious that this is beneficial to the economy. The gain in consumer surplus is $A+B+C+E$. But the loss in rents (via markups) accruing to the domestic producer is $A+B+C+F+G+H$. The net social gain is $E - (F+G+H)$, which is negative in the example illustrated. Even though the foreign firm is more productive and charges a lower price, the economy is worse off by allowing foreign ownership of the local utility in this case. Why? Rents from the markup previous accruing to the local firm now go to the foreign firm. This leakage of rents can more than offset the increased benefits to consumers.

Opening up to foreign provision need not necessarily be harmful, however. If the foreign average cost is sufficiently lower than domestic average cost, then the price fall will be higher, area E will be larger, and there will be net benefits. Alternatively, even in the case illustrated, the domestic government may be able to avoid rent leakage by auctioning off the rights to provide the local service. If auctioning is competitive, then in
Figure 6. FDI with Average Cost Pricing
Figure 7. FDI with Markup Pricing
principle, the foreign firm's markup could be extracted, and true average cost price could be implemented, yielding unambiguous social gains.

However, matters could also be worse than illustrated if it is more difficult for the regulator to monitor costs of a foreign firm than a domestic firm. If the foreign firm imports parts and management skills, and pays royalties for proprietary process to the parent firm, then because of familiar issues of transfer pricing, the foreign firm may be able to inflate its reported costs and push up the effective markup $\pi^*$. This would increase rent leakage and further reduce the benefits (or increase the costs) of foreign provision.
Box 3: Are there good reasons to limit entry?

Entry restrictions are becoming harder to justify in the face of growing evidence of the benefits of competition. In Latin America, for example, countries that granted monopoly privileges to telecom operators of six to ten years to the privatized state enterprises saw connections grow at 1.5 times the rate achieved under state monopolies but only half the rate in Chile, where the government retained the right to issue competing licenses at any time.

Why then do we observe such widespread restrictions on entry? While it is possible to construct special models of market and/or regulatory failure where entry barriers enhance welfare (Laffont, 1999), there are usually more prosaic reasons for the barriers. First, restrictions generally aim to protect the incumbent suppliers from immediate competition for infant industry type reasons, to facilitate "orderly exit" or simply due to political economy pressures. And the result is protection not only of national firms but also foreign incumbents – as in the case of foreign insurance companies in Malaysia, and, most strikingly, the bilateral agreements in air transport. Other instruments, such as discriminatory subsidies or taxes could be better targeted.

Monopolistic or oligopolistic rents are also sometimes seen as a means to help firms to fulfil universal service obligations through cross-subsidization (When South Africa partially privatized its telecommunications utility, it conferred a limited monopoly in return for which it required that the utility increase the number of rural connections). However, governments are increasingly devising means of achieving these objectives without sacrificing the benefits of competition: e.g. by imposing universal services obligations on new entrants or asking for competitive bids for subsidies to serve unprofitable areas. In some cases, a form of "investment pessimism" exists, leading to the belief that promises of oligopoly rents are necessary to attract new investment. However, it is not clear why the market structure needs to be determined by policy, unless there are some initial investments the benefits of which may be appropriated by rivals. Finally, governments may seek to raise revenue (or rents for politicians/bureaucrats) by auctioning monopoly or oligopoly rights. This amounts to indirect appropriation of consumers' surplus. But the static and dynamic inefficiencies consequent upon lack of competition would still exist.

Ideally, governments should not resort to trade restrictions to pursue objectives which are better achieved through other means. In each of the cases mentioned above, entry restrictions are at best a second or third-best instrument to achieve the objective in question, but are chosen because of constraints such as the inability to raise revenue without economic or political cost. It will probably be difficult and not necessarily desirable to outlaw completely barriers to entry. But it may be possible to create a legal presumption against such barriers by requiring that a country which imposes them demonstrate that they are necessary – in the sense that more appropriate instruments are not feasible.
C. Regulation of Professional Services

Most professional service providers, such as doctors, engineers and lawyers, have to be certified in order to practice their profession. The certification requirements can serve as barriers to trade because they raise entry costs for foreign service providers. Foreigners often have to take courses, exams, and sometimes establish a residence to meet local certification requirements. In some cases, such as in law, foreigners are sometimes completely shut out of the market.

Certification requirements, however, cannot simply be dismissed as a trade barrier because they are a response to problems of asymmetric information in these markets. The client often does not have enough information to judge the safety and quality of the service. Even if the client could determine quality and safety at some cost, it can be more efficient to require certification to economize on screening costs.

To illustrate the interaction between trade policy and regulation of professional services, we consider an example where low quality service providers generate externalities. For example, if a bridge collapses or if a public building is not constructed safely, then there will be costs to society at large, not just the contractor. Similarly, in countries with public medical systems, the costs of bad medical treatment will fall not just on the patient but also on taxpayers.

We consider two scenarios. First we consider a case where the domestic regulatory system is initially inadequate. That is, screening is imperfect and there is a mix of good and bad service providers in practice. In this case, we show that trade liberalization without reforms the domestic certification system can lower welfare. Next we consider the effects of trade under a well-functioning certification system. In this case, trade will be welfare-improving. However, the magnitude of the gains depend on the way in which foreigners are screened.

Figure 8 illustrates a market for professional services. There are two types of service providers: qualified (Q) and unqualified (U). Unqualified providers make mistakes that cause external harm. The demand for services is denoted by D. $S^Q$ is the
Figure 8. Trade in professional services with imperfect screening
(long run) domestic supply of qualified personnel (this takes into account training costs). These providers do not generate external harm. $S^{Q+U}$ is the combined domestic supply of qualified and unqualified personnel. The curve $E(MSC^{Q+U})$ measures the expected average social cost of service provision by the mix of qualified and unqualified personnel.

If initially there is no trade, the price is $p_o$ and output is $Q_o$. Now consider the effects of allowing foreigners to provide services under the same rules affecting local providers. For simplicity, we have assumed that the mix of qualified and unqualified personnel is the same among foreigners as among domestic suppliers, so that the expected external costs are the same.

When trade opens up, price falls to $p^{Q+U}$, and consumption rises to $Q_2$. Domestic output falls to $Q_1$ because of increased competition from foreigners. Consumers gain from lower prices and domestic producers lose. The net gain in consumer and producer surplus (ignoring the external harm caused by bad screening) is $A+B$. However, the increase in service provision caused by trade also generates more harm as more mistakes are made by unqualified service providers. The increase in external harm is measured by $B+C+E$.² Hence the net social gain from trade is $A - (E+C)$. Welfare could rise or fall: in the example here it falls.

The point of this example is similar to the point raised in the discussion of entry barriers above. If the domestic regulatory system is initially inadequate, then liberalizing trade can exacerbate the problems arising from imperfect regulation. The ensuing costs can potentially more than offset the gains from trade liberalization.

This does not mean, however, that trade liberalization is a bad policy; however. On the contrary, if domestic regulatory reform accompanies trade liberalization, then gains can be assured. To see this, consider Figure 9. Now suppose that the

² Recall that we have assumed that the domestic and foreign providers generate on average the same external harm. Consequently, there is no change in the harm generated by the displacement of domestic providers - it is only the increase in output beyond the initial output level that generates more harm. If foreigners are on average less qualified than domestic agents, this would generate additional costs; if they are better qualified, this would generate some offsetting gains.
Figure 9. Trade in professional services with reform of screening
government introduces a certification requirement that fully screens both domestic and foreign personnel. Let \( P^Q \) be the price of services provided by qualified foreign service personnel. If trade opens, price falls from \( P_o \) to \( P^Q \), consumption rises to \( Q_2 \), and domestic output falls to \( Q_1 \) (since now only qualified domestic agents supply output). In this case in addition to the net gain in consumer and producer surplus, there are additional gains as the external harm from mistakes is eliminated. Social gains are therefore \( B+E+C+A+F+K+G+H+I \); these are unambiguously positive. Trade liberalization, accompanied by domestic regulatory reform raises welfare.

A final issue is how foreigners should be screened. Trade liberalization in many service industries, and especially in professional services requires both that foreigners be given market access, and that they not be subject to discriminatory barriers. There are several ways of implemented non-discrimination rules.

In a *National treatment regime* governments have the flexibility to implement their own regulations subject to the requirement that the same regulations apply to domestic and foreign suppliers. That is, national treatment regime essential requires non-discrimination. Such as rule does not completely eliminate discrimination, however. For example, an insurance company might be required to establish a local office before it can sell insurance. While both domestic and foreign firms are subject to the same requirement, it may be much easier for local firms to meet the requirement - such a rule essentially imposes a fixed cost that excludes foreign firms who might want to do only a small amount of business locally. Similarly, a requirement that engineers, doctors, or truck drivers obtain domestic licensing and certification can impose additional costs on foreigners who have already gone through a similar certification process in their own country.

In a *mutual recognition* regime, each country agrees to accept services providers who meet the certification requirements of their home country. For example, under a mutual recognition regime, a US resident can be permitted to drive in Canada as long as he or she holds a valid US driver's license. Under a national treatment rule, the US resident could be permitted to drive in Canada only if a Canadian drivers license is
obtained. The advantage of mutual recognition regimes is that they can economize on regulatory costs. The disadvantage is that it can be more difficult for a government to meet its regulatory objectives. If the trading partner has weaker certification requirements, then the average quality of service provision may fall when imports increase. For this reason, mutual recognition is not appropriate for many types of services, and may also only be feasible for countries with very similar approaches to regulation. However, in countries with very different standards, the country with the weak standard can agree to recognize certification from the country with the high standard, but not vice versa.

Harmonization of regulatory standards is another option. In this case, countries agree on common regulatory standards. For example, countries may agree on a common set of rules to regulate insurance companies, and this then may facilitate easy access by insurers to markets in each country. The advantage of harmonization is that it removes the ability of governments to unilaterally adjust standards to favor local suppliers. The disadvantage is that it can add inflexibilities into the system that make it more difficult to change regulations when conditions merit changes. As well, harmonization by its nature eliminates diversity in regulatory approaches. Regulations appropriate for one country need not be the best solution for other countries.

Each of these approaches constrains government flexibility in some ways, and so may raise regulatory costs. They may also conflict with other government objectives. For example, a government may choose to have an exclusively public education or health system. In cases of cultural services, a government pay prefer to explicitly favor local providers. For some types of services, governments may see value in having them provided by producers with a long term vested interest in a local community. The literature on social capital suggests that this can strengthen communities and provide both economic and non-economic benefits. Consequently, for some types of services the conflicts between regulatory objectives and trade liberalization may be difficult to circumvent.
We consider here just one simple example to illustrate how the choice of rules for market access may affect trade and welfare. Consider Figure 10. Suppose the domestic government fully screens professionals, so the domestic supply is $S^Q$. There is no external harm generated. Under a national treatment regime, foreigners could be required to undergo the same (costly) certification procedures as domestic agents. Let $P^\text{NT}$ be the price foreigners would charge under these rules. Opening up to trade (allowing foreigners market access) under this rule would generate gains from trade to area $A$, reflecting a fall in price and increased consumption. Suppose however, that there are three types of foreign service providers: those who have evidence that they are fully qualified because of the certification they have received in their own countries; those who need some retraining to adapt the local market; and those who must undergo the same full certification process as local professionals. Let $S^F$ denote the sum of the domestic and foreign supply curve when the domestic government fully recognizes foreign credentials, and require just enough retraining to meet local standard. The initial flat part of the curve corresponds to the supply of professionals certified in countries which clearly meet the standard; the later flat part (at level $P^\text{NT}$ corresponds to supply from those that need full retraining), and the upward sloping part in the middle reflects those that need only partial retraining.

Under this regime, output is higher and price is lower. The gains from trade are given by $A+B+E+C$. The gains from trade will be larger if the government chooses the most efficient screening regime.

**D. Protection to preserve local product variety**

Much of the above analysis of trade barriers assumed that local and foreign products are essentially the same. However, in some cases, local and foreign products may be very different. This is particularly true in the cultural sector, where a foreign movie or television program is often very different from a local production. Hence while foreigners may object to local content rules in broadcasting and other media as
Figure 10. National Treatment vs. Recognition
protectionist, the motivation for such policies may be to preserve distinctly local cultural products. The issue of the preservation of distinctly local services arises other contexts as well: for example, allowing foreign restaurant franchises (such as fast food chains) may squeeze out local restaurant selling indigenous cuisine.

There are two ways to analyze protective policies in the case where locally produced services are distinct from foreign-produced services: the first allows for externalities in service provision, and the second is to confront the larger issue of optimum product variety.

Let us first consider the case where there are externalities. This is illustrated in Figure 11. Let D denote demand for television programming and let S denote local supply. Suppose that viewing of locally-produced programs has spillover effects - it might promote a sense of community or preserve and enhance social capital. Since all local production is consumed locally in the example presented here, the spillover benefits of local consumption of local products can be represented as a positive externality from local production - hence the social marginal cost of production (private costs less spillover benefits) is given by SMC, which is lower than the supply curve (which represents private marginal costs). In free trade, domestic production is \( x_o \). However, because of the externality, this is too low: the socially efficient level of domestic consumption is \( x_1 \). A tax \( t \) on foreign programs (or an auctioned quota) will implement the efficient domestic output level. However, it does so at the cost of creating a consumption distortion (consumer prices rise). The social cost of the tariff is \( A - B \). Area A is the cost of the consumption distortion. Area B represents the external benefits of increased local programming. Note that if the external benefits (B) are sufficiently high, trade restrictions can improve local welfare.

The analysis in the previous section suggests that a better policy than a tariff or quota would be a production subsidy for local programs. This would indeed achieve the first best: a production subsidy set equal to t would increase output to \( x_1 \) but would not affect total consumption, and so would achieve net social benefits of b, even after the financing cost is considered. It should be noted, however, that a production subsidy
Figure 11. External benefit from local production
discriminates in favor of local suppliers and hence would violate some notions of free trade in services.

An argument for preferential treatment of local suppliers can also be made without resort to the externality argument. Consider Figure 12, which is based on Snape (1977). Let AC denote the average cost of providing a local service (such as a weekly television show, a local restaurant, or local music). We assume there is a fixed cost of production; this leads to downward sloping average costs. MC denotes the marginal cost. Suppose there are initially trade barriers that reduce the number of foreign products that are close substitutes for the domestic service in the domestic market. Given this initial level of protection, the domestic demand for the domestic service is $D_o$. Since the demand curve is above average cost for some levels of output, the local service is viable and will be produced.

Now suppose that a trade agreement eliminates local content preferences, and more foreign firms enter the local market. Because foreign services are imperfect substitutes for the domestic service, the local demand curve shifts in. Let $D_1$ denote the free trade domestic demand for the domestic service. Since $D_1$ is below average cost, the domestic production shuts down. Foreign movies, television programs and music squeeze out domestic products.

Now it might be thought that these products should be squeezed out because domestic demand is below average cost. However, this is not necessarily correct. To see this, suppose that (in free trade) $x_o$ units of the domestic service were supplied at price $p = MC$. Consumer surplus would be $A+B$, and the loss to producers would be $B+C$ (this reflects the costs that are not recovered by charging $p$). The net social gain from provision of the domestic service is $A - C$, which is clearly positive in the example illustrated. That is, even though foreign competition squeezes the domestic service out of the market, it would nevertheless still be socially efficient for the domestic service to be supplied. When there are fixed costs, there is no presumption that free markets will provide optimal product variety. Either trade barriers or domestic subsidies could be superior to free trade in this case.
Figure 12. Local product squeezed out by trade
E. Trade policy substitution between modes of delivery

We conclude our analysis of the interaction between trade policy and regulation with an example where different types of trade barriers apply to different modes.

The exemption of one mode of delivery from taxation while others continue to be taxed is analogous to a preferential trading arrangement. And as in the case of such arrangements there is a positive trade-creating aspect, and a negative trade-diverting aspect. The latter arises when the tax-exempt mode is chosen simply to avoid the tax, even though it is less efficient than the taxed alternative.

This is illustrated with Figure 13. D and S are domestic demands and supplies for a service X. Suppose that foreigners can provide the service directly to local customers through temporary movement of personnel. This would cost \( w^* \) per unit. Alternatively, the service can be provided at higher cost through cross-border trade at price \( p \) per unit.

With free trade in all modes, price is \( w^* \), domestic output is \( x_o \), and consumption is \( c_o \). Now suppose that a tax \( t \) is imposed on temporary movement of personnel. First suppose this was the only way to provide the service (that is, first suppose cross-border trade is not feasible). Then the price would rise to \( w^* + t \), consumption would fall to \( c_2 \) and output would rise to \( x_2 \). Domestic producers gain, consumers lose, and the social cost of the tariff is \( B+F+G+D+I+J \) as usual.

Now suppose that cross-border trade is feasible at price \( p \). Then the imposition of the tax on temporary movement of personnel lead to a switch from this mode to direct trade because direct trade is now cheaper (\( p < w^* + t \)). At first glance, it may appear that this is good thing. Because of the option of cross-border trade, the tariff causes price to rise to only to \( p \) (instead of \( w^* + t \)), and consumption to fall to \( c_1 \) (instead of \( c_2 \)). The switch to the alternative mode has resulted in less of a price increase, and more consumption than in the case where this option was not available. That is, the availability of multiple modes of service provision means that providers can respond to trade barriers.
Figure 13. Substitution between modes of supply
by finding alternative modes, and this can help to keep prices down and reduce the impact of the trade barriers on consumers.

However, there is another effect. Let us consider a cost / benefit analysis of the tariff in the presence of mode switching.

Rise in Producer Surplus: \( E \)
Loss in Consumer Surplus: \( E+F+G+H+I+J \)
Tariff revenue: \( \text{none} \)
Net social \( \text{Loss} \): \( F+G+H+I+J \)

Recall that if cross-border trade was not feasible, the tariff would have led to a social loss of \( B+F+G+D+I+J \). With mode-switching, the loss from the tariff is \( F+G+H+I+J \). The option of switching to cross-border trade in response to the tariff on temporary labor movement will generate welfare gains if \( B+D > H \). However, this need not hold in general, and note that in the case illustrated in Figure 13 we have the reverse: \( H > B+D \).

Area \( H \) represents the cost of trade diversion. In this example, cross-border trade is less efficient than direct provision by temporary movement of personnel. A switch to cross-border trade therefore induces an increase in the cost to the economy of acquiring the service. This increased cost has to be balanced against the lower price that results from mode-switching. If the trade diversion effect dominates, then mode-switching contributes to a decline in welfare.

The point of this example is that the analysis of the effects of trade barriers in services can be complicated by the possibility of mode-switching induced by trade barriers. Even though mode switching generates some benefits, it also creates costs via the trade diversion effect. If the trade diversion effect is not accounted for, then the costs of protection may be underestimated.
VII. Trade agreements vs. Unilateral liberalization

Although it is only recently that trade agreements have begun to focus on liberalizing trade in services, governments have been opening up their markets to foreign service providers for a long time. Moreover, as technology changes, and as global markets evolve, governments have responded to pressures from both producers and consumers to change their rules to allow increased access to their markets. This raises the issue of why governments need to be constrained by trade agreements in services. If designing trade agreements is going to raise complicated issues involving trade-offs between internal regulatory flexibility and trade, why not just let each government decide on its own rules regarding foreign access and avoid the complexity of trade agreements?

There are three major reasons for signing a trade agreement rather than simply liberalizing trade unilaterally.

There are benefits from *reciprocity*. That is, a major advantage of trade agreements is that they can ensure that each country gains increased access to the other's market. If governments are under pressure to protect jobs for engineers in their country, then an agreement to allow foreign engineers into the country may well benefit local consumers, but is likely to encounter resistance from producers. But if it is coupled with an agreement that includes increased access for domestic producers in the foreign market, then the gains from increased trade will be greater and domestic resistance will be lessened.

A second aspect of reciprocity is that it can help avoid trade wars. It is well known that governments can get into self-defeating trade wars, as for example happened during the Great Depression. Each government may face pressure to protect local firms. But if all governments do this, then overall economic activity falls and all countries can be worse off. The situation is much like a prisoners' dilemma. The optimal strategy for an individual government may be to protect, but if they all do it, they are all worse off. Coordinated liberalization can circumvent the problem.
Trade agreements also have important commitment advantages, in both a local and international context. In the local context, the advantage of signing a free trade agreement is that it helps a domestic government stand up to local protectionist interest groups. Producers stand to gain from trade barriers, and although the barriers may lower social welfare as discussed above, the costs are widely dispersed across consumers. Consequently producer lobbies for protection will often be stronger than consumer lobbies for free trade. Governments may therefore face enormous pressure to cave into producers and raise trade barriers that ultimately harm the country. If the government signs a trade agreement, then the costs of caving in are increased. This commitment effect would then have two advantages: trade barriers will stay lower because the government caves in less often; and fewer resources are wasted on lobbying because the lobbyists know the payoff to lobbying has fallen.

The second commitment effect is in the international context. Service providers often have to invest in fixed costs to serve a foreign market, either through direct investment or in obtaining certification or developing local knowledge. Because the investment will be worthless if they are later shut out of the foreign market, firms may be reluctant to make such investments and a country that wants to attract foreign service providers may fail to do so. Even if foreign providers are not shut out, they may feel vulnerable because of what is known as the hold-up problem. Once foreigners have invested in country-specific skills, they may be concerned that the foreign government will try to extract rents from them by imposing discriminatory taxes or other requirements. Signing a free trade agreement can help to increase the credibility of a commitment to not either cut off access once investments have been made, or to not arbitrarily change the rules to extract rents.

Finally, a commitment to a trade agreement allows the development of a rules-based system that helps facilitate trade. Much of the success of internal trade within a country relies on confidence that contracts will be enforced and rules will be applied in a predictable manner. International governments do not exist to fill this role; but trade
agreements lead to the development of institutions that can help settle disputes and increase the consistency with which rules are applied.

On the other hand, although trade agreements have their advantages, whether any given trade agreement is better than unilateralism depends on the terms of the agreement. Trade agreements in services cannot avoid intruding on what used to be thought of as internal domestic policy. Trade agreements require trade-offs between flexibility, domestic sovereignty and international trade. These trade-offs are much more complex and intrusive than in the case of goods trade. Different countries will have different approaches to these trade-offs, and it is not even clear what free trade means in some cases. So although there are a number of reasons to pursue trade liberalization via trade agreements rather than unilaterally; it may also be very difficult to design trade agreements that are both effective and acceptable to a wide range of countries.

VIII. Summary

Trade in services is in some important ways different from trade in goods. First, many services require proximity between the supplier and the consumer, and hence factor mobility is necessary for many international service transactions. Secondly, the limited scope for “border” restrictions implies that domestic regulations have a much stronger influence on trade in services.

However, while interesting twists arise because of the way services are traded and regulated, the basic insights from the theory of trade in goods apply to trade in services. There are likely to be substantial gains from liberalizing trade in services, immediately and in the longer term provided the regulatory framework is adequate.

Notes on the literature. For an in-depth treatment of the motives for trade and trade policy, with a focus on goods trade, see any standard trade theory textbook, such as Krugman and Obstfeld (2003). For an early
application of trade theory to services, see Deardorff (1985). Sapir and Winter (1994) and Copeland (2002) provide reviews of the literature on the economics of trade in services.

References


