Trade, Foreign Direct Investment, and International Technology Transfer: A Survey

Kamal Saggi

What role does trade play in international technology transfer? Do technologies introduced by multinational firms diffuse to local firms? What kinds of policies have proved successful in encouraging technology absorption from abroad and why? Using these questions as motivation, this article surveys the recent trade literature on international technology transfer, paying particular attention to the role of foreign direct investment. The literature argues that trade necessarily encourages growth only if knowledge spillovers are international in scope. Empirical evidence on the scope of knowledge spillovers (national versus international) is ambiguous. Several recent empirical plant-level studies have questioned earlier studies that argued that foreign direct investment has a positive impact on the productivity of local firms. Yet at the aggregate level, evidence supports the view that foreign direct investment has a positive effect on economic growth in the host country.

Economic growth results from accumulation of factors of production or from improvements in technology or both. To encourage the generation of new knowledge, industrial countries have elaborate systems of intellectual property rights (IPRs) in place and conduct the majority of the world’s research and development (R&D). Technologies resulting from R&D spread throughout the world via a multitude of channels. On a fundamental level, international trade in technology differs from other indirect channels of international technology transfer, such as trade in goods and international movement of factors of production. This article critically surveys the literature that explores the roles of trade and foreign direct investment (FDI) as channels of international technology transfer. With respect to FDI, a distinction is made between wholly owned subsidiaries of multinational firms and international joint ventures. Furthermore, the role of FDI is contrasted with that of arm’s-length channels of technology transfer, such as licensing.

Although the literature has done a decent job of outlining the various channels through which international technology transfer occurs, not enough is known, both
in theory and practice, about the relative importance of each of these channels. There is a limited understanding of the role that policy can play in facilitating international technology transfer. For example, the literature continues to debate whether increased openness to trade encourages economic growth. Although Dollar (1992) and Sachs and Werner (1995) find support for the view that open economies grow faster, Rodriguez and Rodrik (1999) are quite skeptical about this conclusion.

As a practical matter, few economists advocate the imposition of trade restrictions. In fact, the general feeling seems to be that traditional analyses may very well understate the true cost of protectionism because most utilize static models and do not capture the dynamic costs of trade protection. Underlying this view is the notion that, somehow, trade, FDI, and interaction among countries in various other forms all play roles not only in improving the global allocation of resources but also in transmitting technology globally. How exactly this transmission occurs is not fully understood, making international technology transfer an active area of research.

Dynamic trade models shed light on the complex relation between technology and trade. These models frequently lead to ambiguous welfare conclusions. The literature (both theoretical and empirical) does not provide a blanket endorsement of trade as an engine for growth because introducing dynamics in an interesting fashion often requires multiple departures from the neoclassical model of perfect markets. Imperfect competition and externalities are central to the new dynamic models of trade, and such distortions can easily lead to perverse results. Of course, the argument cuts both ways. Introducing such elements in the traditional static model also furnishes additional arguments in support of free trade. Nevertheless, it is difficult to make the unconditional case for free trade purely on the basis of logic. For example, see Krugman (1987) for a pragmatic argument for free trade even in the presence of market failures.

It is also not the case that anything can happen if a closed economy opens up to free trade. In fact, the theoretical literature suggests that the scope of knowledge spillovers is a crucial determinant of whether trade necessarily encourages growth (Grossman and Helpman 1995). However, the empirical evidence has been mixed: some studies discover that knowledge spillovers have a limited geographical scope, whereas others find the opposite. The ambiguous nature of this empirical evidence immediately raises the following question: What factors determine the scope of knowledge spillovers? Clearly, the scope of knowledge spillovers must be determined in part by the interaction between innovators (potential suppliers of technology) and those firms and entrepreneurs that seek access to newer technologies through imitation, technology licensing, and other forms of collaboration with innovators. In other words, a fair bit of technology transfer may indeed be endogenous.

In a discussion of the special properties of knowledge as an economic good, Romer (1990) makes the important point that knowledge is a nonrival good: it can be used simultaneously by two different agents. However, this does not mean that knowledge
can be transferred across agents at zero cost. If technology transfer entailed no costs, the room for fruitful policy intervention with respect to assimilation of foreign technology would be quite limited because any technology transfer that would yield even a minutely positive return to any agent would take place automatically. Pack (1992) provides an overview of what can be reasonably expected in terms of technology transfer to developing economies, given that the potential for transfers is large.

The nonrival nature of knowledge only implies that if two agents are willing to pay the cost of adopting a new technology, they can do so without interfering with each other’s decisions. Much empirical evidence indicates that international technology transfer carries significant resource costs (Mansfield and Romeo 1980; Ramachandran 1993; Teece 1977). In his survey of 29 technology transfer projects, Teece (1977) found that on average such costs were approximately 20 percent of the total costs of the project, and in some cases, they were as high as 60 percent.

The fact that international technology transfer occurs through a multitude of channels makes it difficult to arrive at an aggregate measure of the activity and accurately assess its contribution to economic growth. Most research, theoretical as well as empirical, tends to focus on one or two channels of technology transfer. Of these, trade and FDI have received the most attention. If one could somehow rank the different channels of technology transfer in terms of their relative importance, empirical analysis could then proceed by ignoring the relatively unimportant channels. However, given that multiple options exist in theory, the dominance of any one channel in the data would itself require explanation. Indeed, the emergence and expansion of multinational firms, given the existence of alternative arrangements for transacting in technology, has been viewed as a phenomenon that requires explanation. Furthermore, the various channels of technology transfer, though independent to a certain degree, are linked to each other in important ways. For example, the extent to which inward FDI contributes to technology transfer (in addition to international trade) may very well be a function of a country’s trade policy. An important challenge for both theoretical and empirical research is to isolate the marginal contribution of inward FDI to technology transfer and its relation to a country’s trade and investment policies. I discuss the relevant research in this survey to the extent the literature has addressed these questions.

Once a technology has been introduced into a country (by a multinational firm, say), how does it subsequently diffuse throughout the rest of the economy? The presence of trade barriers across countries as well as international differences in market conditions and policy environments imply that technology diffusion within a country should be considerably easier than international transfer of technology. For example, the mobility of labor is severely constrained only at the international level (exceptions include contact with consultants and the return of foreign-educated nationals). Thus labor turnover across firms may be crucial for driving technology diffusion within a country and may not play any role in international technology transfer. This
article discusses the role of technology licensing, imitation, and FDI in the process of international technology transfer and its subsequent diffusion in the host country.

One goal of this article is to help identify the role policy plays in facilitating international technology transfer. The range of relevant policies is clearly quite large. To limit the scope of the discussion, I address the role of trade, FDI, and IPR policies. Given the central questions of interest, I discuss the literature on FDI and IPRs in greater detail than that on trade policy.

Blomström and Kokko (1998) review the theory and evidence regarding spillovers from multinational firms. Unlike the present article, their survey is concerned exclusively with spillovers from FDI and does not deal with trade and the effects of policy on international technology transfer. Furthermore, they do not emphasize the endogeneity of international technology transfer. Blomström and Kokko focus on the internal diffusion of technologies introduced by multinational firms. By contrast, this article emphasizes the interaction between domestic diffusion and incentives for international technology transfer. Furthermore, at least in the context of tradable goods, a relevant issue (not addressed by the Blomström and Kokko survey) is that a complete definition of spillovers from FDI ought to account for technology diffusion that would result in the absence of FDI but in the presence of international trade.

Knowledge Spillovers through Trade

Figure 1 plots worldwide exports of goods and services as a percentage of gross domestic product (GDP) for 1970–96. During this period, the percentage of exports increased from approximately 14.1 to 21.4. An interesting consideration is whether this increase in world trade yielded primarily static efficiency gains or dynamic gains by facilitating international technology transfer. An extensive literature studies the dynamic effects of trade. Much of the relevant work emphasizes two intertwined aspects of the relationship between trade and technology: that trade alters the allocation of resources in an economy and plays a role in transmitting knowledge internationally.

Much of the literature on trade and international transmission of technology derives from closed-economy growth models. For this article, endogenous growth models are those in which economic growth results from the intentional actions of individuals who seek to profit from their investments in technological innovation. In traditional growth theory, capital accumulation is the major determinant of economic growth; a natural conclusion of this research was that unless the return to capital accumulation could stay bounded away from zero, growth would peter out or cease in the long run. A natural implication of this finding is that over time, poor countries eventually converge to the per capita income levels of the countries. However, the evidence on convergence is weak; although some areas—such as the Re-
public of Korea; Taiwan, China; and Hong Kong, China—have achieved enviable rates of growth for sustained periods, most developing economies do not seem to be on a path of convergence toward rich countries (Pritchett 1997).

Standard neoclassical growth models assume costless technology transfer by positing a common production function across countries. The fact that chosen production techniques differ across countries is not evidence against the neoclassical view; when faced with different factor prices (due to differences in factor endowments), firms typically adopt different production techniques in different countries. Thus, the issue is whether all firms can access the global pool of technologies at the same cost. Parente and Prescott (1994) emphasize barriers to technology adoption as a key determinant of differences in per capita income across countries. In their model, although any firm can access the underlying stock of knowledge in the world economy, the cost of such access differs across countries due to differences in legal, regulatory, political, and social factors. Thus in their view, some countries make it inherently costlier for their firms to adopt modern technologies and thereby retard the development of the entire economy. In fact, Parente and Prescott (1994) go on to suggest that trade may affect growth by lowering the barriers to technology adoption.

In contrast to neoclassical models that stress capital accumulation, the new growth theory emphasizes technological change and the accumulation of human capital.

**Figure 1.** Worldwide Exports over GDP, 1972–96

![Graph showing worldwide exports over GDP from 1972 to 1992.](source: World Bank (1999).)
For the purposes of this survey, the literature on R&D-based growth models is clearly more relevant, and I restrict attention to this strand of growth theory. Romer (1990), Grossman and Helpman (1991), Aghion and Howitt (1990), and Segerstrom and others (1990) are among the pioneers of R&D-based models of economic growth. These models provide a coherent framework for the Schumpeterian notion of creative destruction. Although they differ from each other in important ways, the models all share the idea that entrepreneurs conduct R&D to gain temporary monopoly power made possible by patents and other IPRs.

Grossman and Helpman (1991) provide a unifying framework for two widely used strands of R&D-based endogenous growth models: the varieties model, which builds on foundations laid by Dixit and Stiglitz (1977), Ethier (1982), and Romer (1990), and the quality ladders model developed by Aghion and Howitt (1990), Segerstrom and others (1990), and Grossman and Helpman (1991). In a closed economy, growth is sustained in the varieties model through the assumption that the creation of new products expands the knowledge stock, which then lowers the cost of innovation. As more products are invented, both the costs of inventing new products and the profits of subsequent innovators are lower because of increased competition (no products disappear from the market in this model). By contrast, the quality ladders model assumes that consumers are willing to pay a premium for higher-quality products. As a result, firms always have an incentive to improve the quality of products. The important assumption that sustains growth in this model is that every successful innovation allows all firms to study the attributes of the newly invented product and then improve on it. Patent rights restrict a firm from producing a product invented by some other firm but not from using the knowledge (created due to R&D) that is embodied in that product. Thus, as soon as a product is created, the knowledge needed for its production becomes available to all; such knowledge spillovers ensure that anyone can try to invent a higher-quality version of the same product.

Although R&D-based endogenous growth theory is quite appealing theoretically, empirical evidence does not provide a strong endorsement (Pack 1994). In fact, Jones (1995a, 1995b) explicitly tests the empirical implications of R&D-based growth models and finds that the data reject these implications. However, rejecting a particular model of R&D-based economic growth does not imply that R&D is not an important determinant of growth. In fact, a reasonable conclusion may be that although R&D is crucial for the generation of new ideas (and economic growth), early variants of R&D-based growth models do not adequately capture the relationship between R&D and growth. The newer strand of growth theory has not abandoned R&D as a determinant of growth; instead, it has focused on creating models that do not have the “scale effects” that Jones demonstrates are not supported by the data. Roughly speaking, scale effects imply that large economies grow faster than small economies (see Dinopoulos and Thompson 1999 for a discussion of scale effects in endogenous growth models).
R&D-based models of growth argue that new products result from new ideas; therefore, trade in goods could help transmit knowledge internationally. This is the central insight of many open economy growth models. Of course, trade in ideas can take place without trade in goods. Rivera-Batiz and Romer (1991) analyze two different models (the lab equipment model and the knowledge-driven model) of endogenous growth to highlight the role of trade in goods versus trade in ideas. The general conclusion of their analysis is that trade in either goods or ideas can increase the global rate of growth if such trade allows a greater exploitation of increasing returns to scale (in the production of goods or ideas) by expanding market size.

Multicountry models of endogenous growth have two strands: those that study trade between identical countries and those that have a Northern Hemisphere–Southern Hemisphere structure. Although knowledge spillovers are central to both, technology transfer in the sense emphasized here is a central feature only of North–South models. Prominent early works include Krugman (1979), Rivera-Batiz and Romer (1991), and Grossman and Helpman (1991). The literature is now rather large and a complete discussion requires a survey of its own (see Grossman and Helpman 1995). North–South models that emphasize the product-cycle nature of trade have been particularly useful for understanding international technology transfer and merit some further discussion.

Product-cycle models assume that new products are invented in the North and, due to the lower relative Southern wage (endogenous in the model), Southern firms can successfully undercut Northern producers by succeeding in imitating Northern products. A typical good is initially produced in the North until either further innovation (in the quality ladders model) or successful Southern imitation (in both the varieties model and the quality ladders model) makes profitable production in the North infeasible. Consequently, either production ceases (due to innovation) or shifts to the South (due to imitation). Thus, prior to imitation, all products are exported by the North, whereas postimitation they are imported, thereby completing the cycle. These models capture technology-driven trade and have been generalized to consider technology transfer more explicitly. Neither FDI nor licensing (choices available to innovators for producing in the South) was considered in the early variants of these models.

What do R&D-based models of growth imply about the effect of trade on growth? An important conclusion of this line of research is that much importance hinges on whether knowledge spillovers are national or international in nature (Grossman and Helpman 1995). If the spillovers are international, these models endorse the view that trade is an engine of growth. However, when the spillovers are national, perverse possibilities can arise. Note that this perspective is more relevant for North–North models of trade because international knowledge spillovers (of one form or another) are assumed in North–South product-cycle models of trade, where the South is modeled as a pure imitator. In North–South models, the more interesting question is how Southern imitation affects incentives for innovation in the North.
What factors can help account for the explosive growth of economies like Hong Kong, Korea, and Taiwan? Some economists argue that the accumulation of resources has driven economic growth in these countries (Young 1995). Others argue that improvement in productivity (driven partly through trade) has played a large role (Nelson and Pack 1999). However, even if capital accumulation were the driving force, why did it take place at such a high rate? What kept the returns to capital accumulation so high? Perhaps technology transfer (again, partly through trade) kept the marginal product of capital from falling and kept investment rates high (Nelson and Pack 1999).

What does the empirical evidence say about the scope of knowledge spillovers? Should research focus primarily on determining their geographical scope? The frequent agglomeration of R&D-intensive industries (such as in Silicon Valley) suggests that spillovers may be primarily local. However, several studies find that R&D activity in a country is not strongly correlated with productivity growth, suggesting that the benefits of R&D in one country spill over substantially to other countries. Eaton and Kortum (1996) find that more than 50 percent of the growth in some Organisation for Economic Co-operation and Development (OECD) countries derives from innovation in the United States, Germany, and Japan. Yet Eaton and Kortum also report that distance inhibits the flow of ideas between countries, whereas trade enhances it.

In their micro-level study of the semiconductor industry, Irwin and Klenow (1994) find that learning (resulting from production) spills over as much across national borders as it does between firms in the same country. Similarly, Coe and Helpman (1995) and Coe and others (1997) argue that international R&D spillovers are substantial and that trade is an important channel of such spillovers. Using estimates of international R&D spillovers from these two studies, Bayoumi and others (1999) simulate the impact of changes in R&D and in exposure to trade on productivity, capital, output, and consumption in a multicountry model (the International Monetary Fund [IMF]'s MULTIMOD model). Their simulations indicate that R&D can affect output not only directly but also indirectly by stimulating capital investment. Incidentally, this finding is also of interest for the debate regarding the Asian growth miracle.

Keller (1998) casts some doubt on the latter finding by generating results similar to those of Coe and Helpman (1995) for randomly generated trade weights. However, a recent paper by Coe and Hoffmaister (1999) argues that Keller’s weights are not actually random. When alternative weights are used, estimated international R&D spillovers are nonexistent for the case of random weights.

In principle, trade in both consumption and capital goods can contribute to technology transfer, and the empirical studies discussed typically utilize a country’s imports of all goods while attempting to measure knowledge spillovers through trade. For example, when Korea imports a manufactured consumption good, such as an automobile, local firms can absorb some technological know-how by simply study-
ing the design and the engine of the imported automobile. Attempts at reverse-engineering may be important, but they probably contribute less to technology transfer than does trade in capital goods (such as machinery and equipment) that are used in the production of other consumption goods. Xu and Wang (1999) argue that trade in capital goods is more relevant than total trade for measuring knowledge spillovers because capital goods have higher technological content than consumption goods. Xu and Wang measure trade in capital goods by exports of machinery and transport equipment (sttc). They show that although the volume of such trade helps explain cross-country differences in total factor productivity (tfp), trade in all other goods does not. This result fits well with the finding by De Long and Summers (1991) that investment in machinery and equipment has a strong association with growth.

Capital goods trade is a prominent part of world trade, and its importance has increased over time. Table 1 reports worldwide exports of capital goods (as measured by trade in machinery and transport equipment) as a percentage of total exports. In 1975, approximately 23 percent of total trade in the world was trade in capital goods; this ratio was over 30 percent in 1996. During 1975–96, worldwide exports of machinery and transport equipment as a percentage of gdp increased from about 4.2 to approximately 7.0. In 1996, roughly 30 percent of capital goods exports were destined for developing economies. Although the developing economy share of imports of capital goods has increased over time, this increase has not been substantial (from 28.9 percent in 1980 to 30.8 percent in 1996). Within the oecd countries, there is significant cross-country variation in the magnitude of imports of capital goods. In 1983–90, more than 50 percent of U.S. imports from other oecd countries was made up of capital goods; this ratio was only 25 percent for Japan (Xu and Wang 1999). Such variation in the data suggests that using total trade to measure the degree of knowledge spillovers across countries might lead to erroneous conclusions.

<table>
<thead>
<tr>
<th>Year</th>
<th>Percent</th>
<th>Year</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>23.5</td>
<td>1986</td>
<td>27.9</td>
</tr>
<tr>
<td>1976</td>
<td>23.8</td>
<td>1987</td>
<td>28.3</td>
</tr>
<tr>
<td>1977</td>
<td>24.1</td>
<td>1988</td>
<td>28.5</td>
</tr>
<tr>
<td>1978</td>
<td>25.0</td>
<td>1989</td>
<td>28.0</td>
</tr>
<tr>
<td>1979</td>
<td>22.7</td>
<td>1990</td>
<td>28.4</td>
</tr>
<tr>
<td>1980</td>
<td>21.5</td>
<td>1991</td>
<td>28.6</td>
</tr>
<tr>
<td>1981</td>
<td>22.2</td>
<td>1992</td>
<td>27.8</td>
</tr>
<tr>
<td>1982</td>
<td>23.3</td>
<td>1993</td>
<td>28.8</td>
</tr>
<tr>
<td>1983</td>
<td>23.9</td>
<td>1994</td>
<td>30.2</td>
</tr>
<tr>
<td>1984</td>
<td>24.6</td>
<td>1995</td>
<td>30.6</td>
</tr>
<tr>
<td>1985</td>
<td>25.8</td>
<td>1996</td>
<td>30.7</td>
</tr>
</tbody>
</table>

Most theoretical models assume knowledge spillovers from R&D to be national or international in scope and then compare the predictions of the two scenarios. Following this line of argument, the goal of the empirical economist is simply to determine which assumption is appropriate. Yet this approach sits rather uncomfortably with the central tenets of the literature on trade and growth. A major theme of this literature is that technological change occurs due to intentional and costly investments undertaken by firms that seek to profit from monopoly power that results from successful innovation. If this is the case, arbitrage in knowledge, which is basically what the spread of know-how across countries amounts to, cannot be totally exogenous to economic activity. Those agents that invest heavily in creating new technologies face strong incentives to control the spread of their hard-earned successes. If such control were not possible, they would have little incentive to make those investments in the first place. For the theory of trade and innovation to be internally consistent almost requires that inventors partially control the rate at which their technologies spread internationally. Therefore it is misleading to focus on the geographical scope of knowledge spillovers without giving innovators some role in that process.

In addition to the incentives of innovators, other factors determine the scope of knowledge spillovers, including incentives facing potential buyers and imitators of technologies. The literature in this area has not paid adequate attention to the choices that both potential suppliers and buyers of technology face. However, the literature seeks to explain the emergence of multinational firms that play a central role in international technology transfer.

### Explaining FDI: Location and Mode of Production

There are two distinct questions that a firm seeking to serve foreign markets must address. First, is it better to produce the good in the home country and export to foreign markets, or is production abroad more profitable? Second, for production abroad, how should technology be transferred overseas? Firms can choose from a variety of arrangements that differ in their relative use of markets and organizations. One extreme arrangement transfers technology to wholly owned subsidiaries; the other extreme transfers technology to unrelated parties through licensing.

#### Exports versus Production Abroad

When serving a foreign market, a firm can choose from a menu of options. The literature mainly focuses on the choice between exports and FDI, assuming that exports and FDI are substitutes for one another. However, empirical work usually uncovers a complementary relationship between exports and foreign affiliate sales. For example, Lipsey and Weiss (1981) find that sales of foreign affiliates are positively cor-
related with exports at the industry level. Firm-level studies, such as Lipsey and Weiss (1984), Grubert and Mutti (1991), and Blomström and others (1988) also uncover a complementary relationship between trade and FDI.

Does the evidence imply that most theoretical models are flawed? Perhaps not. Blonigen (1999) suggests that a reasonable interpretation of the evidence shows that such studies find net complementarity: aggregation bias in the data simply buries the substitution effects emphasized in theoretical models. Blonigen’s major contribution lies in using product-level data because theory implies the substitution effect at this level. Blonigen uses data on Japanese production and exports to the United States for two types of products: automobile parts and automobiles. His study is particularly useful in the context of Japanese multinational firms located in the United States that import relatively large amounts of parts from Japan and seem quite unwilling to substitute between U.S. and Japanese parts.

Only a study of the type done by Blonigen (1999) can really sort out the complementary nature of trade between intermediate goods and affiliate sales on the one hand and the substitutability of exports of final goods and FDI on the other. Not surprisingly, Blonigen’s results conform nicely to the theory: exports of intermediate goods and sales of affiliates are complements, whereas exports and sales of final goods are substitutes. The only unresolved issue is why aggregate studies find a net complementary relationship. The explanation probably lies in a fact that Ethier (1982) emphasizes: most intraindustry trade between industrial countries involves exchange of intermediate goods. The literature on intraindustry trade as derived from Dixit and Stiglitz’s (1977) model may overemphasize the role of product differentiation and consumer emphasis on variety. As Ethier (1982) notes, actual trade is in intermediate goods needed for production. Thus, if such trade is indeed pervasive, there should be a strong complementary relationship between exports and FDI at the aggregate level.

The theoretical models have also argued that strategic considerations influence the choice between exports and FDI (see Horstmann and Markusen [1992] and Motta and Norman [1996]). The presence of trade barriers creates a tariff-jumping motive for FDI. Bhagwati and others (1987, 1992) argue that the mere threat of future trade restrictions may lead to anticipatory investment (called quid pro quo investment) by foreign firms. However, the preceding research emphasizes the interdependence of decisionmaking between multinational firms. For example, when two firms are exporting to a foreign market, a switch from exports to FDI by one creates an incentive for FDI on the part of the other firm, which finds itself at a competitive disadvantage (Lin and Saggi 1999 call this the competitive incentive for FDI). An old tradition in the management literature describes the interdependence between the decisionmaking of large multinationals as follow-the-leader behavior.

As far as the static choice between exports and production abroad, the theoretical models seem reasonably well developed. However, firms face a dynamic problem, not
just a one-time choice between exports and FDI. Firms may (and indeed do) switch between the two activities over time. Unfortunately, there is scarce literature that explores the dynamics of optimal entry strategies into foreign markets. Roberts and Tybout (1997) highlight the role of sunk costs in determining the dynamic behavior of exporters. Using data for Colombian manufacturing plants, Roberts and Tybout show that prior exporting experience is an important determinant of the current tendency to export as well as the profitability of exporting. Their findings show that sunk costs are indeed relevant for export behavior and that learning is subject to strong depreciation. The entry costs of a plant that has never exported do not differ significantly from those of plants that have not exported for more than two years.

Although Roberts and Tybout (1997) do not consider other modes of serving foreign markets, their insight can be utilized in a more general context. Suppose firms also have the option of FDI. Building on the Roberts and Tybout approach, the choice between exports and FDI is a choice between two different technologies, where exports entail a higher marginal cost and a lower fixed (sunk) cost than FDI. Under uncertainty, if firms do face such a cost structure, an interesting dynamic relation between exports and FDI may emerge.

Saggi (1998) builds a two-period model to examine a firm’s choice between exports and FDI in the face of demand uncertainty. First-period exports yield information about demand in the foreign market. As a result, first-period exports have an option value. That is, if a significant portion of the fixed cost of FDI is sunk, it is optimal for a firm to export in the first period and to choose FDI if and only if demand abroad is large enough.

Clearly, the preceding argument is not specific to demand uncertainty and can be generalized with respect to other types of uncertainty about which sales through exports can yield information. Similarly, exports and initial FDI may be strongly complementary because firms are not likely to shift the entire production process to a new location immediately. If initial investment is profitable, local sourcing may reduce the need for imported intermediates. Over time, such substitution effects may become stronger, and the complementarity between exports and FDI may become weaker (assuming local suppliers are indeed competitive or local production is consistent with comparative advantage considerations).

Of course, generalizing the preceding argument to the case of multiple firms also creates the possibility of information externalities among investors; that is, the experience of one firm may impart lessons to others. Such externalities may be particularly relevant for FDI in many developing and formerly closed economies (China and much of Eastern Europe) that have only recently opened their markets to foreign investors. Firms from industrial countries have little prior experience in operating in these new environments. This lack of experience coupled with the complexity surrounding the FDI decision implies that firms seeking to invest in these markets can
learn valuable lessons from the successes and failures of others. FDI involves hiring foreign labor, setting up a new plant, meeting foreign regulations, and developing new marketing plans; these decisions require adequate information. In this context, decisions made by rival firms can lower a firm’s fixed cost by helping avoid mistakes. For example, Lin and Saggi (1999) use a duopoly model in which the first firm to switch from exporting to FDI confers a positive externality on the subsequent investor by lowering its fixed cost of FDI.

In their survey of Japanese firms planning investments in Asia, Kinoshita and Mody (1997) find that both private and public information play important roles in determining investment decisions. They argue that information regarding many operational conditions (such as the functioning of labor markets, literacy, the productivity of the labor force, and timely availability and quality of inputs) may not be available publicly. Such information is either gathered through direct experience or through the experience of others. Indeed, Kinoshita and Mody’s empirical analysis finds that a firm’s current investment is strongly affected by its own past behavior as well as by the investments of its rivals.

Although the degree of fixed/sunk costs may play a role in determining the choice between licensing, joint ventures, and FDI, other considerations are probably more important. A new foreign plant is the primary contributing factor behind higher fixed/sunk costs of FDI relative to exports. This factor is unlikely to be of first-order importance in determining the choice between different entry modes that are distinguished primarily by the extent of foreign ownership.

**Mode of Operation: Licensing, Joint Venture, or FDI?**

A major question in the theory of the multinational firm is when and why firms choose to internalize technology transfer, thereby forgoing the option of utilizing market based alternatives such as technology licensing. Markusen (1995) and Caves (1996) discuss the relevant economics literature regarding internalization. A vast literature in the field of international business deals with some of the questions posed. By and large, this literature involves empirical tests of the ownership, location, and internalization paradigm developed by Dunning (1988). To limit the scope of this survey, I discuss this literature only to the extent that it offers new insights with respect to the economics of multinational firms (see Caves 1996 for a relatively recent survey of this literature). I focus on the central conclusions of this line of research, particularly those that relate to technology transfer.

Markusen and Maskus (1999) suggest that the literature that attempts to link the emergence of multinational firms with firm- and country-level characteristics can be understood as emerging from a common underlying model—the knowledge capital model. Research that deals directly with technology transfer includes Horstmann and
Markusen (1987, 1996) and Ethier and Markusen (1991). Markusen (1998) argues that the knowledge capital model rests on the fact that knowledge has a public good property, that is, it can be utilized in multiple locations simultaneously. Thus any innovation can then be fruitfully applied at multiple plants dispersed all over the world, giving rise to horizontal multinational firms. Markusen and Maskus (1999) show that there is indeed strong empirical support for this horizontal model of multinationals.

How does the knowledge capital model explain internalization? Once again, the public good nature of knowledge occupies a central role. If licensees (or local partners under a joint venture) can get access to the multinational’s proprietary knowledge, the value of its knowledge based assets can be dissipated either because of increased competition (Ethier and Markusen 1991; Markusen 2000; Saggi 1996, 1999) or because the local partner has inadequate incentives to protect the multinational’s reputation (Horstmann and Markusen 1987). The incentive to prevent the dissipation of knowledge-based assets is reflected in the fact that multinationals transfer technologies of new vintage through direct investment and license or transfer their older technologies through joint ventures (see Mansfield and Romeo 1980). Alternatively, it may be easier to trade older technologies through the market; potential buyers are likely to be better informed about well-established technologies compared with new ones.

In an empirical paper, Smarzynska (1999a) focuses on intraindustry differences in R&D intensity as a determinant of the mode of entry chosen by firms investing in Eastern European countries. Like past work, this study finds that a firm’s R&D expenditure is negatively related to the probability of a joint venture and positively related to direct entry. Furthermore, a firm’s R&D expenditure relative to the rest of the industry is positively correlated with the probability of greenfield entry in high-technology sectors. In low-technology sectors, higher relative R&D expenditure by a firm actually increases the likelihood of a joint venture rather than a greenfield entry. Thus, a firm’s R&D expenditure relative to other firms in the industry and the aggregate R&D expenditure of the industry relative to other industries may interact in subtle ways to influence the choice between alternative entry modes. Smarzynska (1999a) argues that protecting technology is of greater concern in high-tech industries, thereby encouraging technological leaders to adopt direct entry. However it is also possible that in industries characterized by a fast pace of technological change, any technology leakage will hurt a firm for only a short period of time. Furthermore, the formation of joint ventures may be easier in relatively mature host industries because they can more easily find suitable local partners. Thus Smarzynska’s results call for a careful interpretation but raise some interesting possibilities and questions.

Foreign firms may not be the only ones that have valuable information that is subject to the risk of dissipation. Horstmann and Markusen (1996) argue that a potential licensee in the host country may have better information about local demand and could use this information to extract rents from the licensor. Such agency costs can also be utilized to explain the dynamics of optimal entry modes. In his stud-
ies of British multinationals, Nicholas (1982, 1983) finds that 88 percent of the firms sold their products through a contract with a local agent before converting to directly owned sales or production branches. Furthermore, the decision to terminate the licensing arrangement was based on a desire to avoid agency costs. Once the multinational had acquired the information it needed through its alliance with the local partner, continuing the agency relationship was no longer attractive. Similarly, in their survey of Japanese multinationals in Australia, Nicholas and others (1994) find that 60 percent of the firms used a local agent before making a direct investment and 69 percent exported to Australia before making a direct investment of any sort. Such temporary licensing could be viewed as a method of information acquisition on the part of the foreign firm, as opposed to the local firm seeking superior production technology.

In Horstmann and Markusen’s (1996) model, when the multinational firm’s fixed costs of investment are high relative to the agent’s and there is risk of large losses due to low demand, the multinational opts for an initial licensing contract that becomes permanent ex post in case of low demand. Their analysis can be applied to examine the choice between a joint venture and a wholly owned subsidiary, except that cost uncertainty may be more relevant than demand uncertainty for this scenario. For example, if the productivity of foreign labor is in doubt, forming a joint venture may present a low (fixed) cost option. If labor productivity turns out to be high, an acquisition of the foreign partner may be optimal ex post, resulting in the establishment of a wholly owned subsidiary.

However, dynamic issues remain underexplored in the literature. Although the comparative statics of the models provide some partial intuition about forces that are important for dynamic choice, such an approach is a poor substitute for explicit dynamic modeling. Several central questions deserve further research. For example, what determines the sequencing pattern of different activities? Do firms first form joint ventures and then proceed with direct investment? If so, why? To what extent do the dynamic choices of foreign firms result from their efforts at restricting diffusion of their own technology while at the same time maximizing the acquisition of valuable information from local firms? Do host country welfare and the rate of technology diffusion depend on the sequencing pattern?

**FDI, Technology Transfer, and Spillovers**

Although the increase in world trade has received significant attention, the role of FDI has played in the explosion of world trade is not often appreciated. Today, intrafirm trade, that is, trade between subsidiaries and headquarters of multinational firms, may account for one-third of total world trade. The importance of FDI can also be gauged from the fact that sales of subsidiaries of multinational firms now exceed
worldwide exports of goods and services. In 1998, the total estimated value of foreign affiliate sales was US$11 trillion, whereas the value of worldwide exports was $7 trillion (U.N. Conference on Trade and Development [UNCTAD] 1999). Thus, FDI is the dominant channel through which firms serve customers in foreign markets.

Historically much of the flows of FDI occurred between industrial countries (much like most intraindustry trade). For example, during 1987–92, industrial countries attracted $137 billion of FDI inflows a year on average; developing economies attracted only $35 billion, or slightly more than 20 percent of global FDI inflows. Yet developing economies have become increasingly important host countries for FDI, especially because of the large-scale liberalization undertaken by formerly closed economies, such as China. During 1996 and 1997, over 40 percent of global FDI flows went to developing economies (UNCTAD 1999).

Figure 2 plots net FDI inflows as a percentage of gross domestic investment for low- and middle-income countries (those countries with per capita income below $9,655) for 1975–96. During this period, FDI became an increasingly important source of capital for such countries. On average, FDI inflows now constitute approximately 10 percent of their annual gross domestic investment.

Figure 2. Net Inflows of FDI over Gross Domestic Investment, Low- and Middle-Income Countries, 1975–96

The recent surge in capital flows to developing economies, of which FDI has been a significant part, is also reflected in the fact that approximately 30 percent of the total stock of FDI is in developing economies (table 2). FDI is of relatively greater importance to developing economies because of their smaller size. In 1997, the total inward stock of FDI as a percentage of GDP was almost 17 percent in developing countries, compared with less than 12 percent in industrial countries (table 2).

For the purposes of this article, the role of FDI as a channel for transferring goods and services internationally is of secondary concern. Instead, the main issue is the role of FDI as a channel of technology transfer. It is difficult to find fully convincing evidence of the dominance of FDI as a channel of international technology transfer (among those channels that directly involve the owner of the technology being transferred). However, several facts hint in that direction. For example, in 1995 over 80 percent of global royalty payments for international transfers of technology were made from subsidiaries to their parent firms (UNCTAD 1997). In general, technology payments and receipts have risen steadily since the mid-1980s, reflecting the importance of technology for international production.

Table 3 reports data for Germany, Japan, and the United States. For example, from 1985 to 1997, Germany’s receipts of royalties and license fees increased from approximately $0.5 billion to more than $2 billion. For the United States, the increase was even sharper, from $6 billion to more than $33 billion. The data also indicate the importance of FDI for international trade in technology. During 1985–97, between two-thirds and nine-tenths of technology flows were intrafirm in nature. Furthermore, as is evident from table 3, the intrafirm share of technology flows has increased over time. Of course, royalty payments only record the explicit sale of technology and do not capture the full magnitude of technology transfer through FDI relative to technology transfer via imitation, trade in goods, and other channels.

Yet another confirmation of the strong role FDI plays in transmitting technology internationally comes from the interindustry distribution of FDI. It is well known that multinational firms are concentrated in industries that exhibit a high ratio of R&D relative to sales and a large share of technical and professional workers (Markusen

<table>
<thead>
<tr>
<th>Table 2. FDI Inward Stock, 1987–98</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Billions of dollars</strong></td>
</tr>
<tr>
<td>Percent in industrial countries</td>
</tr>
<tr>
<td><strong>Percentage of GDP</strong></td>
</tr>
<tr>
<td>World</td>
</tr>
<tr>
<td>Industrial countries</td>
</tr>
<tr>
<td>Developing economies</td>
</tr>
</tbody>
</table>

Source: UNCTAD (2000).
In fact, it is commonly argued that multinationals rely heavily on intangible assets, such as superior technology, for successfully competing with local firms that are better acquainted with the host country environment.

By encouraging FDI, developing economies hope not only to import more efficient foreign technologies but also to generate technological spillovers for local firms. Not surprisingly, a large body of literature tries to determine whether host countries enjoy spillovers from FDI. It is important to be clear about the meaning of the word *spillover*. A distinction can be made between pecuniary externalities (that result from the effects of FDI on market structure) and other pure externalities (such as the facilitation of technology adoption) that may accompany FDI. A strict definition of spillovers would count only the latter, and this is the definition employed here. In other words, if FDI spurs innovation in the domestic industry by increasing competition, I do not view that as a spillover from FDI because this effect works its way through the price mechanism. However, it is difficult to isolate empirically the pure externalities from FDI from its other effects that work through the market. Furthermore, policy ought to be based on the aggregate effect of FDI on welfare, not just on the extent of positive externalities from FDI.

The central difficulty is that spillovers, as defined here, do not leave a paper trail: they are externalities that the market fails to take into account. Nevertheless, sev-

### Table 3. Receipts of Royalties and License Fees, 1985–97 (millions of dollars)

<table>
<thead>
<tr>
<th>Year</th>
<th>Germany</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Intrafirm</td>
</tr>
<tr>
<td></td>
<td>German parent firms only</td>
<td>Foreign affiliates in Germany</td>
</tr>
<tr>
<td>1985</td>
<td>546</td>
<td>464</td>
</tr>
<tr>
<td>1986</td>
<td>780</td>
<td>597</td>
</tr>
<tr>
<td>1987</td>
<td>997</td>
<td>698</td>
</tr>
<tr>
<td>1988</td>
<td>1,081</td>
<td>883</td>
</tr>
<tr>
<td>1989</td>
<td>1,122</td>
<td>899</td>
</tr>
<tr>
<td>1990</td>
<td>1,547</td>
<td>1,210</td>
</tr>
<tr>
<td>1991</td>
<td>1,515</td>
<td>—</td>
</tr>
<tr>
<td>1992</td>
<td>1,680</td>
<td>—</td>
</tr>
<tr>
<td>1993</td>
<td>1,596</td>
<td>—</td>
</tr>
<tr>
<td>1994</td>
<td>1,720</td>
<td>—</td>
</tr>
<tr>
<td>1995</td>
<td>2,174</td>
<td>1,486</td>
</tr>
<tr>
<td>1996</td>
<td>2,315</td>
<td>1,667</td>
</tr>
<tr>
<td>1997</td>
<td>2,282</td>
<td>1,659</td>
</tr>
</tbody>
</table>

— Not available.

*Source: UNCTAD (1999).*
eral studies have attempted the difficult task of quantifying spillovers. But what are the potential channels through which such they may arise? A more difficult question is whether it is even reasonable to even expect spillovers to occur from FDI. Multinationals have much to gain from preventing the diffusion of their technologies to local firms (except when technologies diffuse vertically to potential suppliers of inputs or buyers of goods and services sold by multinationals).

**Potential Channels of Spillovers**

At a general level, the literature suggests the following potential channels of spillovers:

- **Demonstration effects.** Local firms may adopt technologies introduced by multinational firms through imitation or reverse engineering.
- **Labor turnover.** Workers trained or previously employed by the multinational may transfer important information to local firms by switching employers or may contribute to technology diffusion by starting their own firms.
- **Vertical linkages.** Multinationals may transfer technology to firms that are potential suppliers of intermediate goods or buyers of their own products.

**Demonstration effects.** In its simplest form, the demonstration effect argument states that exposure to the superior technology of multinational firms may lead local firms to update their own production methods. The argument derives from the assumption that it may simply be too costly for local firms to acquire the necessary information for adopting new technologies if they are not first introduced in the local economy by multinationals (and hence demonstrated to succeed in the local environment). Incidentally, the demonstration effect argument relates well to the point made by Parente and Prescott (1994) that trade may lower costs of technology adoption.

Clearly, geographical proximity is a vital part of the demonstration effect argument. As noted earlier, empirical evidence on the geographical scope of R&D spillovers is mixed. However, studies that reach optimistic conclusions with respect to the international nature of R&D spillovers typically involve data from industrial countries and therefore require qualification. Geographical proximity may indeed be crucial for developing economies that are not as well integrated into the world economy and that have fewer alternative channels for absorbing technologies.

The main insight of the demonstration effect argument is that FDI may expand the set of technologies available to local firms. If so, this is a potential positive externality. However, a mere expansion in choices need not imply faster technology adoption, especially if domestic incentives for adoption are also altered due to the impact of FDI on market structure. FDI may expand choices, but it also generally increases competition. The industrial organization literature on market structure and inno-
vation does not provide an unambiguous answer to this question. A rough conclusion is that a monopolist has a stronger incentive to invest in R&D that yields innovations that complement existing technology, whereas competitive firms have a stronger incentive to invest in R&D that yields innovations that replace existing technology. The net effect on the incentives for adopting new technologies may indeed be ambiguous.

Suppose FDI does lower the cost of technology adoption and lead to faster adoption of new technologies by local firms. Does that imply that relative to trade (that is, a scenario in which foreign firms export to the domestic or world market), inward FDI necessarily generates spillovers for the local economy? A point to keep in mind is that technology diffusion may strengthen the competitors of the foreign firms. Foreseeing the consequences of such diffusion, foreign firms may alter the very terms of their original technology transfer. For example, a foreign firm may choose to transfer technologies of lower quality when there is a risk of leakage to local firms. It is conceivable, however, that due to their larger size and other advantages they enjoy in the product market, multinationals can alter the market outcome in their favor despite technology leakage. Thus, a multinational would not have to resort to such strategies.

For example, Das (1987) presents a model in which native firms may learn from the subsidiary of a multinational firm that acts as a dominant firm facing a local competitive fringe in the product market. Wang and Blomström (1992) present a related model. In their duopoly model with differentiated goods, a multinational transfers technology to its subsidiary so that the local firm can learn from the new technologies introduced. Learning occurs both through costless technology spillovers as in the contagion effects that Findlay (1978) first emphasized, as well as through the local firm’s costly investments. The most interesting implication of Wang and Blomström’s model is that technology transfer through FDI is positively related to the level of the local firm’s investment in learning. This result suggests that multinationals respond to local competition by introducing newer technologies faster.

Assuming the rate of increase in efficiency of the local firms to be positively related to the scale of operation of the multinational firm’s subsidiary, Das (1987) investigates the optimal time paths of the multinational’s output and price. She shows that despite technology leakage, the multinational may find it profitable to transfer technology. Huizinga (1995) models a multinational’s incentive for technology transfer where it faces risk of competition caused by expropriation by the government of the host country. The main result is that the multinational responds by lowering the quality of technology transfer even when such transfers are costless.

Because the demonstration effect argument is largely an industry-level argument, relating industry-level variation in R&D expenditures by local firms to the extent of FDI is one method of checking whether local adoption efforts are encouraged through FDI. Of course, such an exercise would have to control for the effect of FDI on market...
structure, and this seems rather difficult. To the best of my knowledge, a convincing empirical exercise of this type has not been performed. Instead, the existing literature has focused on the effects of FDI on TFP in local firms.

**Labor turnover.** Although researchers have extensively studied direct imitation and reverse engineering as channels of interfirm technology diffusion, they have tended to neglect the role of labor turnover. Labor turnover differs from the other channels because knowledge embodied in the labor force moves across firms only through the physical movement of workers. The relative importance of labor turnover is difficult to establish because it would require tracking individuals who have worked for multinationals, interviewing them regarding their future job choices, and then determining their impact on the productivity of new employers. Few empirical studies attempt to measure the magnitude of labor turnover from multinationals to local firms. To the best of my knowledge, there are no empirical studies that attempt to measure the role such turnover plays in improving productivity in local firms.

The available evidence on labor turnover itself is mixed. For example, although Gershenberg’s (1987) study of Kenyan industries finds limited evidence of labor turnover from multinationals to local Kenyan firms, several other studies document substantial labor turnover from multinationals to local firms. UNCTAD (1992) discusses the case of the garment industry in Bangladesh (see also Rhee 1990). Korea’s Daewoo supplied Desh (the first Bangladeshi firm to manufacture and export garments) with technology and credit. Thus, Desh was not a multinational firm in the strict sense; rather, it was a domestic firm that benefited substantially from its connection with Daewoo. Eventually, 115 of the 130 initial workers left Desh to set up their own firms or to join other newly established garment companies. The remarkable speed with which the former Desh workers transmitted their know-how to other factories clearly demonstrates the role labor turnover can play in technology diffusion.

Pack (1997) discusses evidence documenting the role of labor turnover in disseminating the technologies of multinationals to local firms. For example, in the mid-1980s, almost 50 percent of all engineers and approximately 63 percent of all skilled workers that left multinationals left to join local Taiwanese firms. By contrast, Gershenberg’s (1987) study of Kenyan industry reports smaller figures: of the 91 job shifts studied, only 16 percent involved turnover from multinationals to local firms.

To synthesize these empirical findings, the cross-country variation in labor turnover rates itself requires an explanation. One possible generalization is that in areas such as Korea and Taiwan, local competitors are less disadvantaged relative to their counterparts in many African economies, thereby making labor turnover possible. Thus the ability of local firms to absorb technologies introduced by multinationals may be a key determinant of whether labor turnover occurs as a means of technology diffusion in equilibrium. Furthermore, the local investment climate may be such
that workers looking to leave multinationals in search of new opportunities (or other local entrepreneurs) find it unprofitable to start their own companies, implying that the only alternative opportunity is to join existing local firms. The presence of weak local competitors probably goes hand in hand with the lack of entrepreneurial efforts because both may result from the underlying structure of the economic environment.

Glass and Saggi (forthcoming) argue that because superior technology is one of the key intangible assets that permit multinationals to successfully compete with local firms, multinationals may attempt to limit technology diffusion by offering higher wages to their workers relative to local firms. Thus, the wage premiums paid by a multinational can provide a rough estimate of the value it places on the knowledge it transfers to its workers. The more interesting point is that such a premium may either exceed or fall short of the benefit the local economy would enjoy if the multinational were to allow its workers to leave. Note that if the multinational must raise wages to restrict technology diffusion to local firms, the wage premium might not be related to the social value of the knowledge embodied in the workers. Thus, technology diffusion is not necessarily optimal for the local economy. Policies designed to encourage technology transfer do not always increase the welfare of the recipient country.

Local competition policy may also affect labor turnover. For example, Bulgaria’s competition law does not permit an individual to join the management of a competing firm operating in the same line of business as the person’s original employer for the first three years after leaving an enterprise (Hoekman and Djankov 1997). Of course, in many industrial countries, trade secrets laws protect firms against the loss of valuable information to their rival firms. But it is difficult to see how such laws could protect against the kind of basic technology diffusion that results from labor turnover from multinationals in developing economies.

Labor turnover rates may vary at the industry level as well. Casual observation suggests that industries with a fast pace of technological change (such as the computer industry in Silicon Valley) are characterized by very high turnover rates relative to more mature industries. Therefore, cross-country variation in labor turnover from multinationals could simply stem from the global composition of FDI: developing economies are unlikely to host FDI in sectors subject to rapid technological change.

**Vertical linkages.** For quite some time, analysts have recognized that multinationals may benefit the host economy through the backward and forward linkages they generate. However, merely documenting extensive linkages between multinationals and local suppliers or buyers is insufficient for arguing that net benefits accrue to the local economy as a result of FDI. Rodriguez Clare (1996) develops a formal model of linkages and shows that multinationals improve welfare only if they generate linkages over and beyond those generated by the local firms they displace. Yet the question of relevance here is whether the generation of linkages is expected to result in technology diffusion. Although analytical modeling of such issues is scarce, there is
limited empirical evidence in support of the view that multinationals are involved in vertical technology transfers (Lall 1980).

Mexico’s experience with FDI is illustrative of how such a process works. In Mexico, extensive backward linkages resulted from FDI in the automobile industry. Within five years of investments by major auto manufacturers there were 300 domestic producers of parts and accessories, of which 110 had annual sales of more than $1 million (Moran 1998). Foreign producers also transferred industry best practices, zero defect procedures, and production audits to domestic suppliers, thereby improving their productivity and the quality of their products. As a result of increased competition and efficiency, Mexican exports in the automobile industry boomed. Thus, although direct competitors of multinational firms may not realize technological benefits (as evidenced by Aitken and others 1996), suppliers of intermediate goods are likely to benefit substantially.

Even more interesting is the possibility that such vertical transfers when accompanied by spillovers may lead to interaction between upstream and downstream multinational firms that encourages industrial development. Markusen and Venables (1999) develop a model that abstracts from technology spillovers but focuses on the pecuniary externalities that accompany vertical linkages and result in industrial development. Pack and Saggi (2001) emphasize that downstream buyers in industrial country markets benefit from technology diffusion among potential suppliers in developing economies because such diffusion increases competition among suppliers. In their model, by increasing demand in the downstream market, competition caused by technology diffusion in the developing market may induce entry into marketing, thereby increasing competition in the downstream market. Their analysis implies that fully integrated multinational firms may be more averse to technology diffusion than are firms involved in international arm’s-length arrangements. Although they do not model FDI, it is not hard to see how their model can be applied to understand the consequences of technology diffusion under FDI rather than exporting.

**Empirical Evidence on Spillovers**

Early efforts in search of spillovers from FDI proceeded by relating the interindustry variation in productivity to the extent of FDI (Blomström 1986; Blomström and Persson 1983; Caves 1974; Globerman 1979). These studies largely find that sectors with a higher level of foreign involvement (as measured by the share of the labor force in the industry employed by foreign firms or the extent of foreign ownership) tend to have higher productivity, higher productivity growth, or both. The fact that these studies involve data from different countries (Australia for the Caves study, Canada for Globerman, and Mexico for Blomström) lends a strong degree of robustness to this positive correlation between the level of foreign involvement and local productivity at the sector level.
Of course, correlation is not causation and, as noted by Aitken and Harrison (1999), this literature may overstate the positive impact of FDI on local productivity. Investment may have been attracted to the more productive sectors of the economy instead of being the cause of the high productivity in such sectors. In other words, the studies ignore an important self-selection problem. Both trade and FDI help ensure an efficient allocation of global resources by encouraging investment in those sectors in which an economy enjoys comparative advantage. In this sense, Aitken and Harrison’s point is almost necessarily implied by traditional trade theory. However, if trade protection encourages investment in sectors in which a host economy does not enjoy comparative advantage, trade protection may be welfare-reducing. This possibility was relevant for countries that sought to industrialize by following a strategy of import substitution.

Nevertheless, only plant-level studies can control for the self-selection problem that may plague industry-level studies. Taking the argument a step further, the self-selection problem may also arise in plant-level studies: the more productive plants may be the ones that attract foreign investment. For example, Clerides and others (1998) find support in favor of the self-selection hypothesis in the context of exporting. They find that the more productive firms self select into exporting. However, if plant-level studies fail to find a significant relationship between foreign involvement and productivity, the self-selection problem might not be important. It might be important if foreign firms seek out plants with low productivity and bring them up to par with more efficient local plants. In this case, there might be no significant productivity differential between foreign and local firms. This argument seems far-fetched, but it could make sense. Suppose local plants with very low productivity are relatively undervalued by local agents because the skills (technology and modern management) needed to make them competitive are in short supply locally. In this scenario, such plants would be attractive to foreign investors who can, through their technology, generate productivity improvements that simply cannot be achieved by local agents.

What do empirical plant-level studies find with respect to spillovers from FDI? Haddad and Harrison’s (1993) study was the first to employ a comprehensive data set at the level of the individual firm over several years. The data came from an annual survey of all manufacturing firms in Morocco. An important result was that foreign firms exhibited higher levels of TFP, but their rate of TFP growth was lower than that for domestic firms. As the authors note, at first glance, such a finding suggests that perhaps there was some sort of convergence between domestic and foreign firms. However, this was not the case. Although there was a level effect of foreign investment on the TFP of domestic firms, such an effect was missing for the growth rate of the TFP of domestic firms. In addition, when sectors were divided into high and low tech, the effect of FDI at the sector level was found to be more positive in low-tech sec-
The authors interpret this result as indicative of the lack of absorptive capacity on the part of local firms in the high-tech sector, where they may be further behind multinationals and unable to absorb foreign technology.

Aitken and others (1996) undertake a somewhat different approach to measuring spillovers from FDI. The idea behind their study is that technology spillovers should increase the marginal product of labor, and this increased productivity should show up in the wages for workers. The study employs data from manufacturing firms in Venezuela, Mexico, and the United States. For both Mexico and Venezuela, a higher share of foreign employment is associated with higher overall wages for both skilled and unskilled workers. Furthermore, royalty payments to foreign firms from local firms are highly correlated with wages. Most important, the study finds no positive impact of FDI on the wages of workers employed by domestic firms. In fact, the authors report a small negative effect for domestic firms, whereas the overall effect for the entire industry is positive. These findings differ from those for the United States, where a larger share of foreign firms in employment is associated with both a higher average wage as well as higher wages in domestic establishments.

Putting the Aitken and others (1996) findings into the context of previous work, it is clear that wage spillovers (from foreign to domestic firms) are associated with higher productivity in domestic plants. Conversely, the absence of wage spillovers appears to accompany the existence of productivity differentials between domestic and foreign firms. Why might this be so? Any serious explanation of this association requires studying the interaction of the market for labor and goods. Glass and Saggi (1999b) develop a model to capture this interaction (their findings are discussed in the section on labor turnover).

The most recent study on the issue of spillovers from FDI is Aitken and Harrison (1999). This study uses annual census data on more than 4,000 Venezuelan firms. Because each plant was observed over a period of time, the self-selection problem of previous sector-level studies (that is, FDI goes to the more productive sectors) could be avoided. The authors find a positive relationship between foreign equity participation and plant performance, implying that foreign participation does indeed benefit plants that receive such participation. However, this own-plant effect is robust for only small plants, that is, those plants that employ fewer than 50 employees. For larger plants, foreign participation results in no significant improvement in productivity relative to domestic plants.

More interestingly, productivity in domestic plants declines when foreign investment increases. In other words, the authors find evidence of negative spillovers from FDI and suggest that they could result from a market stealing effect. That is, foreign competition may have forced domestic firms to lower output and thereby forgo economies of scale. Note that if loss in output is large enough, local plants may have lower productivity despite enjoying some sort of technology spillovers. Nevertheless, on
balance. Aitken and Harrison (1999) find that the effect of FDI on the productivity of the entire industry is weakly positive. They also note that similar results are obtained for Indonesia, except that the positive effect on its own plants is stronger, whereas the negative effect on domestic plants is weaker, suggesting a stronger overall positive effect.

Djankov and Hoekman (2000) also find a negative spillover effect of FDI on purely domestic firms in industry in the Czech Republic. Interestingly, however, when joint ventures are excluded from the sample and attention is restricted to the impact of majority-owned foreign affiliates (that is, FDI) on all other firms in an industry (including joint ventures), the negative effect loses statistical significance. The authors report that survey questionnaires reveal that joint-venture firms invest significantly more in new technologies than do purely domestic firms. The authors suggest that purely domestic firms might lack the ability to absorb the technologies introduced by foreign firms (due to their lower R&D efforts).

Overall, several studies have cast doubt on the view that FDI generates positive spillovers for local firms. But such findings need not imply that host countries have nothing significant to gain (or must lose) from FDI. The point is that the reallocation of resources that accompanies the entry of foreign firms may not be immediate. Domestic firms should be expected to suffer from an increase in competition; in fact, part of the benefit of inward FDI is that it can help weed out relatively inefficient domestic firms. Resources released in this process will be put to better use by foreign firms with superior technologies, efficient new entrants (domestic and foreign), or some other sectors of the economy. Existing studies of spillovers may not cover a long enough period to be able to accurately determine how FDI affects turnover rates (entry and exit). Furthermore, their design limits such horizontal studies because they cannot clarify linkages and spillovers that may result from FDI in industries other than the one in which FDI occurs.

**Spillovers from FDI: A Recapitulation**

A challenge facing the optimistic view regarding technology spillovers from FDI is to explain how such spillovers can ever be in the interest of the multinational firms. Clearly, under most circumstances, multinationals would rather limit diffusion in the local economy. In fact, the heart of the theory that seeks to explain the emergence of multinationals is that such firms are able to successfully compete with local firms precisely because they possess superior technologies, management, and marketing. Why, then, would multinationals not take actions to ensure that such advantages do not diffuse to local competitors?

Part of the answer must lie in the fact that such actions are costly and may even entail externalities between multinationals. Suppose a costly action (such as litigation in local courts to enforce protection of IPRs) can indeed help limit the loss of
knowledge capital for a multinational. A difficulty arises if all potential multinationals benefit from the curtailment of technology diffusion, whereas the costs fall on only the one who takes legal action. Thus the public good nature of such actions suggests that developing economies hosting multinationals may expect the rivalry among such firms to result in some degree of technology diffusion. Of course, the preceding argument also overstates the case a bit: some loss of knowledge will result despite all precautions. Nevertheless, it is beyond dispute that multinationals can take actions to limit diffusion, and while they are making their decisions regarding where to set up subsidiaries, the expected costs of technology diffusion will enter their calculus of profit maximization.

That being said, the entry of multinationals may indeed benefit host countries even if it fails to result in spillovers for local firms. First, the preceding discussion suggests that spillovers to local firms that directly compete with the multinationals would be the most elusive of benefits that host countries may expect to enjoy from FDI. Second, local agents other than domestic competitors of multinationals (for example, local workers) may enjoy positive externalities from FDI. If so, the total effect of FDI on local welfare may be positive despite the lack of technology spillovers.

Third, spillovers may be of an entirely different nature: local firms may enjoy positive externalities from foreign firms that make it easier for them to export. Such externalities may come about because better infrastructure (transportation, storage facilities, and ports) emerges in regions with a high concentration of foreign exporters. Aitken and others (1997) provide direct evidence on this issue. They conducted a detailed study of 2,104 manufacturing plants in Mexico. In their sample, 28 percent of the firms had foreign ownership and 46 percent of the foreign plants exported. Their major finding is that the probability of a Mexican-owned plant exporting is positively correlated with its proximity to foreign-owned exporting plants. Such spillovers may result from informational externalities and are more likely to lower fixed costs rather than marginal costs of production.

**FDI and Growth**

Regardless of the channel through which technology spillovers occur, the fact that FDI often involves capital inflows along with technology transfer implies that one would expect a positive impact of FDI on growth in the host country. Yet there are several important caveats to this assertion. First, a positive correlation between the extent of FDI and economic growth in cross-country regressions may simply reflect the fact that countries that are expected to grow faster attract FDI because it yields higher returns there. Thus the causation could run from growth to FDI and estimation of a simultaneous equation system may be needed to resolve the issue. Second, multinationals often raise the required capital in the host country, and in such a scenario capital inflows associated with FDI may not be substantial. An optimistic view
of FDI would then look to technology transfer and/or spillovers as the mechanism through which FDI may affect growth. Indeed, Romer (1993) argues that FDI can have a positive effect on growth in developing economies by helping them bridge the idea gap with respect to industrial countries.

Glass and Saggi (1999) examine the question of spillovers from FDI in a product-cycle model. In their North–South model, the demonstration/proximity argument is formalized as follows. Southern firms are assumed to be able to imitate multinationals located in the South at a lower cost than firms located in the North. However, multinational firms are also stronger competitors than firms that produce only in the North because multinationals produce in the same low-wage location as potential imitators. The model delivers the surprising result that a faster flow of FDI increases the aggregate rate of technology transfer to the South only if local firms lack the ability to imitate firms located in the North (that is, if geographical proximity is a prerequisite for imitation). If firms in the North can be imitated, FDI does not alter aggregate technology transfer because imitation focusing on firms located in the North slows down with a hastening of imitation targeting multinationals.

Although the internalization question is a central one in the theory of FDI, almost all theories of FDI and licensing have been developed in either static or partial equilibrium models. A few dynamic general equilibrium models explore the effect of FDI on growth, but these models have ignored the possibility of licensing. Glass and Saggi (2002b) develop a model of FDI that captures the internalization decision and its implications for both the rate and magnitude of innovation. They also examine how policy interventions (taxes or subsidies to FDI) that alter the incentive to internalize production within the firm affect economic growth. They find that the ability of firms to switch modes from licensing to FDI in response to policy changes is vital for ensuring that a subsidy to FDI leads to faster economic growth.

In a comprehensive paper, Borensztein and others (1998) utilize data on FDI flows from industrial countries to 69 developing economies to test the effect of FDI on growth in a cross-country regression framework. Their findings are as follows. First, FDI contributes more to domestic growth than domestic investment, suggesting that it is indeed a vehicle of technology transfer. Second, FDI is more productive than domestic investment only when the host country has a minimum threshold stock of human capital. The latter finding is especially interesting because it clarifies when exactly FDI should be expected to effect growth.

Using cross-section data from 46 developing economies, Balasubramanyam and others (1996) also investigate the effect of FDI on growth in developing economies. They report two main findings. First, the growth-enhancing effects of FDI are stronger in countries that pursue a policy of export promotion rather than import substitution, suggesting that the trade policy regime is an important determinant of the effects of FDI. Second, they find that, in countries with export-promoting trade regimes, FDI has a stronger effect on growth than domestic investment. Both find-
ings relate well to the results of Borensztein and others (1998). The second finding may be viewed as a confirmation of the hypothesis that FDI results in technology transfer.

The findings of Borensztein and others (1998) relate well to Keller (1996), who argues that mere access to foreign technologies may not increase the growth rates of developing economies. In his model, if a country’s absorptive capacity (stock of human capital) remains unchanged, a switch to an outward orientation does not lead to a higher growth rate. Using a model quite different from Keller’s, Glass and Saggi (1998) focus on the issue of the quality of technology transferred through FDI. They argue that investment in imitation by host country firms generates the necessary knowledge (or skill) foundation for FDI, and thus factors that promote imitation can promote a higher-quality mix of FDI. Keller’s (1996) model stresses that a country’s stock of human capital effectively constrains its ability to take advantage of foreign technologies; Glass and Saggi (1998) emphasize that indigenous technological capability in an industry effectively constrains a country’s ability to host foreign technology. Thus, they take a more micro-level view of the constraints on technology transfer relative to Keller (1996), although both studies make similar points. For example, a country may have a fair amount of human capital in the aggregate but may lack the technological sophistication to be able to host high-quality FDI in any particular industry.

Xu (2000) provides yet another confirmation of the argument that, in the absence of adequate human capital, technology transfer from FDI may fail to increase productivity growth in the host country. Using data on outward FDI from the United States to 40 countries, Xu finds that technology transfer from FDI contributes to productivity growth in more developed countries but not in less developed economies because the latter lack adequate human capital. Incidentally, as Xu notes, FDI may contribute to productivity growth due to reasons other than technology transfer. Thus a statistically significant coefficient on some measure of FDI in a productivity growth equation does not necessarily imply that technology transfer is the mechanism through which FDI contributes to productivity growth.

Xu (2000) measures the technology transfer intensity of multinational firms affiliates by their spending on royalties and license fees as a share of their gross output and estimates that, of the total effect of trade (through R&D spillovers) and FDI (through technology transfer) on productivity growth in industrial countries, 41 percent is due to technology transfer. These results for industrial countries confirm the findings of Barrell and Pain (1997), who find that FDI has a positive impact on technological change in Germany and the United Kingdom. Xu and Wang (2000) find that although capital goods trade serves as a channel of technology transfer among industrial countries, bilateral flows of FDI do not. However, Xu and Wang (2000) raise questions regarding these results because of the poor quality of the FDI data.
The Role of Policy

What does the literature say about the role policy plays in the process of international technology transfer? There is a large range of policies; this section focuses on policies on trade, FDI, and protection of IPRs.

Trade Policy

Although the literature on trade policy is voluminous, it does not pay significant attention to the interaction between protection and technology transfer. In fact, most models treat the process of technology transfer in a rather rudimentary way, focusing instead on other aspects of the problem. Here I discuss a few prominent examples of this line of research.

Miyagiwa and Ohno (1995) examine a domestic firm’s incentives for technology adoption when a foreign rival has already adopted a superior technology. They assume that the cost of adoption decreases over time, and they examine how the nature (tariff versus quota) and the duration (temporary versus permanent) of trade protection influence the domestic firm’s incentives for technology adoption. Their most interesting result is that temporary protection (that is, protection that is removed on successful adoption by the domestic firm) actually delays the date of technology adoption. In a related paper, Miyagiwa and Ohno (1999) show that if temporary protection is credible, it may indeed increase R&D relative to free trade. However, if the domestic firm expects that protection will be removed early should innovation occur before the preannounced terminal date of protection, the firm will invest less in R&D under protection relative to free trade. Similarly, as first emphasized by Matsuyama (1990), if the domestic firm expects protection to be extended in case of no innovation by the terminal date, its investment incentives are marred by protection.

The literature also investigates the effect of trade protection in R&D-based models of endogenous growth (see Grossman and Helpman 1991, 1995). As expected from models in which increasing returns, imperfect competition, and externalities play a central role, the results depend on the details of a particular model and require careful interpretation. To the extent that one can draw a general conclusion from such a complex body of literature, it would be that the literature does not provide an unconditional argument against trade protection. The conclusions hinge dramatically on the scope of knowledge spillovers: international knowledge spillovers strongly tilt the balance in favor of free trade, whereas national spillovers create a role for policy intervention that can combat path dependence resulting from a historical accident. For example, if productivity improvements depend only on a country’s own R&D, a case can be made for policies that ensure that industries in which such improvements occur at a rapid rate are not all located elsewhere.
Dinopoulos and Segerstrom (1999) develop a specific-factor variant of the quality ladders model of endogenous growth without scale effects. They examine the consequences of contingent protection, that is, tariffs imposed on imports whenever domestic firms lose their technological leadership to foreign firms who successfully innovate over them. Their approach is interesting because protection in the real world is usually not marginal (for example, antidumping duties may be levied on foreign firms with the explicit goal of providing sufficient relief to domestic industry). Somewhat interestingly, Dinopoulos and Segerstrom find that tariffs that allow domestic firms to capture the domestic market are positively related to the global rate of technological change in the short run.

Grossman and Helpman (1991) also analyze the effects of tariff protection in a two-country quality ladders model. Unlike Dinopoulos and Segerstrom, Grossman and Helpman analyze only tariffs that are too small to allow domestic firms to capture the market. Both models assume Bertrand competition on the product market, so that a low-quality firm can monopolize the market only if a tariff of sufficient magnitude is imposed on higher-quality imports. A small tariff can extract rents from foreign firms but fails to protect domestic firms that have been innovated over by foreign firms. It should be noted that Dinopoulos and Segerstrom’s (1999) analysis assumes that both countries adopt symmetric policies.

**FDI Policy**

There is no simple way of describing the policy environment that faces multinationals in developing economies. In countries that historically emphasized import substituting industrialization—such as most of Africa, Latin America, and Southeast Asia—FDI was either completely prohibited or multinational firms had to operate under severe restrictions. In fact, even where technology acquisition was a major policy objective, multinationals were rarely permitted to operate wholly owned subsidiaries; Japan, Korea, and Taiwan all imposed restrictions on FDI at various points in time. In other words, “outward-oriented” economies were not particularly keen on allowing multinational firms into their markets. Japan’s Ministry of International Trade and Investment (MITI) played an active role in the country’s acquisition of foreign technology. MITI limited competition between potential Japanese buyers, did not allow inward FDI until 1970, never greatly liberalized FDI, and even sometimes insisted that foreign firms share their technology with local firms as a precondition for doing business in Japan. Ozawa (1974) provides a rich description of the role imported technology and local R&D (aimed at facilitating absorption of foreign technology) played in Japan’s economic development.

In contrast to the restrictive stance toward FDI, licensing of foreign technology was aggressively encouraged (Layton 1982). Korea’s experience has been quite similar to that of Japan. For example, annual inflows of licensed technology increased steadily during the 1970s and 1980s. FDI inflows into Korea, which were always relatively
low, stagnated during 1978–83, but annual inflows of licensed technology (as measured by royalty payments) increased steadily during the 1970s and 1980s (Sakong 1993). This slowdown of FDI into Korea was partially a result of restrictive FDI policies instituted by the Korean government during that period (see Hobday 1995 for further details on Korea’s experience).

What is the rationale behind policies that discourage FDI? Pack and Saggi (1997) argue that by prohibiting FDI and placing other restrictions on the conduct of multinationals, government policies in many countries may have effectively weakened the bargaining position of foreign firms. They note that in Japan, MITI restricted many local firms from participating as potential buyers exactly for this reason.

Sometimes policy has also favored licensing and joint ventures relative to wholly owned subsidiaries of multinationals. For example, the Chinese government has been particularly interventionist in technology transactions and has encouraged FDI in the form of joint ventures. Although wholly owned subsidiaries are not prohibited, the policy environment favors joint ventures over such enterprises. Of course, an immediate reason for this might be that all such policies simply reflect protectionism. Large public firms or hitherto protected private firms may not be able to compete with multinationals and may secure protection through the political process. However, is it also possible that joint ventures (as well as technology licensing) lead to more local involvement and therefore greater technology spillovers to local agents.

Saggi (1999) develops a two-period model in which a foreign firm chooses between FDI and technology licensing. The key assumption is that licensing results in greater transfer of know-how to the local firm than does FDI, under which the local firm must compete with the subsidiary of the multinational firm. The main result is that the local firm would have the strongest incentive for innovation if the foreign firm were to follow initial licensing by direct investment. However, in equilibrium, the foreign firm never adopts such a course of action.

Using plant-level data for 1991 for all Indonesian establishments with more than 20 employees, Blomström and Sjoholm (1999) shed light on two important questions. First, do establishments with minority and majority ownership (that is, joint ventures versus wholly owned subsidiaries) differ in terms of their (labor) productivity levels? Second, does the degree of technology spillovers vary with the extent of foreign ownership? The second question is crucial for the purposes of the present study. Blomström and Sjoholm obtain several interesting results. First, as in many other previous studies, they find that labor productivity is higher in establishments with foreign equity than in purely domestic firms. Second, the extent of total foreign production is positively associated with the productivity of domestic firms, suggesting some sort of spillovers from FDI. Third, the degree of foreign ownership affects neither the productivity of firms that get foreign equity nor the extent of spillovers to the domestic sector.

These findings are puzzling. Clearly, the degree of foreign participation does matter in that plants with no foreign investment are less productive. Perhaps the results
suggest some sort of threshold effects in which beyond a certain degree of foreign ownership additional foreign equity affects neither the productivity of those that receive the investment nor the degree of spillovers to local firms. The authors do not report the minimum level of foreign equity (for those plants that do get foreign equity) in their sample. It is important to keep in mind that the study only measures labor productivity and treats some important endogenous variables as exogenous. Overall, it seems fair to say that the question remains open. Several earlier studies document that technologies transferred to wholly owned subsidiaries are of a newer vintage than licensed technologies or those transferred to joint ventures (Kabiraj and Marjit 1993; Mansfield and Romeo 1980; Saggi 1996).

Djankov and Hoekman (1999) also uncover an interesting role for joint ventures and suggest that such enterprises may have greater ability to absorb foreign technologies than do purely domestic firms. Hoekman and Saggi (2000) suggest that although the motivation behind policies that discriminate between licensing, joint ventures, and establishment of wholly owned subsidiaries is not easy to decipher, a plausible interpretation may be that such policies seek to maximize technology transfer to local firms while limiting the rent erosion that results from the entry of multinational firms.

Another policy issue is that many Southeast Asian countries still do not allow free entry of multinational firms and often express preferences with regard to the type of FDI; that is, entry by Pepsi or Coke is viewed differently than entry by General Motors or Texas Instruments. Unfortunately, the literature provides little insight for understanding such policies. Other than the standard argument that certain industries are able to secure greater protection for themselves, perhaps it may also be the case that spillovers to the local economy are higher under certain types of FDI. For example, it might be that domestic content protection policies involve more local firms and therefore generate greater spillovers. However, there is no formal model or empirical evidence to support this position. In addition, this argument is closely related to the idea of industrial targeting in general, and the pitfalls of the government’s ability to correctly identify high-spillover industries are well known.

Despite the subtle policy interventions outlined, when measured by a broad yardstick, overall government policy has become more liberal across the world. For example, as of 1997, there were 1,513 bilateral investment treaties among countries, compared with 400 seven years earlier (UNCTAD 1998). Economic reform in many formerly communist countries has added to the list of countries vying for FDI. Coupled with this rise in treaties, both industrial and developing economies have a proliferation of fiscal and financial incentives to lure in FDI. Such overly optimistic policies carry dangers of their own and may reduce welfare in host countries. Although a case for such policies can indeed be made on the basis of positive externalities from FDI, there is no convincing evidence on this front.

Barry and Bradley (1997) describe Ireland’s experience with FDI. Both favorable policies (reduced taxes and trade barriers, and investment grants) as well as strong...
fundamentals (such as infrastructure and an educated labor force) seem to have played a role in attracting FDI to Ireland. The strong performance of the Irish economy since the mid-1980s is attributable to both strong fundamentals as well as significant FDI inflows.

An alternative case for the use of FDI incentives can be made on the basis of the oligopolistic nature of the markets within which FDI occurs. For example, consider Mexico’s recent experience with FDI in its automobile industry. Initial investments by U.S. car manufacturers in Mexico were followed by investments not only by Japanese and European car manufacturers but also by firms that made automobile parts and components. As a result, competition in the automobile industry increased at multiple stages of production, thereby improving efficiency. Such a pattern of FDI behavior (that is, investment by one firm followed by investment by others) reflects strategic considerations involved in FDI decisions. Because multinational firms compete in concentrated markets, they are responsive to each other’s decisions. An important implication of this interdependence between competing multinationals is that a host country may be able to unleash a sequence of investments by successfully inducing FDI from one or two major firms.

**Protection of IPRs**

Common sense suggests that if any policy variable should affect international technology transfer, it ought to be the host country’s IPR regime. The theoretical literature has often investigated the effect of IPR enforcement on technology transfer and FDI in several endogenous growth models. Other approaches also exist. For example, in a strategic partial equilibrium model, Vishwasrao (1995) argues that the lack of adequate enforcement of technology transfer agreements may encourage FDI relative to licensing. In her screening model, depending on the type of licensee, licensing may or may not lead to imitation. The tradeoff between FDI and licensing is that FDI avoids the risk of imitation at the expense of higher production costs.

To limit the scope of the discussion, I omit models in which technology transfer does not play a central role. Several of the articles are linked through their use of the two models used intensively by Grossman and Helpman (1991). Before turning to these, I discuss Taylor’s (1994) work because it differs from the other studies in that it employs a model of endogenous technological change with Ricardian features.

In a two-country model, Taylor examines two scenarios: one in which IPR enforcement is symmetric across the two countries (it applies to innovators regardless of country of origin) and one in which it is asymmetric (it protects only domestic innovators). Although Taylor conducts the analysis under the assumption of costless technology transfer and equal productivity in R&D in the two countries, his results hold even when these assumptions are dropped, making it possible to apply them to a North–South setting. A subtle qualification must be made: symmetric versus asym-
metric treatment implies both countries adopting one policy as opposed to another. Taylor’s model does not analyze incentives for unilateral adoption of a symmetric policy. His major result is that asymmetric protection of IPR distorts the pattern of world trade and lowers the global rate of growth.

Interpreting the exogenous rate of imitation as a proxy for the level of IPR enforcement in the south, Helpman (1993) shows that a decline in the intensity of imitation promotes FDI (with exogenous innovation). Krugman (1979) addresses the issue as well, although his model has a greater degree of exogeneity than does Helpman’s. The major contribution of Helpman’s work lies in providing the first detailed welfare analysis of IPR enforcement in the South (as measured by an exogenous decline in the rate of imitation) in a dynamic general equilibrium growth model. He shows that a strengthening of IPR protection is not in the interest of the South, and that a weak enforcement of IPR protection in the South may even benefit the North, provided the rate of imitation is not too fast. Lai (1998) extends the Helpman model to allow for FDI and shows that innovation is promoted along with FDI if the South strengthens its IPR protection. The common weakness of both models is that stronger IPR enforcement is modeled as an exogenous decline in the rate of imitation. Nevertheless, Helpman’s model is a tour de force in that it clearly specifies the alternative channels through which a strengthening of Southern IPR protection affects Northern and Southern welfare.

Yang and Maskus (2001) study the effects of Southern IPR enforcement on the rate of innovation in the North as well as on the extent of technology licensing undertaken by Northern firms. A key assumption in their model is that increased IPR enforcement increases the licensor’s share of rents and reduces the costs of enforcing licensing contracts, thereby making licensing more attractive. Consequently, both innovation and licensing increase with stronger IPR protection in the South.

Glass and Saggi (2002a) provide an analysis of Southern IPR protection in a comprehensive product-cycle model of trade and FDI. In their model, Southern imitation targets both multinationals producing in the South and purely Northern firms producing in the North. They treat stronger IPR protection as an increase in imitation cost stemming perhaps from stricter uniqueness requirements in the South. In their model, FDI actually decreases with a strengthening of Southern IPR protection because an increase in the cost of imitation crowds out FDI through tighter Southern resource scarcity. Although products like books, videos, and CDs receive a lot of press about conflicts over IPR protection, imitating most products is not so simple (see Pack and Westphal 1986). Empirical evidence indicates that imitation is indeed a costly activity for a wide range of high-tech goods, such as chemicals, drugs, electronics, and machinery. For example, Mansfield and others (1981) find that the costs of imitation average 65 percent of the costs of innovation (and very few products are below 20 percent).

Less efficient imitation absorbs more resources, although the rate of imitation declines with a strengthening of Southern IPR protection. In addition, the contraction
in FDI tightens resource scarcity in the North: increased production leaves fewer resources for innovation, so the rate of innovation falls. It is worth emphasizing that if strengthening Southern IPR protection increases the cost of imitation, targeting both firms producing in the North as well as multinationals producing in the South, Northern incentives for FDI (at the firm level) are basically unaffected.

It should be clear from the discussion so far that the theoretical literature does not give an unambiguous prediction regarding the effects of stronger Southern IPR protection on the extent of FDI and the rate of growth. Does empirical evidence help resolve the issue? The literature largely has not explored the interaction between optimal policies in the two regions (for a recent exception, see Lai and Qiu 1999).

Consider the effect of Southern IPR enforcement on FDI. Surveys of U.S. multinational firms frequently find that such firms are more willing to invest in countries with stronger IPR protection (see Lee and Mansfield 1996). How does the researcher reconcile the ambiguous predictions of the theoretical models with this empirical finding? There are two ways out. First, increased IPR enforcement can be asymmetrical in that firms investing in a country may expect to have a greater influence in local courts relative to those that simply export. Second, imitation of firms located in the North may not be an option for local firms in some developing economies, as is assumed by some theoretical models. In such a scenario, any increase in IPR enforcement by the South will benefit multinational firms, thereby encouraging them to engage in FDI.

As Ferrantino (1993) notes, all of the proceeding models suffer from a fundamental problem: either FDI or licensing is the only channel through which Northern firms are allowed to produce in the South. A more complete treatment of FDI requires that Northern firms be given the option of transacting in technology through the market. What are the consequences of strengthening IPR protection in the South if Northern firms can choose between licensing and FDI? Does FDI increase with IPR enforcement, or does such a change in policy encourage licensing by lowering the risk of opportunism in market transactions? The latter scenario is equally likely, and studies that ignore the possibility of licensing (or joint ventures for that matter) are likely to overstate the effect of IPR enforcement on inward FDI. In fact, a more subtle analysis may be needed. Increased IPR enforcement by the South may indeed make it a more attractive location for production (thereby increasing FDI relative to exports). However, the technologies transferred for that purpose might flow through licensing rather than FDI, so that the net effect on technology transfer through FDI is ambiguous. Of course, aggregate technology transfer to the South may increase, although general equilibrium effects may also require qualifications of this conclusion (Glass and Saggi 2002a).

Using data for 1982 on U.S. exports and sales of overseas affiliates of U.S. firms, Ferrantino (1993) presents a detailed cross-country study that attempts to identify the determinants of both exports and sales of multinational affiliates of U.S. firms, as suggested by the gravity model. His analysis reveals many insights, but perhaps the most interesting finding is that the U.S. firms export more to their affiliates in coun-
tries that have weak IPR regimes. Ferrantino (1993) suggests that this result may reflect attempts by the U.S. firms to limit technology leakage to their rivals abroad by confining production within the United States. This interpretation fits well with a central theme of this survey: multinational firms will adjust their strategies to optimize against policies and market conditions they face in various host countries, casting doubt on the conclusions of empirical (or theoretical) work that treats FDI as given.

Empirical evidence indicates that the level of IPR protection in a country also affects the composition of FDI in two different ways (Lee and Mansfield 1996; Smarzynska 1999b). First, in industries for which IPRs are crucial (pharmaceuticals, for example), firms may refrain from investing in countries with weak IPR protection. Second, regardless of the industry in question, multinationals are less likely to set up manufacturing and R&D facilities in countries with weak IPR regimes and more likely to set up sales and marketing ventures because the latter run no risk of technology leakage.

These studies present useful findings but are unable to address perhaps the most central question of all: does a country’s IPR regime affect its economic growth? Although there are several theoretical analyses of this question, empirical studies are scarce. One such study is Gould and Gruben (1996), who use cross-country data on patent protection, trade regime, and economic fundamentals. They find that IPR protection, as measured by the degree of patent protection, is an important determinant of economic growth. Somewhat more interestingly, they find that the effect of IPR protection is stronger for relatively open economies than it is for relatively closed economies. In other words, a strengthening of IPR protection is more conducive for growth when it is accompanied by a liberal trade policy.

A possible interpretation of this finding is that by increasing foreign competition trade liberalization not only curtails monopoly power granted by IPRs but also ensures that such monopoly power is obtained only if the innovation is truly global. If firms in other countries can export freely to the domestic market and have better products or technologies, a domestic patent is useless in granting monopoly power. Furthermore, note that trade liberalization alone can improve productivity. Using data from Mexican manufacturing firms, Tybout and Westbrook (1995) find that trade liberalization is associated with higher rates of productivity growth. The results of Gould and Gruben (1996) show that IPR enforcement matters over and above trade orientation and that both have mutually reinforcing effects.

Finally, what does the empirical literature tell us about the effect of IPR protection on trade? Theory informs us that asymmetric IPR protection across countries can distort the pattern of world trade; empirical evidence supports this result. Using bilateral trade data for manufactured goods from 22 exporting countries to 71 importing countries, Maskus and Penubarti (1995) find that within the group of large developing economies, the importing country’s strength of IPR protection (as measured by patent rights) exerts a significantly positive effect on bilateral manufacturing imports in many product categories. In other words, in such countries, weak IPR
protection is indeed a barrier to the manufacturing exports of most OECD countries. Maskus (2000) provides an up-to-date discussion of the empirical evidence on the effects of IPR protection on trade and FDI. Smith (1999) updates the study by Maskus and Penubarti (1995), using data on exports of U.S. states to 96 countries. She makes the interesting point that because countries with strong IPR protection also have sophisticated technological capabilities that facilitate local imitation of foreign technologies, within industrial countries there is an ambiguous relationship between strength of IPR protection and the volume of trade.

Conclusions

This article has covered a rather large terrain. This section highlights the eight main points.

First, the role of trade in encouraging growth hinges critically on the geographical scope (national versus international) of knowledge spillovers. As Grossman and Helpman (1995) note, knowledge spillovers are neither exclusively national nor international; they are probably both to some extent. However, spillovers are more likely to be national in scope for developing economies than for industrial ones. Consequently, whether R&D and high-technology production are carried out in close geographical proximity to such countries may indeed matter for their development.

Second, little is known about the relative role of trade and FDI (with licensing and joint ventures as special cases) as mechanisms of technology transfer. Given that foreign firms opt to produce in a developing economy, FDI seems to be the preferred route and is therefore a prominent channel of technology transfer.

Third, the existence of several channels of international technology transfer raises two important quantitative questions that merit further research. First, is it possible to arrive at an aggregate measure of international technology transfer and its contribution to economic growth in developing economies? Second, can one isolate the role of each channel? For example, how much does FDI contribute to growth in addition to trade? The marginal contribution of FDI has important policy implications and can only be settled through careful empirical studies.

Fourth, a well-developed paradigm (called the OLI paradigm, for ownership, location, and internalization) seeks to explain the emergence of multinational firms, given the existence of viable alternatives such as exports, licensing, and joint ventures. The OLI framework is useful for explaining a one-time choice between alternative modes of serving foreign markets but is virtually silent regarding the dynamics of entry strategies. Future research needs to explicitly consider the dynamic decision problem facing multinational firms.

Fifth, local policy often causes foreign firms to opt for licensing or joint ventures over FDI. There is little or no empirical evidence to support the idea that licensing or
joint ventures are more likely than FDI to lead to increased learning. To be fair, few careful studies have attempted this difficult task. The jury may still be out on this issue.

Sixth, policies designed to lure in FDI have proliferated in recent years, but it is difficult to base the case in favor of these policies on the notion of positive spillovers from FDI to domestic firms. Several recent plant-level studies have failed to find positive spillovers from FDI to their direct competitors. However, these studies require careful interpretation because they treat FDI as exogenous. In addition, FDI spillovers may be vertical in nature rather than horizontal (as is assumed in such studies). Furthermore, all such studies find that the subsidiaries of multinationals are more productive than domestic firms. Thus, regardless of the evidence on the spillover issue, FDI does result in a more effective use of resources in host countries.

Seventh, several studies (both theoretical and empirical) indicate that absorptive capacity in the host country is crucial for obtaining significant benefits from FDI. Without adequate human capital or investments in R&D, spillovers from FDI may simply be infeasible. Thus, liberalization of trade and FDI policies needs to be complemented by appropriate policy measures with respect to education, R&D, and human capital accumulation if developing economies are to take full advantage of increased trade and FDI.

Eighth, empirical evidence supports the argument that IPRs are trade related and that asymmetric IPR protection across countries distorts the pattern of world trade. Furthermore, a country’s IPR policy may alter the composition of FDI at both the industry and firm levels. In industries in which IPRs are crucial, firms may refrain from FDI if IPR protection is weak in the host country, or they may not invest in manufacturing and R&D activities. Last, IPR policy may also lead foreign firms to choose FDI over other arm’s-length modes of technology transfer, such as licensing.

Note

Kamal Saggi is with the Department of Economics at Southern Methodist University. He can be reached via e-mail at ksaggi@mail.smu.edu. This article was written as a background paper for the World Bank’s Microfoundations of International Technology Diffusion research project. The author thanks the World Bank’s Development Research Group for financial support and Amy Glass, Bernard Hoekman, Aart Kraay, Aaditya Mattoo, Howard Pack, David Tarr, Jim Tybout, and three anonymous referees for helpful comments.

References


234 The World Bank Research Observer, vol. 17, no. 2 (Fall 2002)


