ECONOMIC RESEARCH REPORTS

PARETO OPTIMAL SIZES OF INNOVATION SPILLOVERS

by
William J. Baumol

RR# 97-42 **December 1997**

C.V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, NY 10003-6687

Abstract

The paper offers reasons indicating that the spillovers from innovation activity are surprisingly large; and argues that individuals who have invested directly or indirectly in the economy's innovation processes can be estimated, conservatively, to obtain less than ten percent of the total economic benefits contributed by new technology and new products. In contrast to standard analysis, the paper asserts that there is a tradeoff between increased flow of invention and the distribution of benefits because of which zero spillovers cannot be expected to optimal; and that there is a range of values of what will be called the spillover ratio, that is, the share of the benefits that goes to persons other than the investors, such that all values of the ratio within this range are Pareto-optimal. It can be suspected that the high spillover ratio found in reality falls within the range of Pareto optimality. The reason is that, in contrast with most of the literature, here spillovers are considered to be capable of offering social benefits, and do not always simply impede or prevent the attainment of optimality. The results of the paper may offer a step toward reconciliation of the poor innovation performance of our economy that standard theory leads us to expect, and the historically unprecedented actual performance in terms of income growth and innovation of the free market economies.

JEL Classification: O30
Key Words: Spillovers, real wages, optimality in innovation.

File: Inovspi3.art 12-12-97

PARETO OPTIMAL SIZES OF INNOVATION SPILLOVERS

William J. Baumol*

*Director, C.V. Starr Center for Applied Economics at New York University, and senior research economist and professor emeritus at Princeton University. I am very grateful to Ralph Gomory for graphic constructs that he and I originally designed to analyze another subject (Gomory and Baumol, 1997). As usual, I am also deeply indebted to Sue Anne Blackman and Edward Wolff for very helpful comments and suggestions. I also want to thank the Alfred P. Sloan Foundation, the Russell Sage Foundation, and the C.V. Starr Center for their very generous support of this work. There is, of course, a very illuminating literature on the appropriability of the contributions of innovation, going back at least to Scherer (1965), from which I have learned a great deal on the subject. I am also extremely grateful to Professor Jean Gadrey of the University of Lille and several of his colleagues for very valuable comments on this paper.

This paper provides several central results. First, it offers reasons indicating that the spillovers from innovation activity are surprisingly large. It argues that individuals who have invested directly or indirectly in the economy's innovation processes can be estimated, conservatively, to obtain less than ten percent of the total economic benefits contributed by new technology and new products. According to the standard analysis, in a market economy only zero spillovers are compatible with optimality in innovative activity, and spillovers as great as claimed here can be expected to lead to investment in innovation far below the optimal level. However, the remaining basic contentions of this paper assert that this is not so. There is a tradeoff between increased flow of invention and the distribution of benefits because of which zero spillovers cannot be expected to be optimal. Moreover, there is no one level of expenditure that is unambiguously optimal. Instead, there is a range of values of what will be called the <u>spillover</u>

ratio, that is, the share of the benefits of innovation that goes to persons other than the investors, such that all values of the ratio within this range are Pareto-optimal. Consequently, there is no way in which economic analysis alone can choose among them. Rather, value judgments must be employed in making that selection. Indeed, it can be suspected, for reasons that will emerge during the discussion, that the high spillover ratio found in reality falls within the range of Pareto optimality.[1] The reason for these results is that, in contrast with most of the literature, here spillovers are considered to be capable of offering social benefits, and do not always simply impede or prevent the attainment of optimality.

The results of this paper may, perhaps, be a step toward reconciliation of the poor innovation performance of our economy that the standard theory leads us to expect, and the historically unprecedented performance in terms of income growth and innovation of the free market economies that is clearly documented by the economic historians, and that is the feature of our type of economy that has surely (and deservedly) been most attractive to nonspecialists.

I. The Size of the Spillover Ratio and the Welfare of the General Population

My analysis takes off from a pathbreaking paper by Paul Romer (1994) that provides a profounder view of the spillovers generated in the innovation process, indicating that general wage gains resulting from innovation must be spillovers, since they are benefits that go to others than the innovator. I can add the observation that the bulk of the unprecedented rise in the

---

[1]My own value judgment is summed up in the G.B.S. dictum that there is no crime greater than poverty. Consequently, I am inclined to prefer a fairly high spillover ratio, perhaps not far from its current value.

developed world's living standards since the Industrial Revolution can be attributed to innovation, directly or indirectly, and conclude that a very substantial share of the benefits of innovation must have gone to persons other than the innovators. Indeed, since most members of the labor force are not innovators, and since the share of wages in GDP has remained on the order of 75 percent of the total as GDP has risen perhaps twenty-fold, it seems probable that less than 25 percent of the benefits of innovation have been obtained by inventors or those who brought their inventions into use.[2]

Romer studies the role of these spillovers as an impediment to innovative activity, discussing the difficulty innovators encounter in covering their sunk costs and the resulting reduction in innovative activity and output. That is part of the story that I will retell in this paper. However, I will also emphasize another side of the matter, to which Romer only alludes in passing: the inevitable tradeoff between the number of innovations actually produced and the standard of living of the majority of the population. In this scenario, as overall GDP is raised, any increase in workers' standards of living constitutes a rise in the spillovers from innovation that depresses the flow of further innovation. Thus, the more the general public benefits from such growth in GDP, the slower that growth must be.

This is more than just an embellishment of the old story of the tradeoff between output

[2]Some writers on innovation use the term "spillovers" in a more restricted sense, to refer, e.g., to direct gains of knowledge by customers of the industry that supplies the R&D in question (see, e.g., Grilliches, 1979). Such use of the term is, of course, entirely legitimate. However, here, taking the term to represent all the benefits of innovation that do not accrue to the inventor or to those who invested in or otherwise contributed to it is clearly required by the issue under discussion -- the difference between the social and private reward for innovation activity.

and distributive equality. The mechanism under discussion here is very different, and does not involve the disincentive to work that results from a reduction of the marginal return to worker effort. Rather, we are concerned here with the heart of the capitalist growth process: the payoff to innovation and the speed with which new technology and new products become available.

Our scenario is by far the more dramatic. Romer notes in passing that, if the innovator were totally immune to the disincentives of spillovers, then none of the benefits would have gone to others. But, if that were so, then real wages would hardly have risen from their levels before the Industrial Revolution![3] It is almost impossible to imagine how great a difference that would have entailed. The best estimate puts U.S. per-capita GDP in 1820 at less than one seventeenth of what it is today, and even as late as 1870 real per-capita GDP is estimated to have been less than one ninth its current level (Maddison, 1995, pp.196-197). If we assume the most extreme case--that the spillovers from innovation are reduced to (anywhere near) zero--the living standards of the vast majority of the citizens of today's rich countries would have stalled at pre-Industrial Revolution levels. One can hardly accept the notion that it would be socially preferable to achieve a total GDP that is far higher than today's through enhanced incentives for innovation, while the bulk of the population is condemned to near-medieval living standards, but that is where such a premise leads us. Even the fortunate few innovators who might amass unimaginable wealth in such a zero-spillover world would undoubtedly prefer somewhat better conditions for their

---

[3] "This pattern of industrialization without wage gains is what it would take to ensure that the industrialist captures all of the benefits he creates when he introduces machinery....[this] cannot be a historically accurate description of the process of development in industrial countries, for if it were, unskilled labor would still earn what it earned prior to the industrial revolution" (Romer, 1994, p.29).

impoverished compatriots, and would themselves probably be better off not only in terms of the social environment, including reduced violence and disease. In addition, innovators would probably have higher absolute incomes because more can be produced by a labor force that is better fed, healthier and better educated. But any such gain to labor is unavoidably a rise in the percent of the return to innovation that goes to others than the innovator -- it is necessarily a rise in the share of spillovers.

II. Marginal Cost, Total Cost and Adoption of Socially Beneficial Innovations

We have, then, the reasonable notion that even if one measures optimality in terms of some absolute index of size of GDP, with no attention to distribution, that such optimality requires spillovers to be larger than zero and, as I will argue, probably very substantially so. But before turning to our formal model of the socially optimal level of spillovers from innovation, I must first review briefly an important issue that, as Romer notes, goes back to Dupuit: the role of total cost relative to that of marginal cost in "lumpy" decisions (such as the decision to build a bridge or to launch an innovation process). Where such an activity entails substantial fixed and sunk costs or where scale economies make small-scale entry infeasible, we cannot rely on marginal analysis, because marginal data relate only to small adjustments.

That is not to deny the important role of marginal considerations in the theory of innovation. For example, in deciding how much to increase or decrease the budget for R&D on improvement of a particular invention, or how much longer to work on it before releasing it to the market, the usual sorts of marginal calculation apply, as I have shown elsewhere (see Baumol, 1993). However, when deciding whether to build a bridge or to launch a large-scale research project, the pertinent criterion is whether the total yield will exceed the total cost. A profit-seeking firm will not undertake an innovation unless the total revenue is expected to be greater than the total cost, and society should not undertake it unless its expected total benefits exceed total costs.

The very legitimate argument conventionally linking this to the spillovers of innovation, then, is straightforward. It tells us that there are many prospective innovations that promise net social benefits but will nevertheless never be undertaken by private enterprise, even if marginal spillovers, at whatever margin is relevant, are all zero. This is because there are many potential innovations whose prospective total benefits are greater than their total costs, but because a considerable proportion of the total benefits go to persons other than the innovator (in the form of spillovers), no one will find it profitable to carry out those innovations.

In this article I will carry on with the story from this point, and provide a somewhat different ending. I will argue that the spillovers themselves are likely to bring substantial social benefits -- benefits that must be weighed against the forgone innovations before a judgment on optimality can be arrived at.

III. A Model and Graphic Analysis of Optimal Spillovers

I will use two basic premises to facilitate the discussion: 1) that production uses only two inputs, labor and innovation, so that income is derived only from those two sources, and income earners are divided into innovators and (non-innovating) workers; and 2) that, with a given labor force, the production frontier can be shifted outward only by innovation.[4]

I will focus on two alternative scenarios. In the first, the spillover ratio, S = benefit of innovation not accruing to the innovator / the total benefit contributed by innovation, is a modifiable parameter. In the other scenario the value of S will be considered fixed exogenously.

The model will also follow Romer in recognizing the vast set of potential innovations which differ in the magnitude of the total net gain that they offer to society, over and above their sunk costs. Each such innovation, I, also requires a sunk expenditure, C(I), where I is the index assigned to invention I, as described below. Here I assume that both benefits and costs can be translated into money terms and that the total gross benefit, B(I) (not excluding sunk cost), is given by the discounted present value of the stream of the benefits and other costs expected from invention I from now to eternity.

Then, clearly, maximization of the direct benefits from innovation requires that at any given time the economy carry out every recognized prospective innovation, I, for which $B(I) - C(I) > 0$. However, given the spillover ratio, private enterprise will undertake only those innovations for which $B(I)(1-S) - C(I) > 0$. This means that the beneficial innovations, J, for

---

[4]This is not to deny the vast contribution to output that is made by investment in human capital and in plant and equipment. But even here, it seems clear that innovation is, to a large degree, indirectly responsible. If new technology in agriculture, mining and manufacturing had not appeared with the Industrial Revolution and the centuries that preceded it, it is arguable that miserably low per-capita income levels would have prevented any substantial savings and therefore no significant investment in plant, equipment, education or health of the labor force would have been possible. Not even the size of the population and the labor force could have expanded as it did. If so, innovation must be considered the ultimate source of the bulk of the investment in human and physical capital that, along with the innovation itself, is surely responsible for most of the growth in production, production per capita and in productivity that followed the Industrial Revolution.

which C(j)/(1-S) > B(j) > C(j), will be lost to society. That is, roughly speaking, where the story stands in much of the literature.

To take it further, I will utilize a few simple graphs. Figure 1 is a standard depiction of the relationship between the benefits from innovation and the set of innovations that are actually brought to fruition from among those that are currently perceived by prospective innovators. It shows how spillovers limit the number of prospectively useful innovations that are actually carried out. The graph's only surprising feature is the magnitude of the shortfall below the maximum that can plausibly occur. In this graph it is assumed that the spillovers are a fixed percentage (e.g., S = 0.75) of the total future benefits of an innovation. The set of potential innovations currently recognized as possible objectives of R&D activity is taken to be a continuum (or we may prefer to assume that they can be approximated by one). It is also assumed that the sunk investment required to carry out a single innovation is fixed[5] at the level CC'. The horizontal axis, which extends from zero to unity, represents the share of currently recognized innovation possibilities that is actually carried out. Innovations are indexed <u>in descending order</u> of incremental gross benefit, B. Here, gross benefit is defined as the discounted present value of all current and future gains an innovation provides, minus the discounted value of all current and future costs other than the sunk costs needed to carry it out. The descending order of benefits, then, means that B(I) > B(j) iff I< j. For simplicity of presentation the gross benefits curve, B, is taken to be linear. It

---

[5]Professor Gadrey has noted that CC' is unlikely to be horizontal in reality, with more beneficial innovations apt to incur a higher sunk cost, so that the line (curve) may well have a negative slope. It is easy to see that this change leads to no fundamental modification of the analysis.

must have a negative slope throughout because of the way in which the potential innovations are ordered. Then one can easily also draw in the innovator's gross benefits curve, the lower straight line (1-S)B. Point N, where the B and C lines intersect, represents the exhaustion of all recognized innovations that currently promise a net gain to the economy. That is, at N the economy has adopted every recognized innovation that offers benefits that exceed its sunk costs.[6] However, private enterprise will be unable to go beyond point M, with its much lower output of innovation, because spillovers, SB, will prevent it from covering the sunk costs of any additional innovations.[7] The implication is that the level of innovation can be much below the socially efficient level, n.

It should be noted, incidentally, that this is not a problem of private enterprise alone. The public sector can do no better without financing the shortfall out of the, say, 90 percent gain that would otherwise accrue to the labor force. Of course, private innovators can do just as well if at the margin they can be subsidized or use some other source of revenue to decrease the share of total benefit that escapes them as spillovers.

The story becomes much more interesting in Figure 2, in which the share of spillovers is

---

[6]It is the association of point N with optimality in the set of innovations which are carried out that forces me to use the very broad definition of spillovers employed in this paper, rather than a more restricted concept, such as one including only unpaid for benefits derived by an innovator's competitors. Clearly, in itself, an innovation is beneficial to society if its costs are exceeded by the sum of the benefits to <u>anyone</u>, including consumers in other countries, unrelated producers, and members of future generations.

[7]More generally, it is easy to prove that di/dS < 0 at the least remunerative invention I whose production causes no loss at the given S value. For zero profit requires (1- S)B(I) - C = 0, so that (1-S)B'di = BdS, or, since B' = dB/di < 0 by construction, di/dS = B/(1-S)B' < 0.

varied. This time, the horizontal axis represents S, the size of the spillover ratio, which also ranges between S = 0 and S = 1. The upper curve, B*, is the integral of B(I) - C from I = 0 to the profit-maximizing value of I in Figure 1 where (1-S)B(I) = C. In the region to the left of the social maximum the slope of the B* curve must be negative since by construction dB*/di = B-C > 0 and from footnote 7 di/dS = B/ (1-S)B' < 0, so

(1) $$dB^*/dS = (dB^*/di)(di/dS) = (B-C)B/(1-S)B' < 0.$$

It is also easy to verify by taking the second derivative that in this region the B* curve is normally concave, as drawn.[8] This shape would persist throughout the graph if innovation were the only direct input to productivity growth. However, in this model productivity growth is also enhanced, up to a point, by better nutrition and education of the labor force, which are severely reduced as the value of S approaches zero. Thus, the lefthand end of the graph represents a state of affairs with workers ill-fed and ill-educated and so the B* curve begins to fall as S approaches zero.[9] All of this together yields the shape of the B* curve depicted in the graph. It has a unique

---

[8]For, continuing to assume B"= 0 for simplicity, and writing b = dB*/di = B+C, we have by (1) $d^2B^*/dS^2 = [(b'B + bB')(1- S)B'di/dS - bBB']/K^2 = (b'B^2)/K^2 < 0$, where K is the denominator of dB*/dS and, as shown in footnote 3, di/dS = B/(1-S)B'.

However, to the right of S = 1 - C/B(0) at which <u>any</u> investment becomes unprofitable there will be no investment, no costs and no benefits, so that the B* curve will follow the horizontal axis.

[9]An alternative graphic representation employs a separate curve representing the effect of loss of worker productivity as S approaches zero. With such a construct the B* curve will have a negative slope throughout, and the new curve would have to be subtracted vertically from the B* curve to show the net social consequences of variations in S. This representation would contribute clarity to the analysis. In particular, it makes clear that without an effect of S on worker productivity S=0 will return to the range of Pareto optimality, but that such a range will continue to be present. On the other hand, if rising S increases worker productivity throughout most of the interval between S=0 and S=1 then the innovator's maximum may well move far

maximum at S = n, at which current investment in innovation maximizes the net total gain to society.

However, the story is not the same for the two classes into which society is divided -- the innovators and the workers. For example, the benefits added by innovation going to the workers are given by the B*w curve whose expression is SB*, i.e., the product of social benefit, B*, and the share, S, that goes to noninnovators. This lower curve must start off at zero at its lefthand end, where S = 0, and it must approach steadily closer to the B* curve as S increases toward unity, with the two curves meeting at the right, where S = 1. B*w has the derivative

(2) $$dSB^*/dS = B^* + SB^{*\prime},$$

which is positive at the maximum, v, of the upper curve, B*, where $B^{*\prime} = 0$. This means that v, the maximum of B*w, must occur <u>to the right</u> of n, the maximum of B*. The reason is that to the right of the maximum of B*, though the total size of the social output pie is decreasing, the workers' share of that shrinking pie has increased sufficiently to make them better off on balance. Eventually, the size of the social output gain shrinks so much that further increases in S are damaging even to workers. The workers' benefit maximum requires $B^* = -SB^{*\prime}$, which has a straightforward intuitive interpretation in terms of the total benefit pie and the size of the workers' slice. B* is the workers' gain from a unit increase in their share of the pie, while $SB^{*\prime}$ is the loss to them from the accompanying decrease in size of the pie, $B^{*\prime}$, and maximization requires this marginal loss to equal the marginal gain.

---

toward the right, the range of Pareto optimality is apt to narrow, and a high spillover ratio will tend to serve everyone's interests.

Similarly, the graph depicts an innovators' benefits curve, B*I. Its equation is B*I = (1-S)B*. Its relation to the B* curve is perfectly analogous to that of the workers, except that the innovators' curve's behavior going from right to left (as 1- S increases) corresponds exactly to that of the workers' curve going from left to right (as S increases). Thus, S = m, the maximum[10] of B*I, must always lie <u>to the left</u> of the maximum of the social gain curve B*.

Several substantive conclusions follow from this discussion. First, it is clear that innovators do not obtain their maximum reward if they receive <u>all</u> of the benefits their innovations can yield, that is, if spillovers are zero. Rather, their gain is maximized at S = m. Their preferred value of S may even be higher than this if, for example, they derive some utility from altruistic egalitarianism, or if higher wages reduce crime and thereby increase the safety of the innovators.

For the same reason, the welfare of workers will be reduced if <u>all</u> of the benefits of innovation go to them in the form of spillovers, because then of course there will be no innovation and no benefits to accrue to them. They will be better off if S = v, corresponding to the maximum point on their benefits curve, and they may derive indirect gains from a value of S somewhat lower than v if greater innovator wealth leads to such things as increased donations to hospitals, from which workers benefit.

Thus, point n (at which productive efficiency is maximized) is <u>not</u> preferred by either party, and while n is a possible compromise between the two groups, it is hardly the only available compromise position. The entire range $m \leq S \leq v$ is a possible solution since any change from a

---

[10]This can conceivably be a corner maximum at S = 0. That will occur if at this point B* > B*' > 0, so the slope of the innovators' benefit curve, (1- S)B*' - B*, is negative at S = 0.

point within that range must harm one of the groups while benefitting the other. That is, the corresponding range of a utilities-possibility frontier, that is easily derived from our model, must have a negative slope throughout. We therefore have a range of S values all of which are Pareto-optimal, rather than a unique and clearly definable global optimum.[11]

IV. Pigouvian Subsidies, Consumers' Surplus and Lump-Sum Transfers

The preceding analysis raises two related questions: first, whether, at least in theory, one cannot have it both ways, somehow providing the incentive for the socially optimal amount of innovation, and a subsequent redistribution of the resulting gains, thereby making all parties better off. Second, one may well wonder whether all cases of external benefits are not subject to the same argument, with the beneficiaries of the spillovers having interests that conflict with those who provide the externalities, and conflicting also with the welfare of society as a whole, however that may be measured.

The difference here, as Romer rightly emphasizes, stems from the heavy sunk costs of the innovation process. These mean that even if the prices of the products of innovation are adjusted so as to cover the marginal costs of supplying them, the sunk costs will not be covered. Thus,

---

[11]In theory, of course, it would be most desirable for spillovers to be zero, letting the market maximize B* by adopting all innovations that promise positive social net benefits, and then redistributing these maximal benefits of innovation between innovators and workers through a lump-sum tax on the former. In practice, however, no one knows how to design such a tax, defined as one that has no incentive effects on behavior. In particular, if the amount collected from a particular innovator depended on that individual's ability to pay, and therefore varied with the amount the innovator invests, the tax would presumably reduce that investment and move the economy away from the highest point on B*.

even if Pigouvian subsidies are added to those prices, to bring them up to cover all of the pertinent marginal costs, private enterprise will not be willing to invest in a socially beneficial innovation. Here, the obvious test of social desirability is whether the total benefits of the innovation, including consumers' surpluses now and in the future can be expected to cover all of the costs, both those that are sunk and those that are not.

In the more usual version of the externalities story the context is a world of perfect competition, with no sunk costs or economies of scale, so that all that is needed to achieve optimality is a Pigouvian subsidy sufficient to increase the level of the spillover-generating activity by the amount that best serves the interests of society. Since here only relative prices matter, the optimum can be achieved by a tax on other activities that finances the required subsidy. The spillovers continue to go to those who enjoyed them before the tax, presumably in greater quantity than before, and the externality-generating activity is expanded as much as serves the public interest. Consumers' surplus need not be captured and transferred to suppliers of the activity who already are provided the optimal incentive by the product price as modified by the subsidy.

In the case of innovation, some of the spillovers can indeed be internalized through pricing, for example, via royalties obtained from patent licensing. But for reasons just explained, the price that satisfies the marginal optimality requirements will not provide revenues sufficient to cover the sunk costs when only the resulting consumers' surpluses are sufficient to offset those costs.

It is tempting to argue that one can solve the problem in theory by means of a patent system that perfectly protects intellectual property rights and enables innovators to acquire all consumers' surpluses by perfect price discrimination, with others subsequently compensated by lump-sum transfers. Romer emphasizes how difficult (if not impossible) such price discrimination is to carry out in practice. But even if it were easy to carry out, the lump sum transfers would remain a figment of the theorist's imagination. In general, it is possible to argue that there can be no such thing. Even a payment extracted from everyone regardless of wealth, education or any other attribute would presumably affect family sizes and would therefore not be lump-sum, and it is not clear how it could be collected from the penniless. But in the case of innovation, the unreality of the proposal is far more striking. Here the very high magnitude of the spillovers of innovation are particularly relevant. Imagine a setup in which innovators initially receive the, say, 80 or 90 percent of GNP that currently is made up of spillovers. If the remainder of the population is subsequently to be compensated by the ostensibly lump-sum transfers, from whom could the resources conceivably come? Only from the super-rich innovators, who could hardly avoid noticing that the bulk of their innovation payoff is to be taxed away. Surely, this would affect the amount of the resources and effort they were willing to invest in the innovation process. The idea that lump-sum transfers are pertinent to the problem, even in theory, is tantamount to assuming the problem away.

The bottom line, simply, is this: there is no way to escape the trade-off between the incentives required to elicit the "optimal" level of investment in innovation and the desire for the resulting rise in real productivity to benefit everyone, and not just the innovators.[12]

V. Size of the Range of Possible Optima

If the range of Pareto-optimal S values is narrow, then the interests of innovators and workers will be relatively close and conflict will be minimal. On the other hand, a wide range will make it more difficult for policymakers to determine to what degree it is best to tighten intellectual property rights and the strength of their enforcement. I will show next how the length of the range of possible optima can be determined. The conclusion will be that the flatter the B* curve is near its maximum the wider that range will be. There is a simple intuitive explanation that helps to bring out the logic of the issue. Flatness of the B* curve means that changes in the distribution of the benefits from innovation do not materially undercut those benefits -- the size of the total benefits pie shrinks slowly with changes in distribution. This can occur because a substantial rise in the spillover benefits to workers in that case does not greatly discourage investment in innovation, or the innovations that will be lost will be ones that offer a relatively modest productivity contribution, or because there are some other offsets. So, if B* is fairly flat, workers will not be faced with a substantial reduction in total available innovation benefits when their share increases. Similarly, flatness means that innovators will not lose much through a reduction in total innovative benefits resulting from a reduction in the workers' distributive share.

[12]A value judgment also distinguishes innovation spillovers from other beneficial externalities. There seems little equity reason why flower lovers should be protected from charges for their spillover benefits from commercial flower growers. But it is widely considered desirable for the whole population to benefit (without charge) from technical progress.

So, since a flat B* curve means that changes in income distribution will reduce the social output pie very slightly, each of our two groups can benefit substantially at the expense of the other if it manages to get a considerably larger share of that fairly fixed-sized pie. The opposite will be true if B* curves downward sharply from its maximum.

The points m and v and, thereby, the length of the range of possible optima, can be determined graphically. For this, note that for the workers' maximum, $dB^*w/dS = dSB^*/dS = SB^{*\prime} + B^* = 0$. Thus, the maximizing S is obtained from

$$S = -B^*/B^{*\prime}.$$

The calculation indicated by this equation is carried out graphically in Figure 3, where the 45-degree line represents S and the heavy curve represents $-B^*/B^{*\prime}$, which obviously has a vertical asymptote at the value of S corresponding to the maximum of B*, where $B^{*\prime} = 0$. The intersection of the two loci at S = v occurs at the point that satisfies the workers' maximum condition given by the previous equation.[13] The analogous calculation is carried out for the innovators[14] using the other diagonal for (1-S).

We have noted that $-B^*/B^{*\prime}$ is a curve that approaches the vertical line above v asymptotically as S declines, approaching B* max from the right. This is the relevant range in

---

[13]An obvious alternative construction uses the B* curve from Figure 2 and superimposes the curve $-SB^{*\prime}$. The intersection of these two curves gives the workers' optimum, v, with a similar construction for the innovators' optimum, m. This second construction is a bit more complicated graphically, but it is a bit easier to interpret economically as equating $-SB^{*\prime}$, the effect of a very small increase in S on the loss via reduced size of "benefit pie," with the associated marginal gain, B*, from the increased share.

[14]This assumes that at $S = 0$ $B^* < B^{*\prime}$ so that m is an interior maximum. See footnote 10, above.

which to seek v, the maximum of B*w, which has already been shown to lie to the right of B*max. The curve approaches infinity at B*max because near there the denominator, B*', approaches zero. Then, at its other end, S = 1, where B* = 0, the curve approaches zero. The intersection of the S line with the curve -B*/B*' will increase if the curve rises. But this will result from a flattening of B*. The flattening raises the value of B* in any subregion near B*max and, by definition, the flattening also reduces B*'. With the numerator raised and the denominator lowered, it is clear that the -B*/B*' curve will rise except at the end points where S =1 and where B* = Bmax. Thus, its intersection with the 45-degree line will move to the right. This demonstrates the result: a flattening of B* near B*max will bring B*w max further from B*max. Exactly the same argument holds for B*I. This confirms that a flattening of the B* curve near its maximum will increase the distance between the maximum of B*I and that of B*w.

VI. The Importance of Spillovers, and Some Implications for Policy

It should be recognized that there are ways in which policymakers can affect this story. They can either partially offset the effects of the spillovers or they can act to increase or decrease the spillovers from innovations in reality.

Policymakers can reduce the innovation-inhibiting effects of the spillovers without reducing the distributive benefits materially by providing public sector assistance to cover at least part of the sunk cost of innovation. In terms of Figure 1, that would shift the CC' curve downward toward the X axis. Clearly, that will increase the number of innovations it will pay private enterprise to undertake (though the increased taxes needed for the purpose would hardly be lump-sum). In theory, if the public sector would pay for a proportion of the sunk costs sufficient to move CC' through point H where private returns curve (1-S)B lies just above the socially efficient point, n, on the horizontal axis, private enterprise will be induced to invest in that socially efficient amount of innovation. Whether the implied role of the public sector in deciding which innovations to subsidize and the large amount of financing that would then have to flow through the public sector would itself serve as a significant brake on growth and innovation is a matter that must, of course, be weighed in the balance here. Certainly, the taxation required to carry out such a program is hardly lump-sum in character. Of course, to some extent this sort of activity is already carried out by governments, and their focus on basic research, whose spillovers are particularly large, seems quite appropriate. Still, where, as in the U.S., more than 70 percent of R&D is carried out by private industry (National Science Board, 1996, p.84), the government support is far from sufficient to make a fundamental difference for the issue.

The government can also affect the magnitude of the spillover ratio by legislating greater (or smaller) rights to the creators of intellectual property and they can increase or reduce the resources devoted to enforcement of those rights. For example, Japanese patenting laws are far less favorable to inventors than those in the U.S. and probably increase greatly the spillovers from Japanese innovation. No obvious and substantial decline in Japanese innovative activity appears to have resulted from this less protective atmosphere.[15] Even more important, with little protection available from the patent system, Japanese innovators appear to have been driven to

---

[15]Wolff has suggested the hypothesis that this may have resulted because more rapid dissemination resulting from decreased effectiveness of patent protection reduced the sunk cost of innovation sufficiently to offset the effect on profits from the increase in spillovers.

create profitable technology-sharing agreements with competitors and others, a situation that has ensured that inventions are rapidly disseminated and put to use throughout the Japanese economy. The consequence of all this may well have been a net enhancement of Japanese productivity growth (on this general subject and for more details on the relevance of the Japanese rules, see Baumol, 1993, Chapter 12), and illustrates how a spillover ratio substantially different from zero may be desirable in reality and not only in theory.[16]

But, perhaps most important, the argument offered here is surely consistent with views that are widely held, though that may not be immediately obvious. I believe I have never come across a discussion of the consequences of the Industrial Revolution that did not stress its ultimate contribution to the living standards of the population as a whole. Indeed, one can only begin to suggest the flavor of the shocking levels of poverty to which a world without the spillovers from innovation would condemn the labor force. Histories of Europe confirm that for many centuries before the Industrial Revolution the vast majority of the population struggled simply to exist. Most families spent nearly half their food budgets on "breadstuffs" (as late as 1790 in France, according to Robert Palmer, "The price of bread, even in normal times, in the amount needed for a man with a wife and three children, was half as much as the daily wage of common labor" (1964, p. 49), and for most of them the bread was of what was considered a very inferior variety. Still more commonly, it took the form of gruel (in good years) consumed in life-sustaining quantities. But there were many years when even gruel was unavailable. Devastating famines

---

[16] It must be acknowledged that this is hardly the first discussion arguing or implying that spillovers may not in fact lead to socially inadequate innovation expenditures However, so far as I am aware, these analyses are very different from mine.

threatened Europe as late as the beginning of the nineteenth century, and earlier had been a normal fact of life. Fernand Braudel's remarkable history of Europe (1979) documents the depths of human misery before the Industrial Revolution:

> "A few overfed rich do not alter the rule....Cereal yields were poor; two consecutive bad harvests spelt disaster.... Any national calculation shows a sad story. France, by any standards a privileged country, is reckoned to have experienced 10 general famines during the tenth century; 26 in the eleventh; 2 in the twelfth; 4 in the fourteenth; 7 in the fifteenth; 13 in the sixteenth; 11 in the seventeenth; and 16 in the eighteenth.... The same could be said of any country in Europe. In Germany, famine was a persistent visitor to the towns and flatlands. Even when the easier times came, in the eighteenth and nineteenth centuries, catastrophes could still happen....
>
> The poor in the towns and countryside lived in a state of almost complete deprivation. Their furniture consisted of next to nothing, at least before the eighteenth century, when a rudimentary luxury began to spread.... Inventories made after death, which are reliable documents, testify almost invariably to the general destitution...a few old clothes, a stool, a table, a bench, the planks of a bed, sacks filled with straw. Official reports for Burgundy between the sixteenth and the eighteenth centuries are full of references to people [sleeping] on straw...with no bed or furniture, who were only separated from the pigs by a screen.... Paradoxically the countryside sometimes experienced far greater suffering [from famines than the townspeople]. The peasants...had scarcely any reserves of their own. They had no solution in case of famine except to turn to the town where they crowded together, begging in the streets and often dying in public squares...." (Vol. I, pp. 73-75 and 284-286).

The Industrial Revolution wrought changes that defy intuitive grasp. In the U.S. per-capita GDP increased almost ninefold, while productivity per worker-year grew by a factor of nearly 13. Growing education, improved health of workers and investment all doubtless played some role in this. But without the enhanced resources provided by innovation it can be argued that little increased education, health care, food consumption or investment could have been afforded. It is reasonable to attribute the bulk of the rise in per-capita GDP and the rise in productivity to innovation, directly or indirectly. The ninefold rise in GDP per person, then, means that fully 8/9, or nearly 90 percent, of GDP in 1989 was contributed by innovation carried out since 1870. The total contribution of innovation is likely to be even greater than that since pre-1870 innovations, such as the steam engine, the railroad and many other inventions of an earlier era, still add to today's GDP. Moreover, the difference between the increase in productivity and that in GDP is attributable in good part to enhanced leisure, surely a benefit over and above reported GDP, and a benefit also made possible largely by innovation.

At the same time, the share of total investment income in GDP during this period was (depending on classification of investment income) certainly less than 30 percent. But investment in innovation is only a part of total investment. In 1996 R&D expenditures made up only 15 percent of total investment in the U.S., a ratio of about 1 to 7 (Statistical Abstract of the United States, 1997). Thus, if the return on innovation investment were the same (after adjustment for risk) as the return to investment of other types, the return to innovation would have been less than (30)(0.15) = 4.5 percent of GDP. Certainly it is quite implausible that returns to innovators can have been more than 10 percent of GDP. It follows that S, the spillover ratio, must have been greater than 90 percent of the total. This very crude estimate is not out of line with the estimates of private and social returns to innovation (see, for example, Nadiri, 1991; Mohnen, 1992; and Wolff, 1997). Wolff estimates the social rate of return to be 53 percent (in line with previous work on the subject) and the private rate of return to be between 10 and 12.5 percent or less (a figure slightly below earlier estimates) (Wolff, 1997, p. 16). All of this at least indicates the conservatism of the 75 percent figure derived from the share of labor income in GDP. For our story, the precise figure does not matter. What is noteworthy is that the spillover ratio seems clearly to be surprisingly large.

VII. Concluding Comment

Most of us do recognize the spillovers that innovation seems to have contributed, and, more than that, we all seem to agree that this enhancement of living standards is a very desirable outcome. In other words, perhaps without realizing that we were discussing the spillovers of innovation, we have concluded that they are a very good thing-- indeed, that they are the real achievement of the Industrial Revolution and the innovations that drove it. If this is a valid characterization of the prevailing evaluation (with which I fully concur), then it follows immediately that even if zero spillovers had increased innovation they would certainly have been far from optimal.

Here, the standard reaction of economists--that disinterested academicians cannot defensibly take a stand on income distribution--just will not do. Virtually no one aspires to a world in which innovators receive incomes in the trillions of dollars (putting Mr. Gates' income into the shade), while the remainder of the community languishes in seventeenth-century poverty. Of course, there is a value judgment involved, but only cowardice will induce us to reject it. Once agreed to, then, we must also go on to reject the conclusion that spillovers are incompatible with optimality in the growth process. And once the last conclusion is recognized the remainder is a matter of haggling about degree of deviation from zero. I myself believe that the most desirable value of S is very much larger than zero. For, surely, it is widely and appropriately accepted that the main benefit of the Industrial Revolution is the remarkable increase in average per-capita incomes, and in real wages, more particularly.

Of course, it is generally recognized that those innovations that have never been born constitute a loss to society. But the point in the analysis here is that there is an inescapable tradeoff between two desirable phenomena: further increases in innovative activity and investment, versus diversion of the benefits to bring society out of medieval poverty, to spread education and health care, and to finance the better life not just for the fortunate few, but for the population as a whole.

If there is such a tradeoff, and I do not see how that can be denied, we are back in the realm in which economists are most comfortable -- in Lionel Robbins' justly noted words, we are back at the allocation of scarce resources among competing (and desirable) ends. The analysis of such tradeoffs is the meat and potatoes of our professional activity. And an introduction to the analysis of the resulting choice between low and high spillover ratios is what this paper has sought to provide.

References

Baumol, William J., Entrepreneurship, Management and the Structure of Payoffs, Cambridge, Mass.: MIT Press, 1993.

Braudel, Fernand, Civilization and Capitalism, 15th to 18th Century, New York: Harper and Row, vol. 1, 1979.

Gomory, Ralph E. and William J. Baumol, "Maximal Gains from Trade and Conflicting National Interests," mimeo, 1997.

File: Inovspi3.art 12-12-97

Grilliches, Zvi, "Issues in Assessing the Contribution of Research and Development to Productivity Growth," Bell Journal of Economics, Vol. 10, 1979, pp. 92-116.

Mohnen, Pierre, The Relationship Between R&D and Productivity Growth in Canada and Other Industrial Countries, Ottawa: Canada Communications Group, 1992.

Maddison, Angus, Monitoring the World Economy, 1820-1992, Paris: OECD, 1995.

Nadiri, M.I., "Innovations and Technological Spillovers," New York University, 1991, unpublished.

National Science Board, Science and Engineering Indicators, 1996, Washington, D.C.: U.S. Government Printing Office, 1996, NSB 96-21, p.84.

Palmer, Robert, The Age of Democratic Revolution, Princeton: Princeton University Press, Vol. 2, 1964.

Romer, Paul, "New Goods, Old Theory and the Welfare Costs of Trade Restrictions," Journal of Development Economics, Vol. 43, 1994, pp. 5-38.

Scherer, F.M., "Firm Size, Market Structure, Opportunity and the Output of Patented Inventions," American Economic Review, Vol. 59, 1965, pp. 1097-1125.

U.S. Bureau of the Census, Statistical Abstract of the United States, 1997, 117th ed., Washington, D.C.: U.S. Government Printing Office, 1997, pp. 447, 455, 603.

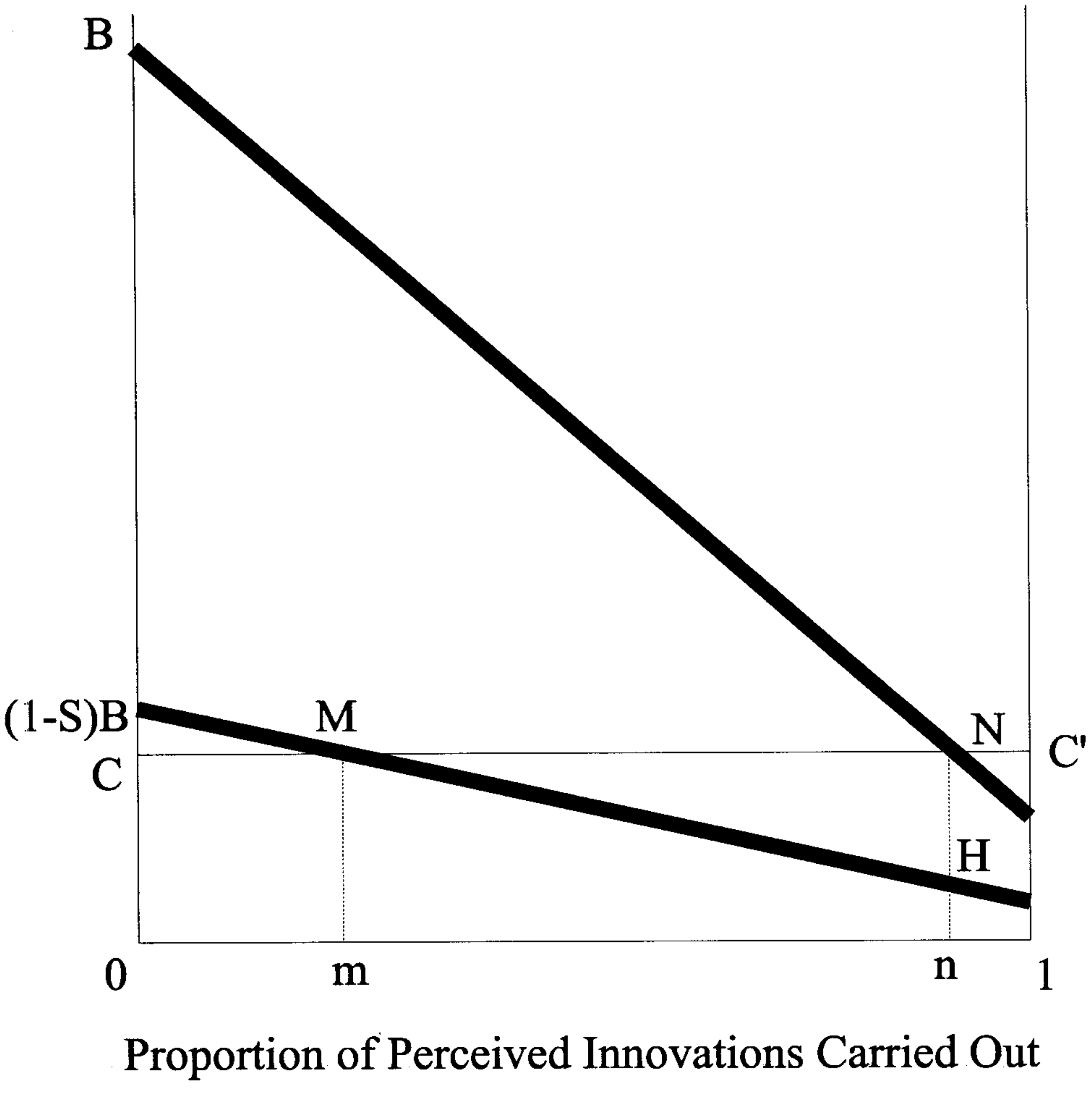
B
(1-S)B
M
N
C'
C
H
0
m
n
1
Proportion of Perceived Innovations Carried Out

Figure 1

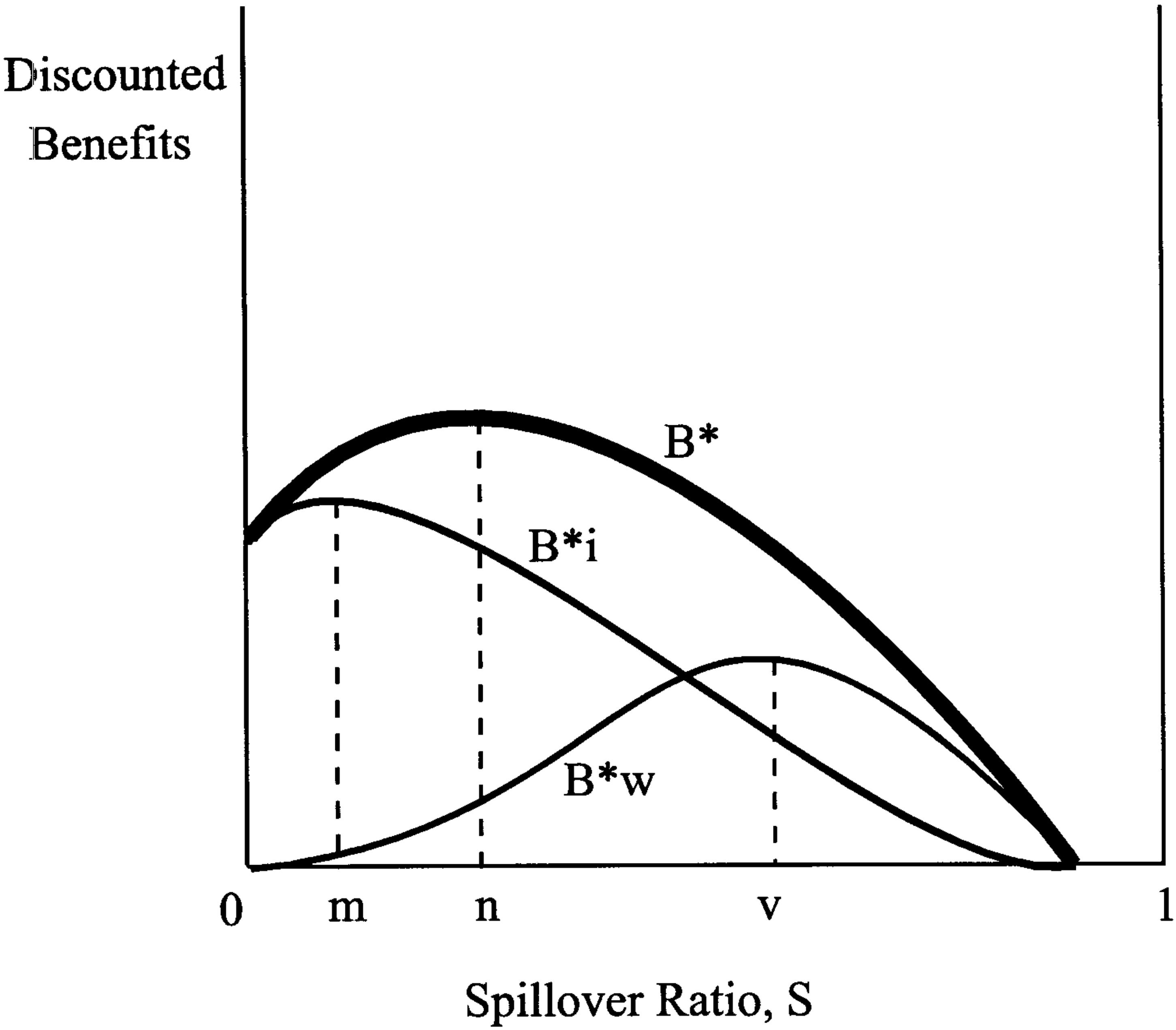

Figure 2

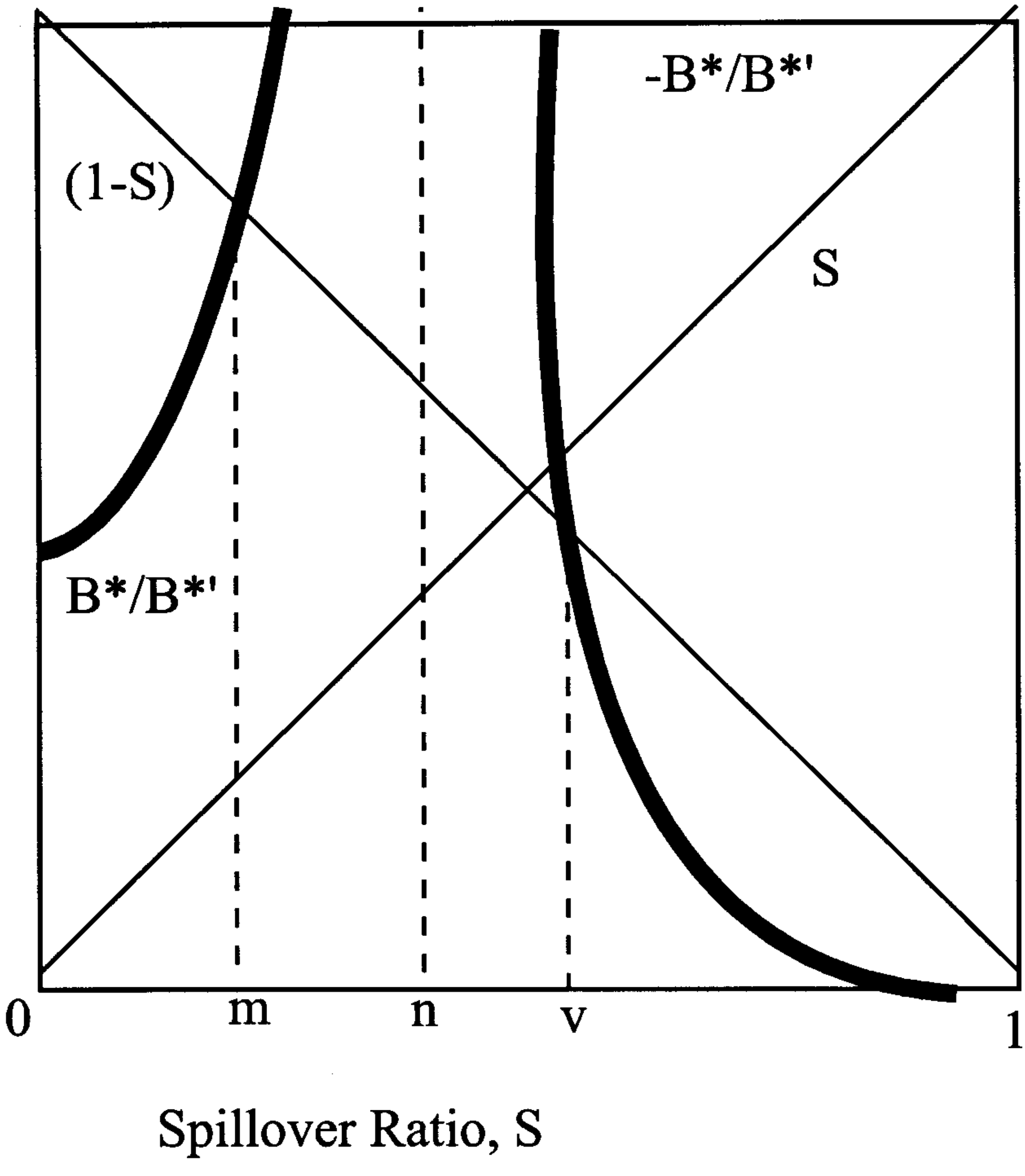

Figure 3