A

Foreign Direct Investment and Technological Spillovers in the Developing Nations

One of the themes of this book is that developing nations have been liberalizing their policies toward FDI over the last 10 years. A principal reason has been the promise of accompanying "technological spillovers." This appendix surveys what exactly these are and to what extent the evidence shows that FDI in developing nations does in fact create such spillovers.

A technological spillover is a special case of a positive externality, which is a benefit generated by an economic activity that is captured neither within the activity itself (i.e., owners or workers do not directly benefit in the form of higher profits or wages) nor directly by users of the activity's output—in other words, the benefit is *external* to the activity.[1] Thus, for example, the benefits to customers of lower prices for a product or service that a new technology makes possible are *not* externalities. If, by contrast, a firm creates a useful new technology that is subsequently copied or learned by competing firms, such that all products or services in the relevant market embody the technology, then the benefits to users of those products and services would be externalities.

1. Likewise, if an economic activity creates a cost that is neither born by the activity nor by users of the output, this would be a *negative* externality. An example of a negative externality is pollution, which generates a cost in the form of reduced health of those living near its source.

More generally, an externality is any benefit or cost received or borne by agents external to the activity that creates the benefit or cost. Direct users of the output of the activity are considered "internal" agents, as are all agents associated directly with the activity (including shareholders).

If, for example, the benefits take the form of lower prices, these lower prices would be the product of both the new technology and competition among firms employing the technology.[2] The relevant "economic activity" in this instance is the research and development (R&D) that produced the technology.

The foregoing is an example of a "horizontal" technological spillover because the firms that realize or generate the benefits compete with the firm that created the technology.[3] In practice, it will be difficult to divine what exactly is a gain from the horizontal spillover and what is simply a gain from internally generated improvements in the technological efficiency of a firm. For example, if firm X introduces a product embodying a new technology in response to a similar new technology previously introduced by firm Y, is this a spillover? In principle, firm X's technology is a spillover if the underlying knowledge was gleaned from firm Y; but in fact, firm X's technology is likely to be the outcome of some combination of its own R&D inputs plus incorporation of knowledge gleaned from firm Y. This ambiguity, among other reasons, bedevils the empirical measurement of spillover effects. Vertical as well as horizontal technological spillovers can occur. For example, if a manufacturing firm transfers technology to firms supplying inputs that enable these firms to deliver their inputs at a lower cost and these lower costs are then passed on to customers other than the original firm, then a spillover has occurred. There is no spillover, strictly speaking, from the supply of lower-cost inputs to the firm that originally transferred the technology, these being internal rather than external benefits. However, most developing countries consider the enhancement of locally owned suppliers' efficiency to be an unmitigated blessing, whether or not spillovers in the strict sense occur.

Indeed, increased efficiency of local firms through technology transfer is one of the main reasons developing countries seek FDI. To enhance this transfer, as well as simply to encourage the growth of indigenous suppliers, developing countries have often imposed "local-content" requirements on multinational firms. However, the World Trade Organization (WTO) now bans imposition of most new local-content requirements under the Agreement on Trade-Related Investment Measures (TRIMs;

2. Strictly speaking, those benefits that would have accrued to users of the product had the product been offered solely by the original innovator (who presumably would have charged monopoly prices) should be deducted from the total, as these benefits would be captured internally.

3. The externality must be distinguished from the benefits that accrue simply from additional competition. For example, suppose that a new firm enters a market and brings with it new technology. Prices subsequently fall. The fall in prices might be attributable to the new technology, including spillover of this technology to competing firms. But it might also be simply the result of more competition.

see chapter 5), and developing countries having such requirements must phase them out.

Vertical spillovers can also be achieved via technology transfer from a multinational firm to "downstream" operations servicing that firm (for example, distribution and retailing). Once again, in principle, a spillover occurs only if external agents realize at least some of the benefits from the technology. This would occur, for example, if the efficiency-enhancing technologies that distributors absorbed from a multinational firm could be applied to products and services sourced from firms other than the multinational itself. However, again, most developing countries seek such technology transfer as an end in itself, whether or not spillovers result.

As already suggested, technological spillovers are difficult to measure, as are all externalities. Externalities generally do not, as it were, leave a "paper trail" that is easily detected. Thus, most evidence that such spillovers from multinational firms' direct investment exist is circumstantial—if it can be established that technology transfer occurs, spillovers are assumed to accompany this transfer.

Among the earliest published evidence of spillovers was John H. Dunning's now-classic work (1958) on manufacturing direct investment by US firms in the United Kingdom (discussed extensively in the main text of chapter 3). While the United Kingdom is hardly a developing country, the efficiency of UK manufacturing firms over a wide range of industries was significantly lower than that of US firms (at least at the time of the study), and so arguably the results have some relevance to developing countries, where much the same thing can be said about relative efficiencies. Dunning showed that UK affiliates of US firms were more efficient in terms of labor productivity than domestically owned rivals, evidence that US firms were transferring to their affiliates technologies (including perhaps managerial prowess) that were not available to locally owned firms.[4] Moreover, Dunning demonstrated that local rivals were eventually able to catch up to the US firms in terms of productivity, thus suggesting that horizontal spillovers in fact did occur.

There was also evidence of vertical spillovers, achieved largely via transfer of technology or other information from the US-affiliated firms to local suppliers. The vertical spillovers varied from industry to industry and included reductions in unit cost (sometimes resulting from expansion and/or greater specialization in output and associated scale economies as well as from technology transfer), improvements in the product itself, and improvements in management, including in logistics.

While Dunning was able to quantify the productivity differences as

4. Dunning also showed that the productivity of the US affiliates in Britain was typically less than that of the operations of the same firms in the United States itself.

between US-affiliated firms and local horizontal rivals, much of the evidence on vertical spillovers was based on supplier firms' qualitative responses to questionnaires. These firms typically indicated whether or not their dealings with US-affiliated firms had resulted in tangible changes (e.g., unit-cost reductions or product improvements) without indicating the magnitude of the changes. Much the same can be said about later evidence regarding vertical spillovers: the evidence often is of a qualitative rather than a quantitative sort.

The Dunning study, although limited in focus and by now quite dated, nonetheless remains unique in terms of its thoroughness. It also established substantial direct and indirect evidence that FDI does generate technological spillovers even in relatively advanced economies. More recent studies compile largely indirect evidence that these spillovers occur in developing nations as well. We next turn to some of this evidence.

In a very recent study, Borenszstein, De Gregorio, and Lee (1994) present indirect but compelling evidence that FDI does generate technological spillovers. Their methodology is to use panel data from 69 developing countries to test the effects of FDI from the OECD nations upon growth of these nations. They first develop a model based on concepts of endogenous economic growth, wherein FDI is treated as a source of new varieties of capital embodying growth-enhancing technologies. They then test the following equation, derived from the endogenous growth model:

$$g = C_0 + C_1 FDI + C_2 FDI \times H + C_3 H + C_4 Y_0 + C_6 X$$

where g is growth in a particular developing country, *FDI* is inflow of FDI into that country from the OECD countries, H is a measure of human capital, Y_0 is income at the beginning of the period, and X is a vector of other variables that might act as determinants of growth. Because the data are pooled cross-sectional/time series ("panel data"), each variable implicitly bears a country and time subscript. Only FDI from the OECD nations is considered because these nations perform most of the world's R&D, much of which is in fact performed by multinational firms based in these nations. The implicit assumption is that if FDI from the OECD nations into the developing nations is positively associated with higher growth, the link between the growth and the FDI is the technology that foreign investors transfer into the local economy.[5]

The chief empirical findings are (1) that FDI contributes positively to economic growth and that the effect of FDI on growth is higher than

5. The endogenous growth model developed by the authors provides a theoretical framework under which this might happen, but the mechanics of this model are, in the view of this author, somewhat unrealistic. Fortunately, the tests they provide do not depend upon the mechanics of the endogenous growth model being exactly correct.

that of domestic investment, (2) that FDI interacts with human capital—in particular, a minimum threshold of human capital must be obtained before FDI makes a positive contribution to growth,[6] and (3) that FDI does not displace ("crowd out") domestic investment but rather seems to supplement it. Finding (1) is consistent with other recent studies indicating that FDI to developing countries is associated with technology transfer (e.g., Blomström, Lipsey, and Zejan 1992; De Gregorio 1992). Findings (2) and especially (3) provide indirect evidence of technological spillovers, for two reasons.

First, if spillovers do occur, one would expect that the country must have "absorptive capacity"—that is, trained people able to learn and apply the knowledge flowing out of the multinational firm and into external firms. Finding (2) is consistent with this expectation.

Second, if multinational firms and their customers internalized all of the benefits of technology transfer to developing countries, it might be expected that local affiliates of these firms would "crowd out" domestic investment because their greater efficiency would enable them to expand at the expense of less efficient rivals. While finding (3) does not necessarily negate the idea that local affiliates of multinational firms displace domestic rivals, the complementarity between FDI and local investment shows that, to the extent that such "horizontal" crowding out does happen, it is dominated by the capital investment activity of other local firms whose operations expand with those of the multinational enterprises. One can deduce that these other local firms are probably suppliers, distributors, and retailers whose operations are linked to those of the multinational firm and, indeed, that these local firms benefit from technology transferred by the firm.

At the end of the day, however, the results of Borenszstein et al. do not conclusively demonstrate that these links do exist. The best that can be said is that these results are consistent with such linkages and provide circumstantial evidence of spillovers.[7]

The Borenszstein et al. findings are consistent with those of another recent study by Coe, Helpman, and Hoffmaister 1994. Their basic finding is that productivity tends to be higher in developing countries with strong trade links to OECD countries than in developing countries without such links. The authors attribute this to the fact that high-productivity

---

6. In the regressions, using the seemingly unrelated technique, FDI appeared positive and significant when the interactive (*FDI* x *H*) variable was omitted. When *FDI* was omitted but the interactive term included, the interactive term is positive and significant. And when both terms are included, the *FDI* term is negative but insignificant, whereas the interactive term is positive and significant, indicating the interaction is critical.

7. Missing from their tests is whether domestic capital is more efficient in the presence of FDI than otherwise; such a test, if positive, would be even more suggestive of spillovers than the evidence presented.

developing countries import products embodying technology from the OECD countries. However, because FDI to developing countries from the OECD countries is highly correlated with trade flows, the higher productivities could be associated with technology transfer brought about by FDI rather than trade.[8] Thus, two of the authors, Coe and Hoffmaister, are attempting to separate trade and FDI effects, but their results had not been reported at the time of this writing.

Wei (1996) provides evidence that FDI in China has produced technological spillovers. Wei's data are specific to urban areas in China, covering 434 of them over 1988-90. He first shows that of three types of firms in China—notably, state owned enterprises (SOEs), township and village enterprises (TVEs), and foreign-invested firms—foreign-invested firms grew faster than other categories. The total share of these firms in industrial output is rather small, however, growing from less than 1 percent in 1988 to about 1.7 percent at the end of 1990. But, across the urban areas, this share ranges from as low as zero to as high as 62 percent. Thus, what Wei sought to do was to determine if the share of foreign-invested firms was a significant explanatory variable for growth differentials among the urban areas after accounting for other factors that might explain these differentials. These other factors included growths of population, the capital stock, and human capital. Wei used three variables to capture the effect of foreign-invested firms: unadjusted stock of FDI as a percentage of total capital stock, an adjusted stock of FDI as a percentage of total capital stock (where the adjustment accounted for differences in treatment of depreciation between the two aggregates), and share of foreign-invested firms' output in total output. The regression coefficients of the second two (but not the first) measures are statistically significant, indicating that FDI does seem to be an explanator of growth differentials among Chinese urban areas.

These results hold when Wei introduces dummy variables to account for certain urban areas being either "special economic zones" and "open coastal cities" or "comprehensive reform experimenting cities." In earlier work, Wei showed that although all three of these categories of cities have grown at above average rates in China, the differences in growth of output per capita between the first of these two subsets and other cities is well explained by differences in growth of capital inputs, while the second subset has not grown at faster rates after controlling for labor input.

These results are all consistent with the existence of technological spillovers, but, again, they hardly prove that such spillovers exist. However, Wei presents additional empirical results that provide stronger, albeit still circumstantial, evidence. In regressions with the growth rate of town-

8. Likewise, it is possible that the growth effects Borenszstein et al. attributed to FDI are in fact the consequence of importation of goods embodying new technologies.

ship and village enterprises' output by urban area as the dependent variable, he finds that two of his FDI measures—adjusted stock of FDI as a percentage of total capital stock and share of foreign-invested firms' output in total output—are statistically significant as explanators of differences in TVE growth. He notes that this positive association could be the result of additional competition created by the presence of foreign-controlled firms, but he believes that the more likely possibility is technological spillovers.

Wei's findings on growth rates are consistent with those in several earlier studies performed cross-sectionally within countries. For example, Blomström (1986) showed that labor productivity across industries in Mexico was correlated with the presence of foreign-owned firms after controlling for classical determinants of productivity such as capital intensity, quality of human capital, industrial structure, and existence of scale economies. More recently, similar results for Morocco have been reported by Hannad and Harrison (1993), who note that in sectors with a high presence of foreign-controlled firms there also tends to be lower variance of productivity, a fact suggesting the existence of horizontal spillovers.

Each of the studies examined above is essentially a cross-sectional investigation—that is, they investigate differences across countries (or, in the case of Wei, across urban areas). It would be helpful to the study of spillovers if it could be established in a time-series analysis whether FDI is associated with any of the characteristics of spillovers within a single nation. For example, has productivity growth in any given sector in a developing country accelerated following the entry of multinational firms? Alas, such quantitative analysis has not been performed in developing countries.

There have been qualitative case studies involving specific industries and a time dimension that bear on the issue of spillovers. One of the earliest of these was of the automotive industry in India, Peru, and Morocco by Lall (1980), who showed that the local content (the percentage of value added originating locally) of automobiles produced in India tended to rise over time, indicating that the multinational firms producing the autos increased the extent of their vertical integration with local suppliers. Lall attributed this rise to governmental policies that encouraged local content. However, his investigation did not cover whether vertical spillovers actually occurred as a result of the linkages.

Several subsequent studies have been performed—for example, of the automotive industry in Nigeria by Landi 1986 and of six manufacturing industries in several developing countries by Halbach (1988). These tend to confirm Lall's observation that global firms in these countries tend to extend their vertical linkages over time. Halbach in particular reports that the amount of local subcontracting (i.e., purchasing of inputs from local suppliers) depends upon the level of indigenous technology and

the skill of the local work force, findings roughly consistent with those of Borenszstein, De Gregorio, and Lee (1994). Halbach also reports that government policy had a significant impact on the extent of vertical linkages, consistent with the Lall finding. In two studies of the consumer electronics industry in Singapore and other southeast Asian nations performed five years apart, Lim and Pang (1977 and 1982) report that vertical linkages between multinational firms and local suppliers were not significant at the time of the earlier study but had grown substantially just five years later.

The Lim and Pang results are of special interest because they concern an industry that was very export-oriented (unlike the industries studied by Lall, Landi, and Halbach; multinational firms in these industries in fact were operating behind high protectionist walls and served local markets only). Presumably firms in an export-oriented industry will source components from the most efficient producers they can find or create. The fact that multinational firms operating in the electronics industry in Southeast Asia turned increasingly to local suppliers suggests that these suppliers were highly efficient. It is not unreasonable to surmise that part of the story behind the efficiency was technology transfer from the multinationals. Indeed, in subsequent years, the electronics industry in Southeast Asia has become very competitive internationally, and the structure of the industry is characterized by numerous linkages among large networks of firms, some foreign-controlled and others under local ownership. The linkages these networks create cross national boundaries, with Singapore largely serving as the hub. The development of the network of firms is highly suggestive that significant spillovers occurred as the result of linkages between the multinational firms and local suppliers.

What can be said, in conclusion, about technological spillovers resulting from FDI in the developing countries? It is certainly safe to say that these positive externalities are elusive: they are difficult to observe, let alone measure. But it is probably also safe to say that they do exist. While the evidence is hardly conclusive, it is all consistent with the idea that spillovers do occur and points toward likely yield positive, significant benefits to the countries in which they happen. Thus, developing nations that encourage inward FDI in the hopes of capturing spillovers almost surely are not chasing a will-o'-the-wisp.

B

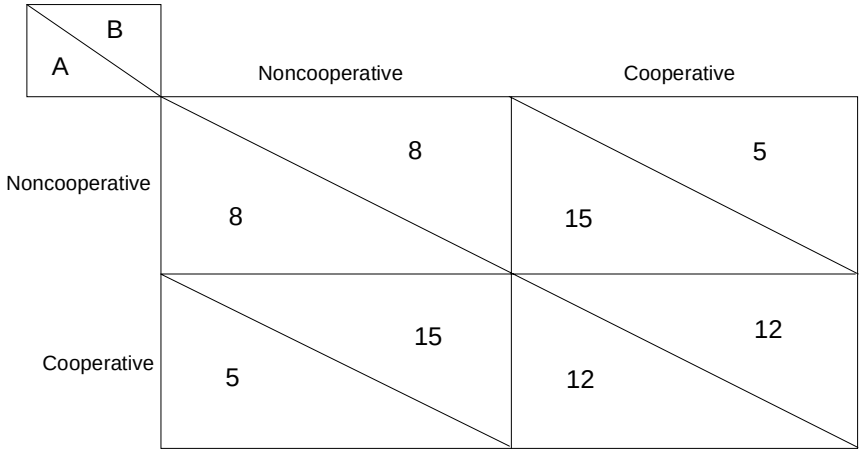
A Game Theoretic Approach to Sanctions

Game theory offers one explanation for the weakness of the old GATT procedures—and by extension, for a potential weakness of an international investment accord enforced by sanctions. Although termed a "cooperative arrangement," the WTO mechanism can be thought of in the jargon of game theory as an arrangement whereby "players" (the signatories) play a repeated noncooperative game. The game is "noncooperative" because there is no external agent with powers to enforce the rules. The institutional arrangement nonetheless has some potential to be "self-enforcing" because the game has something of the structure of the "prisoner's dilemma." In a prisoner's dilemma, if each player plays noncooperatively—that is, tries to maximize unilateral gains subject only to the observed strategies of other players—the outcome will be (Pareto) dominated by any number of cooperative solutions.

What does this mean? The essence is that if each player attempts to assess the likely strategy of other players and then to play a strategy that maximizes unilateral interests, the result will be that all players collectively fail to maximize the total payoff that they could otherwise achieve.

The simplest case is a game wherein each of two players can each play one of two moves. Figure B.1 indicates the payoff to each of two players (player "A" and player "B") of such a game for each pair of possible moves. The cells inside the matrix indicate the payoffs to A and B from cooperative and noncooperative moves, with the payoff to A on the left of each cell and the payoff to B on the right. Thus, if A seeks to maximize unilateral payoff, A will choose a noncooperative move irrespective of which move B makes. That is, if B also chooses the noncoop-

Figure B.1



erative move, A receives 8 units of reward whereas if A had chosen the cooperative move, A would receive only 5 units. But if B chooses the cooperative solution, then A receives 15 units of reward for choosing noncooperation but only 12 units for choosing cooperation. Because B faces an exactly similar payoff structure, B is driven to choose a noncooperative move, and thus each player would receive a payoff of 8.

The paradox is that if both players had chosen cooperative moves, they would each have received 12 units of reward, and hence both would have done better. The secret seems to be for them to devise an enforceable agreement in which each would choose the cooperative move *and* which would penalize either party if that party attempted to cheat on the other by selecting the noncooperative move on the assumption that the other stuck to the agreement (in which case, the cheating party would raise the payoff to 15, whereas the payoff to the honest party would be reduced to 5).

In terms of game theory, the objective of international rules on trade or investment is to force countries (the "players") to play a cooperative strategy whereby an optimal outcome can be achieved. For this argument to make sense, it must be established that international trade or investment has the character of a "prisoner's dilemma" game. Whether this is truly the case is open to some question, but what seems to count is that nations tend to think that it is. That is, nations the world over seem to think that they can benefit from some sort of unilateral implementation of policies that selectively close their markets to trade or investment or regulate these in ways contrary to the interests of other nations. Likewise, there is wide consensus that all nations can achieve a

"Pareto-dominant" outcome (the 12-12 outcome in the illustration above) if they can collectively agree never to carry out such policies.

The trick is enforcement of such an agreement when each nation believes it could do a little better by welshing. One means of enforcement would be to hire an outside agent (some sort of "policeman") to compel each player to live by the rules. But, in the world today, no such policeman is available, at least not for the enforcement of commercial law. So, the alternative is some sort of self-enforcing agreement.

The literature on game theory has firmly established that a self-enforcing agreement is achievable if the game has no known termination and a punishment scheme can be constructed in which it would always be in the interests of all nondeviant parties to cooperatively punish a deviant player *and* the punishment is sufficiently harsh that no player would choose to deviate given the certainty of punishment (Fudenberg and Maskin 1986). Such a punishment scheme is termed "subgame perfect."

Even under reformed WTO dispute settlement procedures, sanctions would rarely if ever qualify as subgame perfect. When a country that takes a dispute to the Dispute Settlement Board, after failure to resolve it via consultation, there is some chance the board will rule that the "defendant" country has violated an obligation. Then the country that brought the dispute may apply sanctions. But no other countries are obliged to apply them, and, for the sanctions to have "bite," they must be initiated by a large country. Hence, the "deviant player" cannot be sure of certain and stiff punishment.

Other international agreements also fail to produce subgame-perfect punishment strategies. For example, NAFTA provides for no sanctions (or any other form of punishment) in the event that a member nation fails to uphold or enforce awards a tribunal makes against it under the dispute settlement procedure of chapter 11 on investment. The NAFTA commission can advise a member that it has violated its obligations but cannot penalize that member for it. In this regard, NAFTA is a purely voluntary agreement.

Can a subgame-perfect strategy, or something reasonably close to it, be devised and implemented in an international investment accord? One possibility game theory suggests would be for parties to such an agreement to obligate themselves to punish one of their number that violated an obligation, but it seems unlikely that nations would ever bind themselves in this way.

Another possibility might be to empower the dispute resolution body to fine a party in violation of a basic obligation, the value of which should be equal to or greater than the value to the offending party of continuing the violation. Because these parties are sovereign states, there would be no means to compel payment. Furthermore, it would be tricky at best to calculate what was the value of continuing misbehavior. None-

theless, imposing a fine would doubtless increase the opprobrium associated with violations and thus might enhance the moral suasion of a dispute settlement board or panel. Furthermore, the amount of the fine would signal the world what the board believed to be the cost of the violation in monetary terms, allowing the general public (including that of the violating country) to judge the severity of the violation.

In the end, it remains difficult to devise a sanctioning procedure that is subgame perfect. And thus international agreements, including the proposed accord on investment, will continue to be difficult to enforce. The prime value of such agreements rests, and will continue to rest, in setting standards for proper international commercial behavior and not in the creation of tightly enforceable legal instruments.