

Chapter 2. *Economic Rationale of Support for Commercial Innovation*

Ever since the path-breaking research of Robert Solow (1956), economists have known that secular growth is due mostly to technological change, rather than to factor accumulation. Indeed, a vast array of subsequent empirical research during a half century has shown conclusively that at least half of the growth in per capita income, in virtually every country studied, is associated with the growth of total factor productivity (TFP) rather than other, more traditional, factors. However, attaching the label of technological change to the famous “residual” (i.e., TFP growth) begs the question of what it contains exactly, and perhaps, more important, how it evolves over time, and the nature of the economic forces that determine its course and pace.

Indeed, one of the frustrating aspects of the early phase of economic thinking about these matters was that the growth of TFP (arguably the single most important economic phenomenon) appeared to economists as an impenetrable “black box” and seemed to occur outside the realm of economic forces. A long and very fruitful research agenda pioneered by Griliches, Jorgenson, Denison, Rosenberg, and their associates sought to open this ‘black box’ to understand its contents. However, it was only with the extensive development of endogenous growth theory in the late 1980s (Romer 1986, 1990; Lucas 1988; Grossman and Helpman 1991, etc.) that the economic profession came to accept the view that innovation, spillovers, and R&D were indeed the key factors driving self-sustained, long-term economic growth and, moreover, that these factors were generated from within the economic system, responding to economic incentives.⁴

Recent literature examines the interplay between competition and innovation and their impact on growth (Aghion et al. 2002). Although Schumpeterian growth models predict that only firms with market power would have the resources and incentives to innovate, these empirical findings suggest that, contrary to the Schumpeterian assumption, in the more developed economies, among the incumbent firms closer to the “technology frontier,” competition does encourage innovation. In a further article Aghion, Carlin, and Schaffer (2002) look particularly at the relationship between competition, innovation, and growth in transition economies: by using different methodologies and the BEEPS survey, they show that competitive pressures raise innovation in both new and incumbent firms, subject to hard-budget constraints for incumbent firms and availability of financing for new firms. In a more recent article, Aghion and Howitt (2005) discuss Europe in comparison with the United States and argue that since Europe has recently come closer to the world technology frontier, it would benefit from a competition and labor market policy that not only emphasizes competition among incumbent firms, but also stresses the importance of entry, exit, and mobility. For this report, it is important to note that the closer firms in ECA countries move to the technology frontier, the more

⁴ Work on education and technological change by Nelson and Phelps (1966) mentioned that technological progress was key to growth and highlighted the difference (for growth) between human capital stock and accumulation. However, it was only in the late 1980s that those views were widely shared.

competitive pressures and market structures will play a role in the innovation capability of a country. The companion report (ECAKE II) will elaborate on the impact of competition and innovation, particularly innovation financing.

This conceptual framework molds our analysis: on the one hand, the view of the centrality of innovation and knowledge creation in the growth process and, on the other hand, the understanding that these are economic factors that may be shaped and influenced by properly designed economic policies. Building on that view, a recent paper by the World Bank (Chen and Dahlman 2004) seeks to decompose “knowledge” into a wide array of indicators and assess their contribution to growth. By using an array of indicators, each of which represents an aspect of knowledge, as independent variables in cross-section regressions that span 92 countries for the period 1960 to 2000, the paper shows that knowledge is a significant determinant of long-term economic growth. It finds that an increase of 20 percent in the average years of a population’s schooling tends to increase the average annual economic growth by 0.15 percentage point. In regard to innovation, it finds that a 1 percent increase in the annual number of U.S. Patent and Trademark Office (USPTO) patents granted is associated with an increase of 0.9 percentage points in annual economic growth. Last, Chen and Dahlman find that when the ICT infrastructure, as measured by the number of phones per 1,000 persons, is increased by 20 percent, annual economic growth tends to increase by 0.11 percentage points.

One corollary of the developments just sketched was the emergence of a soundly based and carefully articulated economic rationale for public support of R&D and innovation, which is by now widely accepted among academic economists and practitioners. The basic argument for public support of R&D is that innovation is a critical factor for growth (and hence *inter alia* for poverty alleviation), but a well-functioning market economy cannot generate by

itself the optimal levels of R&D. There are two main sources of market failure with respect to R&D⁵: (1) partial appropriability (due to spillovers), which does not allow inventors to capture all the benefits of their invention, and (2) information asymmetries, for example, the difference between the information that an inventor looking for financing has about an invention and the information that the potential financier has, which leads to a “funding gap.” These failures inhibit private firms from investing enough in innovation and R&D, thus depriving the economy of one of the key levers for sustained growth.⁶

2.1 Partial Appropriability (spillovers)

A basic feature of knowledge creation is that the returns from investments in it are not fully *appropriable* by the original investor. Knowledge has significant public good attributes, that is, once created it can be used repeatedly by multiple actors at no or very low extra costs. Firms

5 For a full list of rationales for state interventions in fostering knowledge creation see the flagship study of the World Bank’s Latin American and Caribbean Studies, De Ferranti et al. (2003). They list other important aspects of knowledge creation that prevent markets from generating the optimal level of knowledge: (1) the long-term and risky nature of R&D investments, (2) lumpiness of innovation, and (3) coordination failures. See Baumol, 2002, for a description of the features of the free market economy (market structure, productive entrepreneurship and rule of law, markets for technology trading and reasons why R&D expenditure might be efficient despite substantial externalities on innovation) that explain its effectiveness in promoting innovation and growth.

6 Clearly though, it is not enough to spell out such an economic rationale: for it to lead to policy, it must be weighed against the *costs* of government intervention, namely, the well-known problems of “industrial policy,” capture and corruption, which constitute the so-called government failures. We shall discuss those later on.

making investments in knowledge creation capture only a portion of the benefits created. They do not receive compensation for the “spillovers” that their innovative efforts generate, that is, for the positive externalities of their actions on other firms and agents. Further, new technologies confer benefits to the purchasers of new products (consumers and producers alike) that often exceed any increase in the selling price that can be sustained. These nonappropriable benefits are also referred to as *spillovers* to consumers and are of particular importance in the context of “general purpose technologies” (as described further in section 2.3). Both types of spillovers, namely, the purely technological externalities and the excess benefits to buyers, imply that the social returns from innovations may be far larger than the private returns (Jaffe 1998).

As a result of this gap, innovators operating in a market economy will invest in R&D less than the socially optimal amount; the extent of underinvestment depends of course on the extent to which social returns exceed private returns, and that may vary widely across fields, technologies, stages along the innovation cycle, and so on. Empirical studies have shown that typically the social rates of return of R&D expenditures are very large, often several times larger than private ones (Klette, Moen, and Griliches 1999). Moreover, these studies show that the returns from R&D exceed by a wide margin the returns from other types of investments, in particular, from investment in physical capital. This implies that a government role in increasing the amount of resources devoted to R&D at the economywide level can have significant social benefits.

Spillovers may occur in many different ways, one of them being the mobility of R&D personnel. The process of innovation and its commercialization in an enterprise builds the human capital of its employees. Employees acquire R&D skills and an understanding of technologies and markets that are partly general, that is, they go beyond the specific knowledge

embodied in the innovation and protected by intellectual property rights (IPR). Employees that move from one firm to another carry with them this human (or innovation) capital, which may benefit their new employers beyond the increment in wages that the mobile employees may receive. If mobility takes the form of migration, then the origin countries may be unwittingly “subsidizing” the destination countries through these spillovers; thus the mobility of inventors is an important transmission mechanism for spillovers, and hence a channel that should be closely monitored because it may have both positive and negative effects on any given country.⁷

Openness to trade and FDI increase the probability of receiving spillovers that originate elsewhere. As Coe and Helpman (1995) have shown, large economies tend to benefit the most from international spillover flows mediated by trade. Countries can increase their productivity by importing goods (especially capital equipment) from foreign, more advanced technologies (Coe, Helpman, and Hoffmaister 1997). Another potential source of technology spillover is FDI, although investors frequently “keep their knowledge at home” (Blomstrom and Kokko 1999). That is beginning to change (e.g., R&D is moving to India), although the international principals still maintain control of the innovations via patents registered abroad. The impact

⁷ The spillovers-based argument clearly holds for large, mostly closed economies: being closed there is no risk of spillovers slipping out, and being large there is a high probability that at least some other local economic agents will benefit. In small open economies, spillovers may spill *out* of the country and benefit external firms and consumers rather than the local economy. Any policy designed to promote R&D should aim not only at increasing total R&D, but also at increasing total R&D in a way that incentivizes *local* spillovers rather than external leakages, develops absorptive capacity, and ultimately affects the productivity of a wide range of sectors in the local economy.

of FDI is indirect, via “spillover effects,”⁸ owing to the presence of multinationals—first, because they create linkages with domestic firms and, second, because their presence spurs domestic producers to invest in new technology to compete with the foreign-owned firms. For example, in the Czech manufacturing sector during the 1995–98 period, the indirect effect of R&D via the development of the absorptive capacity was found to be far more important than the direct effect of innovative R&D in increasing productivity growth of the firm; it was also found that R&D and intraindustry spillovers from FDI go hand in hand. (Kinoshita 2000). In Poland, so far, spillover effects leading to technology improvements in firms are observed only in a few industries, such as the auto industry, in which foreign R&D is high.⁹ To be able to capture these international spillovers, the country needs to develop “absorptive capacity” (see Cohen and Levinthal 1989), which entails, inter alia, investing in local indigenous education and innovation, which will be the subject of Part II of this study.

Another result of partial appropriability is “coordination failure.” Often, technical advances in a given field require complementary advances

⁸ Spillover effects (from neighboring countries or industries) arise when production affects the economic activity of other local firms or their employees. Positive spillover effects occur through the supply of new information, new technologies, managerial practices, and so on. Thus the “social” gain is larger than the profit or productivity gain made by the “source” company.

⁹ This is consistent with Kinoshita’s (2000) finding concerning Czech enterprises’ data—in oligopolistic sectors, such as electrical machinery and radio and television, there is a significant rate of spillover as a result of having a large foreign presence. Also, R&D investment has a higher rate of return in these sectors. However, less oligopolistic sectors, such as food and nonmetallic minerals, show no evidence of spillovers despite the large presence of foreign investors in these sectors.

by numerous distinct parties. Any one party may find that it is not worthwhile to develop one component of the system unless it can be sure that others will develop complementary components. If these parties do not have a mechanism to coordinate their investments, it is possible that no investment will occur. Government support may tip the balance such that multiple actors will invest in R&D independently. Innovation instruments can also be designed specifically to remedy coordination failures in innovation by encouraging “consortia” of universities/research institutes and firms or by promoting technology “clusters.” We will discuss encouraging consortia in chapter 3.

2.2 Information Asymmetries and the “Funding Gap”

A second source of market failure related to the creation of knowledge has to do with asymmetric information between inventors and external agents (e.g., investors such as banks). Innovative activities entail by necessity a fundamental information asymmetry, certainly *ex ante*, that is, at the stage at which the inventor formulates the idea and seeks funds to develop it. It can be assumed that inventors have sufficient knowledge of the technology and of the details of the planned innovation, of their true abilities to carry it out, and of the efforts they are willing to put into developing the innovation. However, there will always be a significant gap between what the inventor knows and what an external agent can gauge, even if the information on those crucial matters is well documented.

In particular, there will be significant information asymmetries in that respect between the inventor and mainstream financial intermediaries, such as banks and institutional investors, who lack the capacity to verify the specific technical information and claims of the entrepreneur. Potential investors will therefore be skeptical of the likely returns on investments in developing new technologies. Entrepreneurs

who could offer attractive returns may have no credible way of conveying such potential to risk-averse investors.

The information asymmetry makes it very hard for a creditor or equity investor to predict the returns from a potential investment in innovative ventures, which implies that such funding is not likely to be forthcoming. In the absence of demonstrated cash flows or other collateral, a typical start-up company or individual innovative entrepreneur will not have access to traditional sources of finance—this is the so-called “funding gap”. At the most basic level, the “funding gap” implies that entrepreneurs face stiff constraints in the funding of innovations and therefore will not invest (or will invest too little) in innovative projects that may have high social returns.

This “funding gap” has been studied in most detail in the United States, but the findings have direct implications for the ECA region as well. Figure 4 shows an estimated breakdown of actual funding sources for early stage technological development (ESTD) in the United States (see Auerswald and Branscomb 2003). ESTD is the most problematic phase in the innovation process and is defined as the link between invention and innovation, when a new product and market are identified. In this stage, product specifications appropriate to the identified market are demonstrated. Production processes begin to be developed, allowing estimates of production costs. At the end of this stage the entrepreneurial venture has articulated a business case.”

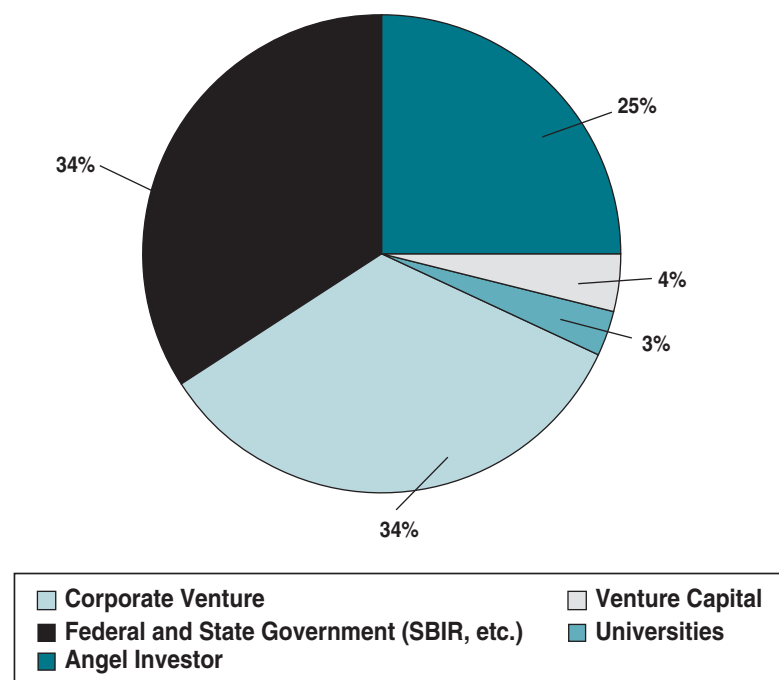
Figure 4 emphasizes the importance of internal financing by enterprises, government funding, and “angel investors” in the ESTD stage. But most important, it emphasizes the virtual absence of more mainstream intermediaries such as banks, private equity, and other institutional investors. Although the percentages are for the United States

only, the figure illustrates that, typically, even in one of the most advanced and innovative economies, early-stage finance of innovative projects is undertaken directly by firms, if they have the resources, or by very specialized institutions, with a significant role played by the government.

Not surprisingly, internal funds account for the biggest share of ESTD financing in the United States because that is the most straightforward way of overcoming information asymmetries. Established enterprises know the track record of their own inventors/employees and, typically, have a better understanding of the market and the commercial potential of internally proposed innovations than do outside agents. Enterprises use the cash flows generated by established operations to finance innovation or source external funds on the basis of their balance sheet strength.

“Angel investors” are another important source of ESTD funding in the United States and to some extent in Europe. The term “angel investor” refers to successful entrepreneurs that look for new opportunities to invest private

Figure 4. ESTD Funding in the United States



funds (earned from their own previous innovations) and are willing to invest in ESTD projects in technological fields that they understand well (having “been there and done that”). Studies of the behavior of “angel investors” frequently find that they are often heavily involved in the commercial decision making and that this “business support” function can be as important as the financing. Managerial advice and commercial control over the ESTD entrepreneur are typical characteristics of the “angel investor” and venture capital funding models, as well as, of course, in internal funding models.

Given the short history of capitalist accumulation and profit-generating enterprises in ECA, internal financing by enterprises and “angel investors” is rare in the region and does not provide a viable basis for promoting innovation. The absence of “angel investors” is problematic not only from a funding perspective, but also given their role as sources of managerial expertise, as information brokers, and as access points to formal and informal networks of entrepreneurs and innovators. The role of government is therefore different in ECA countries than in OECD countries. The lack of “angels” and internal financing is acute, and the capacity of government agencies to fill their place is extremely limited. The Finnish case study in box 2 and the discussion in chapter 3 provide possible options for interventions that compensate for the absence of local “angel investors” by promoting international “angel investor” networks and building public information marketplaces and networks.

ESTD requires patient and high-risk tolerant investment capital to fund early, prerevenue stages of research, development, and commercialization. Yet filling the “funding gap” requires specialized investors with the skills to evaluate and directly manage the risks of ESTD (“angel investors” or innovative managers in firms that are willing to invest retained earnings accumulated in other activities in the highly risky innovative projects) or governments with a broader public policy objective of capturing some of the

spillovers associated with ESTD. In the absence of positive internal cash flows and “angel investors”, even if appropriability is adequate to yield a reasonable profit expectation, it may be impossible to secure the capital necessary to develop a new technology. Typically, in developing countries, the information asymmetry and “funding gap” problem is much more acute than in developed economies.

2.3 The Impact of General Purpose Technologies (GPTs)

Technological change contributes to growth wherever it happens, but there are certain technological advances that have played a critical role over the long term in fostering growth in the economy as a whole. Indeed, in any era there are a handful of (or even a single) “general purpose technologies” (GPTs) that drive growth by spreading over the different sectors of the economy and prompting them to innovate further (see Bresnahan and Trajtenberg 1995 and Helpman and Trajtenberg 1996). Progress in the adopting sectors feeds back, in turn, into the GPT sector, providing incentives for further advances in the GPT itself, thus setting up a positive, self-sustained virtuous cycle.

During the past two decades innovation has commonly been associated with the tremendous technological advances that have taken place in what is generally referred to as “high tech” and in particular in Information and Communication Technologies (ICT). Indeed, the advent of the personal computer and of the Internet, cell phones, and the digitization of words, voice, and image in a wide array of existing and newly created media and, above all, the inexorable march of “Moore’s Law” have revolutionized the way we produce, communicate and consume virtually everything. The preeminent GPT of our era is undoubtedly ICT, and as such it is enabling and fostering economic growth in developed countries as well as in many transition and developing countries.

As a result of the wide-reaching impact of ICT, many developing countries focus their innovation support specifically on developing an ICT sector. This approach misinterprets, however, the role and impact of GPTs on innovation. The way a GPT fosters economywide growth is not simply and not mainly by innovation taking place only in the GPT itself; rather, economywide growth occurs when a wide and ever-expanding range of *other* sectors adopt the advancing GPT and, as a consequence, improve their own technology. The best example is the revolution in retailing brought about by WalMart via the massive adoption of ICT-based methods, which made a sizable contribution to the productivity growth of the U.S. service sector and the economy during the second half of the 1990s. Indeed, the GPT sector itself is bound to be small relative to the economy as a whole (e.g., think of the steam engine-producing sector or the electricity sector), and however fast it innovates and grows in itself, it can never pull the whole economy on its own. In that sense, the often-used analogy of the GPT as a “locomotive” is wrong and misleading. Moreover, if the rest of the economy fails to widely adopt the GPT or fails to make complementary innovations and investments in the adopting sectors, economywide growth will just not materialize. That touches on the possibility of the emergence of a “dual economy,” an economy in which one or several sectors are highly innovative and dynamic but are mostly disconnected from the rest of the economy. The result is that only those parts of the economy connected to the innovative sector or sectors (“enclaves”) benefit and that spillovers to the rest of the economy are lacking, which could lead to increasing socioeconomic inequality.

The dual economy picture could be problematic not only from an equity viewpoint; it may also affect the future growth potential of the economy by restricting the pool of skilled labor and otherwise creating frictions and tensions that are detrimental to growth. In India

and China the impact of “enclave” innovation on the rest of the country remains to be seen: some argue that the benefits of such innovation, mostly exporting to foreign markets, are being reaped mostly by a small share of the population and regions engaged in these industries as well as foreign customers. Yet, the increase in exports surely improves the current account and the fiscal situation, thus allowing expansion of services for those who are poor: building schools and hospitals, fighting diseases, and so forth. Moreover, the skills acquired by the entrepreneurs or employees of the innovative companies increase the human capital of the country of origin even if the new product is exported and the inventing start-up is acquired by a multinational corporate investor.

A key issue then in countries that are not at the frontier of the GPT (“secondary countries”) is how to allocate R&D and other innovative inputs so as to leverage the growth potential of the prevalent GPT. Just trying to jump onto the bandwagon of ICT innovation per se is far from enough and may not necessarily be the most effective strategy. A more effective strategy will be to encourage ever-expanding segments of the economy to adopt ICT in ways that increase their own productivity. These types of complementary actions (i.e., adoption of ICT, local innovations in traditional sectors, etc.) may appear to be less overtly “innovative” (and therefore may not be deemed as worthy of public support) but will eventually constitute the key to economywide growth. Intensification of ICT adoption, in particular, in services and traditional manufacturing sectors, requires a conducive business climate that combines open borders to trade with liberalized product and labor markets, robust financial systems, developed regulatory framework, and sufficient quality of human capital.

Nevertheless, it can be argued that in some more advanced middle-income countries, developing the legal and institutional framework (e.g., protection of IPR) so as to enhance the

development of a local ICT industry, joining forces with ICT multinationals, and otherwise encouraging the ICT-producing sectors may play an important role in the process of economic development and economywide innovation for two reasons. First, an indigenous ICT sector allows for the concomitant development of local technological skills, managerial expertise, and world-class standards in ICT. Second, local ICT strategies require the opening up of the economy, which itself brings in inflows of capital, expanding trade, and so on. The spillovers related to a thriving local ICT sector may play a crucial role in prompting the rest of the economy to innovate and open up.

Thus, growth-oriented innovation policies have to proceed from a far wider perspective than just promoting the ICT sector per se. Indeed, the guiding conceptual framework should be that of GPT, which emphasizes the spread of (in our era) ICTs throughout the economy, and the “innovational complementarities” that ought to materialize for economywide growth to take place. The development of the ICT sector may be in some cases an effective stepping stone, but by no means the final destination. In fact, the recalcitrant problem may lie in eliciting adoption and innovation, not in ICT producers but in those that could benefit from its use (Jorgenson and Vu 2005; Piatkowski and van Ark 2005).

2.4 Exports-versus Local Markets-Oriented Innovation

The issues of “high tech” versus the rest of the economy and of exports- versus local market-oriented innovations are connected, as touched on in the previous discussion. For developing countries, especially, the role of exports- versus local market-oriented innovations deserves further scrutiny in evaluating a rationale for innovation support. In the era of globalization, there are widely held perceptions that there is no such thing as “local needs” or “local markets,” particularly not in innovative technologies. This

view holds that virtually all relevant markets are global, and hence local innovators should aim at serving global demand rather than local niches. It cannot be denied that the ICT sector is pre-eminently global, both in inputs and outputs. Furthermore, the extent of global specialization and cost arbitrage is increasing over time, leading to further productivity gains and faster innovation. For many countries, linking with this vast, enormously complex and extremely dynamic ICT machine can be a worthy policy goal. That approach, however, does not preclude supporting locally oriented innovation, which can be desirable and even critical for growth.

Globalization does not imply homogeneous demands to be served by uniform products and services. There is increased recognition of the inherent heterogeneity of preferences (and of “needs,” even if this notion is ill defined in textbook economics) in specific markets and of the vast opportunities to increase both consumer surplus and profits by catering to this heterogeneity. In fact, advances in ICT and in the Internet, in particular, are often heralded as providing the means for such mass customization, that is, for tailoring products and services to the specific preferences of individuals, without sacrificing scale economies.

If this heterogeneity trend holds true for markets in (advanced) countries, it surely holds across markets, across countries, and across the development divide. The needs to be served in developing countries differ from those of developed countries in a wide array of markets, and in some areas may be radically different. Applying a single strategy for R&D and innovation that focuses on plugging into the global network of ICT to supply the demand emanating mostly from developed countries neglects a huge market opportunity. There are vast areas of economic activity in which innovation is needed to serve local needs and local demand, in which “local” may mean a large fraction of the developing world population. A few examples can illustrate that point:

In the area of health care, the incidence of diseases in less developed countries differs significantly from the Western world, with the prime example being the prevalence of tropical diseases (e.g., malaria, parasites, yellow fever). Given the dearth of access to medical care and even to elementary medicines, the largest market opportunities in developing countries are innovative ways of delivering simple, cheap, easily administered preventive medicine. Innovation in sophisticated technologies (e.g., MRIs, stents, “orphan” drugs for rare diseases) are virtually irrelevant for those countries, and in some cases may have unintended consequences (such as the widespread use of ultrasound in India to select male newborns).

In the ICT sector and, in particular, in software, developing country markets rarely need what is typically viewed in advanced economies as innovation, such as adding more features to already highly complex and cluttered software packages. Instead, innovations are needed to simplify operations and ensure “sturdiness” and backward compatibility, so that barely literate workers could use software, computers, and computer-based tools in a reliable fashion, and use older versions as well. Likewise, local innovations could improve and reduce the costs of satellite-based broadband to deliver Internet services to farmers in isolated villages or develop search engines tailored to their prime needs, for example, having real-time information on prices of crops and on agricultural inputs.

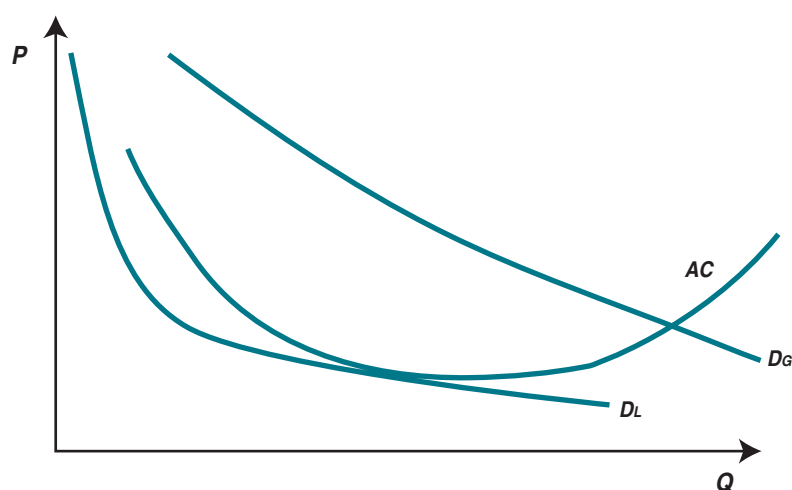
It could be argued that if it were profitable to engage in innovations oriented toward local needs, then market forces would lead to it, and therefore there is no need to intervene in that respect. Figure 5 illustrates why that may not be the case:

DG denotes the demand from high-income countries (the “global” demand); DL is the local demand; and

AC , whose shape is driven by a fixed cost of innovating (e.g., R&D), is the average cost curve facing local entrepreneurs. Absent intervention, the local entrepreneur will surely develop an innovation to serve the global demand, because doing so would result in positive profits, whereas as it stands serving the local market would not cover the fixed cost.

Is it optimal then to leave it at that? Not necessarily. A small R&D subsidy may tip the balance and make it profitable to innovate for the local market, and the local surplus generated may be significantly larger than the subsidy. Recall that the “global” consumer surplus (under the DG demand curve) is irrelevant from the standpoint of the local economy, only the profits count, whereas if serving the local demand *both consumer and producer surplus* should count equally. In particular, the social gains of serving the local market in regard to consumer surplus may be very large, as is likely to be the case in the area of medical care (e.g., developing a malaria vaccine). Moreover, local spillovers may be in some cases more significant and more widespread if innovating for the local market, if only because of demonstration effects, but that of course remains to be established empirically.

Figure 5. Exports-versus Local Markets-Oriented Innovations



Source: Trajtenberg 2005