NBER WORKING PAPER SERIES

THE WELFARE ANALYSIS OF PRODUCT INNOVATIONS WITH AN APPLICATION TO CT SCANNERS

Manuel Trajenberg

Working Paper No. 1724

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1985

I acknowledge with deep gratitude Professor Zvi Griliches' continuous support and guidance throughout my research endeavors. I am also grateful for the generous financial support provided to me this summer by the USIA through its American Research Fellowship Program, and for the good auspices of the NBER Summer Institute. Participants in the NBER Productivity Workshop held in July 1985 offered useful suggestions and comments. The research reported here is part of the NBER's research program in Productivity. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #1724
October 1985

The Welfare Analysis of Product Innovations with an Application to CT Scanners

ABSTRACT

The main goal of this paper is to put forward a methodology for the measurement of product innovations using a value metric, i.e., equating the 'magnitude' of innovations with the welfare gains that they generate. This research design is applied to the case of Computed Tomography (CT) Scanners, a revolutionary innovation in medical technology. The econometric procedure centers on the estimation of a discrete choice model (the nested multinominal logit), that yeilds the parameters of a utility function defined over the - changing - quality dimensions of the innovative product. The estimated flow of social gains from innovation is used primarily to compute a social rate of return to R&D, to explore the interrelation between innovation and diffusion, and to trace the time profile of benefits and costs, the latter suggesting the possible occurance of 'technological cycles'.

Manuel Trajtenberg
Department of Economics
Tel-Aviv University
Ramat-Aviv
Tel-Aviv 69978
Israel

CONTENTS

I. Introduction

Technological change has long been regarded as an economic phenomenon of paramount importance, be it in accounting for the secular growth of the economy at large, or the dynamic performance of individual sectors. Yet, our understanding of it remains precarious, partly because theory had been for too long in the grip of the 'circular flow' mode of thought (in Schumpeter's words), but more importantly, because of the pervasiveness of measurement problems. This means, first, that many of the basic notions used to describe and analyze technical change do not have clear empirical counterparts (e.g., stock of knowledge, appropriability, spill-overs, etc.), and second, that the data required to quantify those notions - when at all possible - are for the most part unconventional and difficult to obtain.

Those problems are the most severe when it comes to the study of product - as opposed to process - innovations, since these come about and manifest themselves in dimensions other than prices and quantities (the familiar turf of economists), and cannot be easily reduced back to them. This is rather disturbing, particularly so in view of the fact that product innovations have become in the last few decades the most prevalent form of technical change.[1] Moreover, it is quite clear that the measures of technical advance commonly used, do not and cannot capture the impact of this type of innovations upon the economic magnitudes of interest, be it 'real' income or productivity growth (see Griliches [1979]). Thus, we are at risk of being constantly misled by economic indicators that purport to gauge overall economic performance, but in fact overlook a main ingredient of such performance. At the micro level, in

---

[1] See, for example, Mansfield [1968] and [1977], and Scherer [1980].

turn, we lack for the same reason basic empirical evidence - or 'stylized facts' - about the innovative process, and hence the proliferation of theoretical models, often yielding contradictory results, that cannot be either supported or refuted.

Responding to those concerns, it is our main goal here to conceptualize and give an operational content to the notion of product innovations, in order to enable their empirical measurement. Mindful that the 'acid test' of any research design lies in its actual implementation, we apply the resulting methodology to a specific innovation in the realm of medical technologies, namely, Computed Tomography (CT) scanners. Even though there is little - if any - novelty in the theoretical and econometric building blocks of the proposed methodology, we know of no previous attempt to explicitly measure product innovations in an economically meaningful way.[2] Thus, we dwell extensively on a variety of implementation issues and, although these are discussed with reference to a specific case study, what can be learned from them is hopefully of general interest.

A remark regarding the meaning of an economic 'measure' of innovations: given that the 'output' of innovative activity does not present itself in countable units of any sort,[3] a quantifiable dimension for innovations can only be defined in <u>value</u> terms, i.e., in terms of their impact upon social welfare. In other words, the question of 'how much' innovation there was or occured in a

---

[2]The only study that comes close is Mansfield <u>et al</u> [1977], but product innovations are in fact treated there as process innovations.

[3]Strictly speaking, such output consists of 'bits of new and socially useful knowledge', which is clearly an all but impossible notion to quantify. Patent counts will not do either, in view of the enormous variance in their scientific, technological and economic significance.

certain field during a certain period of time is exactly equivalent to, and can only be interpreted as asking, how much additional social surplus (i.e., consumer plus producer surplus) was generated by those innovations. Other measures such as counts of patents or of 'important innovations', rates of change of attributes, sales of new products, etc., could at best play the role of proxies, and their accuracy as such can be judged only by relating them to the value measure put forward here.

II. The Definition and Measurement of Product Innovations

The natural conceptual framework in which to analyze product innovations is the so-called 'characteristics' approach.[4] In it, products are thought of as bundles of attitudes that enter explicitly in the utility function of consumers, rather than as abstract entities having only a quantity dimension. The analytical advantages of such an approach - pertinent to the study of innovation - are immediately apparent: first, it vastly reduces the dimensionality of the optimization problem confronting agents in markets for differentiated and technologically progressive products, since a relatively small number of characteristics can span a much larger and even growing number of products. Second, it allows one to analyze explicitly substitution effects, and hence to derive the demand for new - actual or potential - products. More generally, the representation of products as vectors in attributes' space rescues the notion of product quality from the sterile domain of exogeneously given tastes, thus allowing to separate analytically between quality changes and changes in

---

[4]Actually, this is not a unique, universally agreed-upon (sub)paradigm, but an umbrella under which a diverse collection of analytical set-ups have been formulated, e.g., Lancaster's original demand model and his subsequent generalization (Lancaster [1966], [1971] and [1979]), the household production function (Becker [1965], Muth [1966], etc), the "hedonic" model of Rosen [1974],

preferences.[5] Quite surprisingly, however, the characteristics' approach has rarely been applied to the study of product innovation: most existing applications have to do instead with static product differentiation. We hope to make some progress here in filling in the gap.

We start therefore by representing a product - or brand - within a given 'product class'[6] as $s_i \equiv (z_i,p_i)$, $z_i = (z_{i1},z_{i2},\ldots,z_{im})$ being the vector of relevant characteristics (that is, $U_i = U(z_i)$, $\partial U/\partial z_{ij} \neq 0$), and p_i its price. Thus, the choice set from which the consumer selects the most preferred brand is $S = (s_1,s_2,\ldots,s_n)$.

Following the discussion on the nature of characteristics in Trajtenberg [1979], we distinguish between concatenable and non-concatenable characteristics,[7] the former being formally defined as

$$z_{ij} = f^j(x_i), \qquad \frac{\partial z_{ij}}{\partial x_i} \neq 0 \qquad \text{for} \quad x_i \geqslant 0$$

and the latter as

---

etc.

[5]That is, in traditional demand analysis the qualities of products are embedded in the structure of preferences, and therefore quality changes manifest themselves as, and cannot be disentangled from changes in tastes. An important implication of being able to separate between the two phenomena is that it is then much more plausible to assume that preferences, defined over attributes rather than over a given set of goods, do in fact remain stable in the face of a changing universe of products. Needless to say, the stability assumption is indeed central for much of micro theory.

[6]By 'product class' we mean a well-defined set of close substitutes located in a common space of attributes, and comprising a separable utility branch. We shall return to this rather recalcitrant issue when discussing the specification of the decision tree.

[7]This terminology, borrowed from the theory of measurement (see Krantz et al [1971]), was meant to focus attention on the physical properties that underlie the different kinds of measurement, and their implications for economic

$$z_{ij} = g^j(\theta_i), \qquad \frac{\partial z_{ij}}{\partial x} = 0 \qquad \text{for all} \quad x_i > \theta_i$$

where θ_i denotes the 'natural unit' of product i and x_i its quantity. Typical examples of concatenable characteristics are proteins in food products, or carrying capacity of vehicles, i.e., the amount of the characteristics available to the consumer is a monotonic function - usually linear - of the quantity of the product(s) that he/she chooses to consume. Non-concatenable characteristics, on the other hand, are much closer to the intuitive notion of 'quality', that is, they are properties inherent to the product as such, and do not vary with its quantity (e.g., speed of vehicles, aperture of photographic cameras, etc.). Therefore, different amounts of characteristics can be obtained only by switching products, and not by quantity adjustments. This distinction carries interesting implications for economic behavior and competition, as suggested in Trajtenberg [1979]. However, for our purposes here the relevant issue is that, for there to be a distinct and meaningful notion of product innovation, some of the characteristics (i.e., at least one) have to be non-concatenable. Otherwise, the choice set would be homogeneous of degree zero in prices and characteristics, which in turn implies that, regardless of their preferences, consumers would necessarily be indifferent between price reductions and proportional increases in the per-unit quantity of all characteristics. But

---

behavior. Concatenation is an operation by which objects are connected with respect to some common attribute, alowing for 'extensive measurement' (e.g., the placing of rods edge to edge for the measurement of length). Other authors have used terms such as additive, or combinable (and their respective antinomes) to make similar - if less precise - distinctions between types of characteristics.

if so then the distinction between process innovations (normally associated with costs and hence price reductions) and product innovations (which is intuitively linked to changes in attributes) all but vanishes, at least from the viewpoint of their effects. Furthermore, consider the case where there is a change in, say, product $s_i = (z_i, p_i)$ such that $s_i' = (\lambda z_i, \lambda p_i)$, $\lambda > 1$. If all characteristics were concatenable, then the transition from s_i to s_i' could not be regarded as an innovation at all, for it is totally inconsequential for behavior, and more importantly here, for welfare (it should be clear that $V(s_i, y) = V(s_i', y)$, where $V(\cdot)$ is the indirect utility function and y money income). To sum up, the point is that the notion of product innovation is inextricably related to and presupposes the existence of an independent quality dimension, and since non-concatenability is essential for the latter, it is by extension a sine qua non for the former.

With this caveat in mind, we define product innovation simply as dS/dt (or just ΔS), that is, as improvements over time in existing products, and/or additions of new products to the set.[8] This is, however, just a first step: given the lack of commensurability between attributes of products in a given set, and more so between attributes of products belonging to different product classes, ΔS cannot render by itself meaningful economic measures of product innovations. In fact, ΔS comprises just some of the raw data needed to construct such measures. If, for example, S refers to computers, how can an x percent increase in, say, speed of computation be linked to a y percent increase in storage capacity? And moreover, how do these advances compare to an

---

[8]Actually, ΔS may consist at times of 'negative' product innovation, i.e., of the deterioration of existing products, and/or the net contraction of the choice set. The definition above was meant to be general and encompass those cases as well, and hence to be more precise one should substitute in it

improvement in the gas mileage of cars, or in the fideltiy of a stereo set? This resembles the problem arising in the construction of any aggregate measure of economic performance, such as GNP: without appropriate price indices, the rate of change of individual outputs are of little help. Furthermore, these prices are supposed to reflect both marginal benefits to consumers and marginal costs, so that the resulting value measures have normative significance. Similarly, what we need to know in the present context is not, say, how much faster and 'smarter' are today's computers, but what is the 'worth' of such improvements to consumers, and the value of resources spent to bring them about.

Thus, what we want is a 'surplus' - or net social benefit - function $W(S)$, such that the extent of product innovation (which is, to insist, tantamount to the social gains accruing from it) would be measured by

(1) $$\Delta W_t = W(S_t) - W(S_{t-1})$$

Assuming for a moment, and for illustrative purposes only, that $W(\cdot)$ is linear in the characteristics,[9] the problem resides in finding a set of 'marginal utility' coefficients, β_j, such that the changes (innovations) occurring from one period to the next in the set of products offered to consumers would be evaluated by

(2) $$\Delta W_t = \Sigma_i \Sigma_j \beta_j [(z_{ij})_t - (z_{ij})_{t-1}]$$

'changes' for 'improvements.'

[9]That will certainly not be the case in the analysis below; the linearity assumption should be seen just as an expositional device.

However, these β_j's are not directly observable but have to be inferred from an indirect piece of evidence, namely, the relative frequencies of actual choices (i.e., the market shares of the different products in S), which presumably reflect the structure and distribution of preferences in attributes' space.[10] The main task ahead is thus to model the demand for products in S, so as to be able to formulate an appropriate W function, and estimate something akin to those β_j coefficients.

a. <u>The scope of ΔW and its limitations</u>

In the foregoing discussion we have referred to changes in characteristics, without mentioning prices explicitly. In fact, ΔS was meant to comprise changes in prices as well, and hence ΔW_t refers to the 'net' social benefits from having the set S_t rather than S_{t-1}, all things considered (think of price in (2) as another characteristic, and the corresponding coefficient as the marginal utility of income). It could be argued, however, that this is too broad a measure, and that for some purposes the impact of changes in characteristics (the 'pure' product innovation component) should be separated from the welfare effect of price changes. After all - so the argument goes - price changes often reflect phenomena that have little to do with innovation per se, e.g., changes in oligopolistic pricing strategies, variations in input prices, etc. Even though such a view is not without merit, we shall still stand by the broad definition of ΔW, primarily because of <u>pragmatic</u> reasons: First, almost any plausible W function will not be additive-linear in its arguments, and therefore

---

[10]To put it in intuitive terms: If, for example, consumers are observed to choose more often TV sets having sharper images rather than those with larger screens (controling for all other attributes and price), then the β_j associated with 'image quality' will be large relative to that of 'screen size'; this in

it is not clear how the two effects could actually be sorted out, without resorting to arbitrary procedures (local approximations would be helpful only for small changes, and not for discrete displacements of the choice set). Second, a great deal of what goes into ΔS usually consists of additions of new products, and the dropping out of outmoded ones; for those 'births' and 'deaths' it is simply meaningless to talk of a separate price effect, since by definition there is no prior or posterior reference point to which the current price of the new good (or the last price of a displaced brand) could be compared. This problem is even more serious in view of the fact that in many instances it is very difficult to distinguish between a change in an existing product from, say, s_i to s_i' on the one hand, and the simultaneous birth of s_i' and death of s_i on the other hand. In any case, it should be borne in mind that the whole point of the methodology to be presented here is to deal with technologically dynamic sectors, that is, with classes of products where innovation is the dominant force, propelling directly or indirectly most of the observed changes, be they in characteristics or in prices. Quite apart from the practical issues just mentioned, it is clear that for those sectors the broad definition is indeed the most appropriate.

A brief comment on the scope of W: as suggested previously $W(\cdot)$ stands for what is usually called the 'total surplus' function, that is, the sum of consumer and producer surplus. There is, however, an important difference

---

turn will be reflected in the benefits from innovations in TV technology, which are to be computed as some function of the improvements in image quality and screen size, weighted by their respective marginal values.

between them: profit (producer surplus) is a well-defined magnitude whose measurement does not pose special difficulties, except for the standard practical problems of reconciling accounting with economic profits, and obtaining the appropriate data. The fact that we are dealing with innovative sectors rather than with static products is, quite clearly, of no consequence whatsoever in that respect. The problem lies entirely in the specification and estimation of the - changes in - consumer surplus; thus, we shall for the most part ignore profits and associate ΔW with the gains from innovation that are not appropriated by the innovating firms, but passed on to consumers.

Finally, two limitations of the approach put forward here should be brought into the open. First, the proposed measure of gains from innovation, ΔW_t, actually excludes the possibility of directly assessing 'radical' innnovations, that is, the introduction of entirely new products that cannot be fitted into existing product classes, but span instead new classes of their own. The reason is plain and simple: in order to compute ΔW, there has to be an S and an a W(S) to begin with, that is, something to compare the innovation to (the null set does not qualify for obvious reasons), and a yardstick to evaluate the difference between old and new. However, these are precisely the conditions that are not fulfilled by radical innovations.[11]

It could be argued that any new product, be it as revolutionary as it may,

---

[11] It is interesting to note that Kuznets had clearly foreseen this limitation long ago; in a seminal paper on the nature of innovative activity he wrote:

> "...If an invention involves a new product, a rough approximation to its economic magnitude seems to be possible only if the new product can be treated as a substitute for the old one so that it again becomes feasible to estimate the additional yield and seek for a defensible economic basis for evaluating it." (Kusnetz [1962], p. 28).

could be seen as belonging in some way or another to existing product classes,[12] and would therefore be amenable to this kind of welfare analysis. That may well be the case at some ultimate level of abstraction, but the relevant issue is, once more, pragmatic rather than conceptual: is there enough in common between the presumed breakthrough and existing product sets, so that the innovation could be meaningfully evaluated in the context of preferences already defined over those sets? Or, alternatively, does its appearance bring about a reformulation of preferences, very much like quality changes do in the context of traditional demand theory? If the latter is true, then we cannot estimate directly (i.e., via ΔW) the gains resulting from the introduction of the radical innovation, that is, the discrete jump between not having anything of the kind, and having the first commercial units. Of course, we can estimate the benefits stemming from improvements from then on, provided only that enough variants (i.e., different brands) of the new product appear in the market, so as to make the parameters of the emerging preferences statistically identifiable. However, the limitation is not unsurmountable: we shall show later on a way of obtaining indirect estimates even of those elusive initial gains, that rely upon the interaction over time between innovation and diffusion.

The second limitation has to do with the fact that, by confining our attention to the evaluation of changes within a given set S_ℓ, we leave out spill-over effects, that is, improvements in other product sets (or processes) resulting from the original advances in S_ℓ. There is nothing in the approach itself that precludes, in principle, the inclusion of such externalities: if

[12]To put it differently, it can be argued that there is a limited number of higher-level, all-encompassing 'wants' a la Karl Menger, or utility branches a la Gorman or Strotz, so that any conceivable object, if desired and hence consumed at all, will fall by necessity into one of those categories.

there was information tracing these ripple effects from, say, ΔS_ℓ to ΔS_k, then one could compute also ΔW_k, add it to ΔW_ℓ, and assign the total to the original innovation. It is, however, very unlikely that such information could be obtained, and moreover, it is extremely hard to determine how much of ΔS_k was actually due to ΔS_ℓ, as opposed to independent developments in the other sector itself. Likewise, as time elapses the once specific and localized scientific and technological advances underlying ΔS_ℓ become part of the general stock of knowledge, making the identification of causal nexus between particular innovations, and the proper assignment of down-the-line benefits a rather hopeless task. Thus, the measures arrived at by looking only at ΔS_ℓ should be regarded as conservative estimates, or lower bounds to the 'true' social gains generated by any particular innovation.[13]

III. The Formal Derivation of the Surplus Function

As mentioned above, ΔW_t stands for the gains, in terms of consumer-surplus, generated by changes over time in the set of available products. Thus, the function $W(S_t)$ is to be obtained by integrating an appropriately defined demand system, which characteristics would depend in turn upon the nature of the choice set S, e.g., on whether it is continuous or discrete, whether it comprises a finite or an infinite number of porducts, etc.

In order to gain some intuitive understanding of the matter, let us review first the simple case whereby S consists of a single product having one quality

[13]An innovation might also have what could be regarded as 'negative externalities': Its appearance may render existing products obsolete 'before their time' or, looked at from a more basic viewpoint, it may speed up the depreciation of the social stock of knowledge. It is not as yet clear whether these effects should be taken into account, and if so, how.

dimension, q.[14] Denoting by $x(p,q)$ the demand function, and assuming away income effects, the benefit to consumers of a 'small' quality improvement is:

$$\frac{\partial W}{\partial q} = \int_{p}^{\infty} \frac{\partial x(s,q)}{\partial q} \, ds \tag{3}$$

or, for a sizeable change from ,say , q to q',

$$\Delta W = \int_{p}^{\infty} [x(s,q') - x(s,q)]ds \tag{3'}$$

Note that (3) and (3') measure the additional area under the demand curve, brought about by its upward shift in response to the change in quality. To put it differently, (3) stands for the marginal willingness to pay for quality improvements, which is precisely the conceptual tool needed to evaluate product innovations. It is worth keeping this in mind since the interpretation of ΔW is basically the same in all the models of interest.

A brief note on the continuous case. Strictly speaking, this refers to choice sets having infinitely many products that span a continuous spectrum in attributes space. More pragmatically, the assumption can be taken to mean that there is a sufficiently large number of contiguous products (and corresponding markets) so that, for all practical purposes, consumers and producers can be thought of as engaging in marginal optimization. This case has been examined in a seminal paper by Rosen [1974] and, notwithstanding some remaining problems, it seems that the framework and tools developed there could be adapted for the assessment of product innovations. Needless to say, that would be warranted

---

[14]For a detailed analysis of various versions of this case, see Willig [1978]; see also Spence [1975] for a particular application of the same construct.

only in those instances where at least the mild version of the continuity assumption is empirically justified.

In the discrete case it is assumed that the choice set consists of a finite number of different brands, these being located far enough from each other in attributes space so as to preclude marginal adjustments.[15] This is, in our view, the case that corresponds more closely to technologically progressive products, both theoretically and empirically,[16] and we shall therefore concentrate on it. Besides its referring to a property of the choice set, discreteness is commonly used also to describe a behavioral pattern, i.e., when consumers purchase a single unit of a single product out of the set, making the choice problem exclusively qualitative rather than quantitative (incidentally, note that neither aspect of discreteness entails the other). We shall assume throughout that discreteness holds also in this latter sense, although the analysis can be easily extended to accommodate cases of continuous/discrete choice as well (see, for example Hanemann [1984]).

Having thus defined the setting of the problem, we can now draw from the extensive literature, on discrete choice in order to specify the demand side and derive the surplus function.[17] For the benefit of readers unfamiliar with this

---

[15]Recall also that we assume non-concatenable characteristics, and hence we rule out the possibility fo spanning a piece-wise continuous choice set from a finite number of brands, as in Lancaster [1971].

[16]Simply put, the theoretical argument is that investment in R&D constitutes a fixed cost, and hence in any innovative sector there will be in equilibrium a <u>finite</u> number of firms (brands), n^*. Moreover, we know from the theory of monopolistic competition that n^* is inversely related to the height of those fixed costs, and hence the more R & D intensive a sector is, the smaller - ceteris paribus - n^* would be. Thus, it is more reasonable to characterize technologically progressive sectors as having discrete rather than continuous choice sets; this is borne out by what we usually observe in the real world.

[17]We rely here primarily on McFadden [1981], and Small and Rosen [1981].

literature, and in order to set the stage for the subsequent analysis, we review first the main ingredients of these models.

The basic hypothesis underlying discrete choice models of demand for differentiated products is that consumers maximize a random utility function,

$$\text{(4)} \qquad \max_{i,m} U_i = U(z_i, m; h) + \epsilon_i$$

$$\text{s.t.} \quad p_i + m = y, \quad s_i \in S$$

where m stands for a composite 'outside' good whose price has been normalized to unity, h is a vector of observable attributes of the individual, and ϵ_i is an i.i.d. random disturbance (the rest of the notation has been defined above); for notational simplicity the subscript for individuals has been omitted everywhere. The error term ϵ_i encompasses unobserved - and presumably less important - attributes of the individual and the product, and perhaps also an idiosyncratic and irreducible element of randomness in tastes. Thus, individual choices cannot be predicted with certainty, but depend upon the - unobserved - realizations of ϵ_i. From the researchers' point of view, this implies having to construct probabilistic rather than deterministic demand functions, conditional on S and h .

The problem in (4) is formally solved in two stages: first, the consumer chooses the optimal quantity of m given s_i, and then the optimal $s_i \in S$. Since by assumption only one unit of the product is purchased, the first stage is trivial, i.e., the optimal quantity of the outside good - conditional on s_i - is just $m^* = y - p_i$. Substituting it for m in (4) renders the conditional indirect utility function

$$(5) \qquad V_i^c = V(z_i, y - p_i; h) + \epsilon_i \equiv V_i + \epsilon_i = \underset{m}{\text{Max}}\ U_i$$

which the consumer maximizes in the second stage: $\underset{i}{\text{Max}}\ V_i^c$, s.t. $s_i \in S$. Given the discreteness of S, product s_i will be chosen <u>iff</u> $V_i^c \geqslant V_j^c$ for all $j \neq i$, that is, <u>iff</u>

$$(6) \qquad V_{i-j} \equiv [V(z_i, y - p_i; h) - V(z_j, y - p_j; h)] \geqslant (\epsilon_j - \epsilon_i)$$

Denoting by $f(\epsilon_i)$ the density function of the residuals and by $F(\epsilon_i)$ the corresponding cumulative distribution, and recalling that they are assumed to be i.i.d., we can define the probability of choosing s_i as

$$(7) \qquad \pi_i \equiv \Pr(s_i \mid S, h) = \int_{-\infty}^{\infty} \Pi_{j \neq i} F(\epsilon_i + V_{i-j}) \cdot f(\epsilon_i) d\epsilon_i$$

The exact form of π_i will depend, of course, upon the distribution of the error terms: if, for example, these are normally distributed, then (7) is just the cumulative normal, and the resulting model is the Probit. On the other hand, if the residuals conform to the type I extreme-value (or Weibull) distribution, then π_i will be logistic, i.e.,

$$(8) \qquad \pi_i = \exp(V_i) \Big/ \sum_{j=1}^{n} \exp(V_j) \qquad\qquad i = 1, \ldots, n.$$

This is the well-known conditional multinomial logit model, hereafter referred to as the MNL. The Probit and the logit - and some variants of them - have been and remain the most commonly used choice models, at least within the random utility framework. In our case the Probit had to be discarded from the outset for pragmatic reasons: we had to estiamte models involving choice sets of up to twenty alternatives, whereas with present-day computer capabilities the Probit can successfully handle only up to four or five. Thus, we shall use throughout the MNL, paying due attention to the limitations of the model that stem primarily from the underlying assumption of 'Independence of Irrelevant Alternatives' (this will eventually lead us to the nested MNL).

The n equations in (8) constitute a system of probabilitistic demand functions of the individual, that is, π_i stands for the probability that a consumer with personal attributes h will choose product i, this being a function of the price and attributes of the product in question, as well as of the prices and attributes of competing brands. It is easy to prove that this is indeed a well-behaved demand system, exhibiting all the properties of conventional - i.e., deterministic - demand functions, and hence the notion of consumer surplus applies to it as well, and can be computed by integration. The task is greatly simplified by assuming away income effects; thus, we specialize the underlying utility function to be additive separable in the group products (those in S) and in the outside good m, rendering a conditional indirect utility function of the form $V_i^c = \alpha(y - p_i) + \phi(Z_i, h) + \epsilon_i$, where α stands for the - constant - marginal utility of income. Substituting in (8),

$$(9) \qquad \pi_i = \frac{e^{\alpha(y - p_i) + \phi(z_i;h)}}{\sum_{j=1}^{n} e^{\alpha(y - p_j) + \phi(z_j;h)}} = \frac{e^{-\alpha p_i + \phi(z_i;h)}}{\Sigma\, e^{-\alpha p_j + \phi(Z_j;h)}} \qquad i = 1, \ldots n$$

Thus, income drops out altogether from the choice probabilities, since it affects equally the utility level of all alternatives, and not their relative desirability (note, however, that y can still appear as a personal attribute in $\phi(\cdot)$). The identity of hicksian and marshallian demand functions in (9) allows us to obtain the surplus function W(S,h) simply by integrating under these demand functions, the integral being path independent. The result is[18]

$$W(S,h) = \ln\left[\sum_{i=1}^{n} \exp(-\alpha p_i + \phi(z_i;h))\right]/\alpha \tag{10}$$

As suggested earlier, the surplus function as defined in (10) is the key element in the computation of welfare gains from innovation, and therefore it is worth mentioning in brief some of its properties (we examine them in detail in a separate paper). First, note that even though the consumer ends up buying one unit of one specific product, the surplus function refers to the whole set S, i.e., it is calculated as if the consumer were to buy fractions π_i of each and every product in the set. Reasons aside, this is a very convenient property for our purposes, for it allows us to trace the welfare effects of changes (innovations) in the entire set of products, regardless of what product each consumer actually purchases.

Second, it is easy to show that (10) can be decomposed as follows

$$\alpha W(S,h) = \Sigma\, \pi_i V_i + [-\Sigma\, \pi_i\, \ell n\, \pi_i] \equiv E(V) + \psi \tag{11}$$

where E(V) stands for the expected value of the deterministic component of the conditional indirect utility function (recall that here $V_i = -\alpha p_i + \phi(z_i,h)$, and ψ for the well-known entropy measure. The interesting aspect of (11) is that it brings to light the existence of a taste for 'horizontal variety', that is, the

---

[18]Actually, the full expression in (10) should include the income y as a constant of integration (recall that utility is linear in income, and that, in dividing by α, we are normalizing $W(\cdot)$ so as to express it in money terms); however, we shall ignore y throughout, since we are only interested in the

type of preferences underlying (10) are such that the welfare of the consumer is enhanced by the proliferation of seemingly identical brands that may differ only in their unobserved (and thus presumably less important) attributes. As we show in Trajtenberg [1983], Ch.II, this taste for variety is embedded in θ, and it stems from - and is formally equivalent to - the assumption of Independence of Irrelevant Alternatives. Thus, it may be of interest to actually decompose the gains ΔW into the two components $\Delta E(V)$ and $\Delta\psi$ and, since the latter stands for the benefits due only to increased variety, one may want, for some purposes, to net them out and use the partial measure $(\Delta W - \Delta\psi)$.

a. Incorporating the hedonic price functions into the MNL model

Throughout the foregoing discussion we have ignored a prominent feature of markets for differentiated products, namely, the fact that prices and attributes usually exhibit a systematic relationship, embedded in the so-called hedonic price function:

$$p_i = p(z_i) + \tilde{p}_i \tag{12}$$

where $p(z_i)$ is the systematic component, and $\tilde{p}_i$ and i.i.d. error term, to be referred to hereafter as 'residual price'. The existence of such a relationship poses a serious multicollinearity problem in the estimation of the choice probabilities defined in (9): since both price and the vector z_i appear there

---

changes ΔW_t. Notice also that $-\partial W/\partial p_i = \pi_i$, and hence (10) is indeed the correct solution.

as explanatory variables, their individual coefficients cannot be estimated with any precision. As we are reminded all to often, multicollinearity is a problem affecting the data entering the model, rather than the model per se, and hence cannot be easily circumvented. In this case, however, the question is not one of mere statistical association between certain variables, but has to do instead with more substantive conceptual issues. Thus, the solution that we shall put forward involves providing the functions V_i with more structure, building upon the role that the hedonic price function plays in the agents' optimization problem.

To recall, hedonic price functions are not just a statistical regularity, but the result of simultaneous and interdependent equilibria occuring in a string of continguous (sub)markets for differentiated products, as shown in Rosen [1974]. From the point of view of the individual consumer deprived of market power, the hedonic price function is to be thought of as an exogeneously given, usually nonlinear budget constraint, i.e., it is the locus of feasible 'consumption bundles' in price-attributes space,[19] and should be incorporated as such in the analysis of consumer behavior. That is indeed done in the continuous case (as in Rosen [1974]), in a rather straightforward manner: assuming a continuous and non-stochastic hedonic function $p = p(z)$, and the same indirect utility as in (5) above, we simply substitute $p(z)$ for p in the V function, rendering the garden-variety maximization problem[20]

---

[19]The consumption bundles are actually $[(y-p_i),z_i]$ rather than $[p_i,z_i]$, but this transformation makes no analytical difference.

[20]The vector of personal attributes is of no consequence for the issue under consideration, and we shall therefore disregard it throughout this discussion.

(13) $$\max_{z} V = \alpha[y - p(z)] + \phi(z)$$

$$\text{FOC:} \quad p'(z) = \phi'(z)/\alpha$$

The rather obvious point made explicit by (13) is that the consumer cannot independently choose both a vector z <u>and</u> a price p, but instead the choice of the former uniquely determines the latter. Consequently, in the regression analysis proposed by Rosen and based upon the FOC in (13) (for the demand equations), only the characteristics z appear as explanatory variables, and therefore the collinearity problem does not arise.

We contend that the essence of the foregoing analysis applies to the discrete case as well; obviously, the hedonic price function is, in that context, neither continuous nor fully deterministic, but it still plays the role of a - stochastic - budget constraint, and should therefore be brought into the analysis as such. Substituting (12) for p_i in V_i, and ignoring y,

$$V_i = -\alpha[p(z_i) + \tilde{p}_i] + \phi(z_i) = \phi(z_i) - \alpha p(z_i) - \alpha\tilde{p}_i$$

or, defining $V^n(z_i) \equiv \phi(z_i) - \alpha p(z_i)$,

(14) $$V_i = V^n(z_i) - \alpha\tilde{p}_i$$

The term $V^n(z_i)$ can be interpreted as the net utility conferred by product i (that is, net of the expected cost of the product), and $\tilde{p}_i$ as an extra charge/ discount resulting from random deviations of actual prices from predicted market equilibria. Having thus stated the problem, the behavior of consumers is now seen to depend upon z_i and $\tilde{p}_i$, rather than upon z_i and p_i, this being an intuitively appealing reformulation: given the existence of a hedonic function, the price variable largely replicates the information already conveyed by z_i; hence, only the component of p_i that is orthogonal to z_i, $\tilde{p}_i$, can affect the choice behavior, and qualifies as a legitimate explanatory variable in a model that attempts to account for such behavior. Thus, (14) offers a conceptually plausible solution to the collinearity problem, provided only that a suitable specification for $V^n(z)$ is found. The following proposition furnishes the required structure: $V^n(z)$ can be closely approximated by the sum of a linear and a quadratic form, provided that it has an interior maximum, i.e., $V^n(z) \cong z'\beta + z'Gz$, where G is a symmetric matrix, if there is a $z^* > 0$ such that

(i) $$\frac{\partial V^n(z^*)}{\partial z_j} = \frac{\partial \phi(z^*)}{\partial z_j} - \frac{\alpha \partial p(z^*)}{\partial z_j} = 0 \qquad j = 1,\ldots,m$$

(ii) the Hessian matrix $[\frac{\partial^2 V^n(z^*)}{\partial z_j \partial z_h}]$ is negative definite.

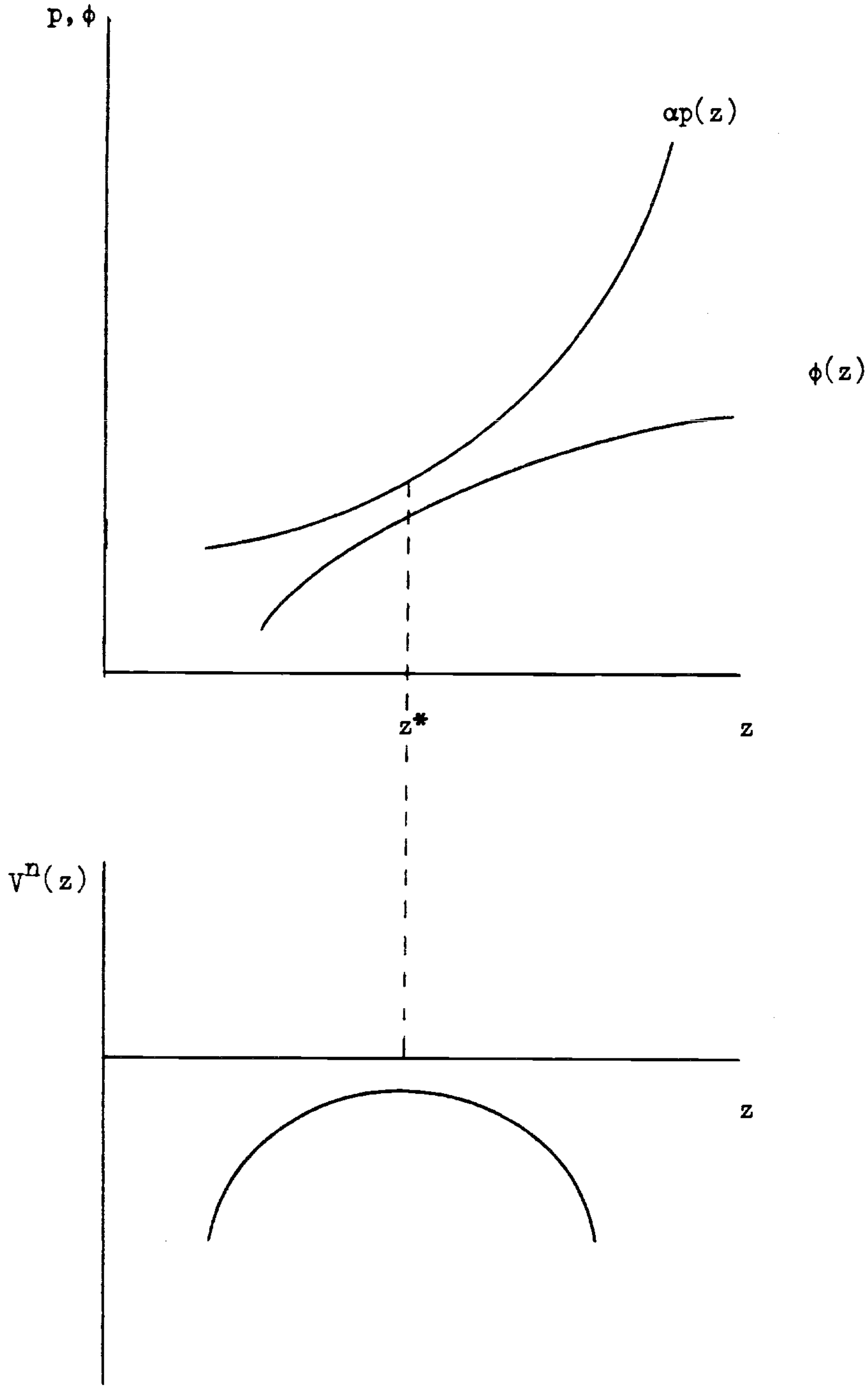
When these conditions hold, the approximation $z'\beta + z'Gz$ obtains in a straightforward manner from a second-order Taylor expansion about z^*. Normally we would expect $\phi(z)$ to be concave (or quasi-concave), and the hedonic function to be convex (as has been found to be the case in many empirical studies), in which case $V^n(z)$ would necessarily meet the required conditions.[21] Figure 1 illustrates such a case for a one-dimensional z. Notice that, since we operate

---

[21]Actually, p(z) may be concave and the stated conditions could still hold: all that is required is, loosely speaking, that p(z) be 'more concave' than $\phi(z)$. If that is not the case then the expected optimum z^* will be in a corner, and hence the proposed approximation will probably be less precise; likewise, the estimated coefficients may turn out to have the wrong signs.

FIGURE 1

Quadratic Approximation to 'Net Utility'



in the context of discrete choice, tangency between $\phi(z)$ and $p(z)$ is not required, and z^* does not necessarily stand for the actual choice; rather, products in the vicinity of z^* have, ceteris paribus, a higher probability of being selected than those further apart (thus, the Taylor expansion is done about the expected, not the actual, optimum).

This specification of the 'net utility' leads to the following model,

$$\text{(15)} \qquad \pi_i = \exp V_i / \Sigma \exp V_j \ ,$$

$$V_i = z_i'\beta + z_i'Gz_i - \alpha\tilde{p}_i ,$$

$$\tilde{p}_i = p_i - p(z_i)$$

implying a two-stage procedure: first, estimate the hedonic price function and compute the residuals $\tilde{p}_i$; second, enter $\tilde{p}_i$ as an independent variable in (15) and estimate the MNL model.[22] Quite clearly, this is not the most general formulation possible: a simultaneous equation framework could be more appropriate, depending upon the presumed behavior of the supply side (e.g., the pricing and product design behavior of firms) and related issues such as whether list versus actual prices are used, the length of the periods analyzed (and hence the likelihood of within-period adjustments), etc. Such a framework may comprise a system of demand and supply equations, replicating in the context of discrete choice what Rosen [1974] does in the continuous case, or just the demand system as in (15), but estimated simultaneously with the hedonic price equation, instead of doing it in two stages. Since our main interest lies

---

[22]It is interesting to note that Cowling and Rayner [1970] also made use of the residuals from hedonic regressions as explanatory variables in a demand equation; even though their model and econometric tools were rather rudimentary, their approach was insightful and constituted a step in the right direction.

elsewhere (i.e., in obtaining the gains from innovation), we shall abstract for now from simultaneity issues, thus avoiding excessive complications. Formally, the working assumption here is that each choice set S_t, and hence each hedonic price function, is determined prior to the beginning of period t and does not change throughout the course of the period; fortunately, this corresponds quite closely to the case analyzed below.

There is, however, a related problem that may present itself when implementing (15): the 'residual price' $\tilde{p}_i$ may be correlated with left-out attributes that do affect the choice behavior (i.e., that do belong in $\phi(z)$), and hence its coefficient might overestimate the marginal utility of income, α. We shall have more to say about this problem, and propose ways of dealing with it, when discussing the results of the case study.

Finally, it is interesting to note that the rather striking omission of hedonic price functions in the existing literature on discrete choice seems to be quite accidental: the case originally and most often analyzed in that literature is the choice of transportation modes; in that context, however, the hedonic price function is virtually non-existent, i.e., there is almost no relationship between fares and the attributes usually considered there (e.g., driving time, route access, etc.). Thus, the practical problem of multicollinearity that motivated the foregoing analysis simply did not arise. Quite clearly though, the transportation case is, in this sense, the exception rather than the rule.

IV. The Case Study: Computed Tomography (CT) Scanners

As mentioned in the Introduction, the methodology outlined above will be applied to a particular innovation, namely, the case of Computed Tomography (CT) scanners (also known as C.A.T. scanners). The following is a brief description of the innovation and its background, its main characteristics (to be used in the econometric analysis), data sources, and selected quantitative indicators.

Although it is generally recognized that medicine as a whole has advanced enormously in recent times, it is perhaps less of a commonplace that one of the areas to have experienced the most progress is <u>diagnostic</u> medicine. Particularly striking has been the pace of innovation during the last decade in imaging technologies, these referring to a vast array of instruments and procedures designed to provide visual information of the interior of the human body, going back to Roentgen's discovery of x-rays in 1895, and the subsequent development of radiography. CT scanners, widely hailed as one of the most remarkable innovations of recent times, came to epitomize this ongoing revolution in diagnostic technologies, and set the stage for subsequent innovations, the last 'wonder' being MRI (Magnetic Resonance Imaging). Public recognition of the significance of CT climaxed in 1979, when the Nobel Prize in Medicine was awarded to the two scientists who pioneered the system.

Research on Computed Tomography began in 1967 at the British electronic company EMI, the first operational prototype was built and installed in 1971 in a London hospital, and the first commercial system, called the EMI Scanner, was installed in the U.S., in June 1973. Aware of the revolutionary nature of the innovation and anticipating a vast market, a number of U.S., European and Japanese companies rushed to enter the new field. In the period 1974-1977 nearly 20 firms stepped in, ranging from the giants in electronics (G.E., Siemens, Hitachi, Philips), to pharmaceutical companies (Pfizer, Searle, Syntex), to small, specialized firms (AS&E, Elscint, etc.). They engaged in fierce product competition, each trying to capture a share of the growing market by offering ever improving performance, thus bringing about a staggering pace of technological advance.

In contrast to the traditionally cautious and often reluctant attitude of the medical profession to the adoption of innovations, the diffusion of CT scanners in the U.S. proceeded at a very fast pace.[23] This happened at a time of mounting concerns regarding the spiralling costs of health care, and hence brought to the forefront the intricate policy issue of how should society allocate resources to that kind of new, very sophisticated but very expensive medical technologies. An intense public debate followed, prompting the government to take a series of regulatory measures, primarily through the implementation of CON (Certificate of Need) requirements. The full impact of these measures was felt in 1978-79, bringing about a sharp downturn in the market for CT, and the exit of many firms, including EMI. This was to be, however, a short-lived occurrence: after a major reshuffling of the industry and the easing of regulatory controls, the market rebounded, settling on a 10-15 percent annual growth rate. The structure of the industry has since stabilized as well, with General Electric as the well-established leader (holding about 50 percent of the U.S. market), and some ten other firms scrambling for the other half.

---

[23]CT scanners were the fifth fastest among a set of twenty innovations in various fields for which diffusion studies have been performed, and by far the fastest among the subset of medical innovations; see Trajtenberg and Yitzhaki [1985].

a. Quality dimensions of CT scanners

Although CT scanners are highly complex systems, it is relatively easy to identify their most important 'performance' - as opposed to 'technical' - characteristics,[24] these being scan time and image quality. In order to gain some understanding of what these attributes stand for, it may be helpful to think of a CT scanner as a photographic camera, that 'takes pictures' of very thin cross-sections (or 'slices') of organs in the body. Since internal organs are subject to involuntary motions, the faster a CT scanner can complete a scan,[25] (i.e., 'take a picture'), the less will be the 'blurring' - or distortion - in the picture caused by those motions (as would be the case in photography). Thus, increasing the scanning speed widens the range of organs that can be successfully visualized (recall that some organs are quite stable whereas others are subject to fast motions): the scan time of the first CT scanner was of five minutes, and hence it could be used only for studies of the brain (a motion-free organ). As scan time dropped to twenty seconds and less, other body organs could be seen and, with speeds of up to one second, present-day systems can render good images of almost any section of the body, except the heart. We shall denote scan time by SPEED, and take it to be the minimum scan time

[24]Performance characteristics are those that affect directly the utility of consumers, whereas technical (or physical) attributes are to be thought of as inputs entering into some transformation function that produces the former 'final' characteristics. The distinction is in many instances not clear cut and, even though performance characteristics are mostly to be preferred in economic analysis, they are often difficult to measure, whereas there is usually plenty of data on technical attributes, and in many cases they can serve as good proxies for the others.

[25]A scan is done by a rotational motion around the patient of a mechanism on which there are mounted an x-ray source on one side, and an array of detectors opposite to it. The x-ray tube emits a narrow beam that goes through the section under examination, and then hits the detectors on the other side; these read and quantify the outcoming energy, and rally the digitized information to a

technically possible in a CT scanner (most systems can be operated at several speeds), measured in seconds.

The 'output' of a CT procedure, that is, the final product to be used in diagnosis, consists of a series of computer-generated pictures displayed on a TV-like screen. Thus, the 'acid test' for the performance of a CT scanner is the 'quality' of the received images, i.e., the accuracy and richness in detail of the diagnostic informational content of the pictures. This is, quite clearly, a broad and ill-defined concept that does not render itself easily to quantitative, objective assessment; there is in fact a voluminous literature (and a great deal of controversy) in Medical Physics on this issue, and the search for evaluation standards still goes on. Since satisfactory, comprehensive measures were nowhere to be found, we designed an econometric procedure to estimate indices of image quality (very much in the spirit of total factor productivity indices) and applied it successfully to experimental data on six CT scanners, provided to us by researchers in radiology (see Trajtenberg [1984]). However, we were unable to obtain the required data for <u>all</u> the CT scanners marketed throughout the period studied (i.e., about 50 different systems), and hence we could not use the proposed index as a measure of image quality in this study. Instead, we had to content ourselves with a measure of <u>spatial resolution</u>, that is, the ability of a system to record detail, or distinguish small objects. More precisely, spatial resolution refers to the

---

computer. This procedure is repeated many times as the mechanism rotates, taking readings at each angle in the rotation. Finally, the data thus gathered are processed by the computer, and the reconstructed picture is displayed on a cathode-ray tube.

size of the smallest object in the received image that can be just visualized in high-contrast regions, i.e,. where there are pronounced differences in density between the small object and its surroundings.[26] That is, to be sure, just one dimension of the imaging performance of a system (although it is usually correlated with the other dimensions) and therefore can be taken only as an imperfect and partial indicator of overall image quality; still, it is the best proxy available for almost all scanners. We denote spatial resolution by RESOL, and measure it in millimeters.

Scan time and image quality are, as said, the two most important performance dimensions, standing well above other potentially relevant attributes (e.g., those that have to do with the mode of operation and options of the systems, patient throughput, data storage capabilities, etc.). Of these we chose to consider, primarily for pragmatic reasons,[27] two additional attributes: reconstruction time and gantry tilt; however, the latter proved to be statistically insignificant, and hence was eventually discarded.

Reconstruction time refers to the time interval between the end of the scan and the display of the image, i.e,. the length of time that it takes the computer to process the massive amount of data generated during the scan, and 'reconstruct' the final picture out of these data. Clearly, faster reconstruction times increase the efficiency of a CT scanner, both by reducing the overall time of a CT examination (thereby increasing the number of patients

---

[26]Thus, spatial resolution is the limiting capability of the system, i.e., it is the ability to visualize small objects in 'the best of conditions'.

[27]The major difficulties that prevented us from considering additional attributes were: (a) lack of data covering a sufficiently large number of scanners; (b) small variance in cross-sections, i.e., some of the features of interest were incorporated in most CT scanners at roughly the same time, and hence their effect within a given period could not be successfully estimated;

per day that can be scanned) and by enhancing the operator's ability to monitor and adjust the system. We denote reconstruction time by RTIME and, as with scan time, define it to be the minimum of the available range, measured in seconds. (For future reference notice that the three characteristics considered are defined so that 'less is better').

b. Data and sources

We have gathered by ourselves, from primary and secondary sources, an all-encompassing data bank on CT scanners, and related technological, economic and institutional issues, covering the nine-year period since the first announcement of the innovation in 1972 and up to the end of 1981 (some of the data extend further, up to 1983). The main data sets thus compiled are:

(i) The technical and performance characteristics, and the prices of all CT scanners developed and marketed up to 1982, by year (this is what we denoted earlier by S_t).

(ii) Detailed sales data, that is, information on each individual installation, including the identity of each hospital (or private clinic) in the U.S. that acquired a CT scanner, the date when the scanner was ordered and installed by each user, and the precise scanner model bought. Repeat purchases (i.e., users buying more than one unit over time, either for replacement or for additions to

---

(c) the qualitative nature of some of the attributes: their inclusion would have required the use of many dummy variables, exceeding the limit number of independent variables allowed. In any case, a preliminary statistical examination showed that many of the excluded attributes are indeed highly correlated with the included ones, and therefore not much is lost by leaving them out.

capacity) as well as upgrades of existing units are also recorded in detail.

(iii) The attributes of hospitals, e.g., size, affiliation, services offered, budget, etc.

(iv) Information about the manufactures of CT scanners, including annual R&D expenditures on CT and patents granted.

In assembling these data sets we had to resort to a wide range of sources, primarily the following:

(a) A questionnaire sent to all CT manufacturers, and follow-up personal contacts with officers from those companies.

(b) Articles in the scientific literature (i.e., in Radiology and Medical Physics), journals related to the medical sector such as Modern Health Care and Diagnostic Imaging, reports from consulting companies (e.g., A.D. Little, Eberstadt, etc.), and a variety of other publications.

(c) Public and government agencies, primarily the Bureau of Radiological Health (BRH) at the Food and Drug Administration (FDA), the Office of Technology Assessment of the U.S. Congress (OTA), the U.S. Patent Office, and the American Hospital Association.

(d) A telephone survey of a few hundred hospitals; personal contacts with faculty and researchers at various medical schools.

All but two of the companies that were still active by 1981 in the CT market answered the questionnaire, at least partially: Elscint, General Electric, Omnimedical, Picker, Siemens, Technicare (of Johnson and Johnson), and Toshiba. The two that did not were CGR (from France) and Philips; the latter had to be excluded from the analysis for lack of reliable information, but not much is lost since the company had sold less than twenty scanners in the U.S. by 1981. Of those that had exited the market, only one (Varian) responded to the questionnarie, whereas the other six (Artronix, AS&A, EMI, Pfizer, Searle and Syntex) could not be reached. We are quite certain that the installation data set, comprising more than 2,000 observations (including upgrades) covers about 98 percent of all CT installations in the U.S. up to July 1981, and that it is very accurate.

c. A first look at the data

We present in Tables 1 and 2 various indicators of the technological and market evolution of CT scanners over time. All figures refer to the U.S. market only; nevertheless, the trends displayed by them are fairly representative of the world market as well (with the possible exception of Japan) since the U.S. accounts for some two-thirds of the total. A 'year' in these tables (and throughout the paper) refers not quite to the calendar year, but to the period from November 1 - of the previous year - to October 31. The reason is that most new scanners are introduced at the annual meetings of the Radiological Society of North America (RSNA), which takes place during the month of November.

In Table 1 we present separate figures for Body and Head scanners: as their names suggest, the former are systems capable of scanning almost any organ of the body, whereas the latter are designed to scan the brain only. As can be clearly seen in these tables, ever since their introduction in 1975 body scanners came to dominate the scene in CT, both in sales and in terms of technological advance. Moreover, the trends displayed by the two types of scanners have been diametrically opposed: whereas head scanners became simpler

Table 1

Characteristics and Prices of CT Scanners - 1973-1982

YEAR	PRICE[a] ($K)		SPEED (sec.)		RESOL (mm.)		RTIME (Sec.)	
	H	B	H	B	H	B	H	B
1973	310	-	300	-	3.1	-	300	-
1974	370	-	300	-	1.7	-	30	-
1975	379	365	285	195.0	1.8	1.6	20	8
1976	374	471	105	63.0	1.7	1.5	60	83
1977	354	573	95	19.0	1.7	1.3	54	38
1978	167	620	96	7.1	1.6	1.2	29	30
1979	154	667	150	6.6	1.5	1.1	19	29
1980	154	739	115	5.5	1.5	1.0	27	31
1981	150	827	115	4.9	1.5	0.8	27	32
1982	150	850	115	2.6	1.5	0.7	27	27

H: Head scanners; B: Body Scanners

[a] "PRICE" is the <u>weighed</u> average of the prices of all scanners in the market, using annual sales as weights. The figures for the characteristics are simple average.

Table 2

CT Scanners -Selected Indicators

YEAR	Unit Sales	Sales in $M	B-Scanners: % of Sales	New Adopters: % of Sales	No. of Firms	No. of Models	Herfindahl Index
1973	16	5	0	100.0	1	1	1.00
1974	74	27	0	98.7	1	1	1.00
1975	221	82	44.8	97.7	4	5	.43
1976	374	167	76.2	84.8	9	14	.38
1977	385	208	84.9	84.1	12	23	.23
1978	248	122	72.4	84.4	10	23	.28
1979	273	141	70.2	76.0	9	22	.26
1980	270	169	80.4	65.3	8	17	.32
1981	392	302	91.6	50.0	8	14	.31
1982	428	344	93.5	n.a	8	16	.30

and cheaper over time, body scanners exhibited a tremendous pace of technological advance and a corresponding steep rise in prices. Since 1981 the market segment occupied by head scanners has shrunk to negligible proportions, and it is very likely to remain at that.

According to virtually all indicators, competition in the market for CT scanners reached its peak in 1977, e.g., fourteen new scanners were introduced that year, the number of firms reached 12, and the Herfindahl index hit an all-time low of .23. Sales also attained a local maximum then, being surpassed only in 1981, when replacement sales became substantial, compensating for the natural decline in the number of new adopters (due to the leveling-off of the diffusion porcess). To repeat, our mean interest lies in the evolution of prices and characteristics over time, as reflected in the summary figures of Table 1.

V. The Econometric Analysis

The cornerstone of the econometric analysis leading to the computation of gains from innovation consists, as already noted, of the estimation of the MNL model of choice of CT scanners. We shall make use for that purpose of the two-stage procedure developed in III.a above, and estimate the model separately for every year in the period 1975-1981.[28] In order to keep computational complexities down at a manageable level, we shall ignore the vector h of individual attributes and their effect on choices (except in the estimation of the MNL for 1975). In other words, we shall estimate the

---

[28]In 1973 and 1974 there was in fact only one CT scanner in the market, and therefore there is nothing to estimate for those years. As to the last year, we have detailed data only up to June 1981, accounting for less than half the total sales for the year. Unless stated otherwise, all figures for 1981 will thus

choice probabilities only as functions of the attributes and prices of scanners - these being common to all buyers[29] - and omit the interaction between these variables and the individual h's. Otherwise the function $W(S_t)$ would have to be computed for each buyer in every period, and then integrated over h; given that there are about 300 observations on average per annum, such procedure would increase the computational burden by an order of magnitude. Thus, the results from the estimation of the MNL model refer to the 'average' or 'representative' user of CT scanners, this including both hospitals and private clinics. Likewise, we do not consider here firm-related variables that have a bearing on choices, but that do not have a clear interpretation in terms of welfare, such as 'reputation' (measured for example by cumulative sales) or advertising outlays.[30]

As pointed out earlier there are two different types of CT scanners, namely, head and body systems. An important issue in formulating the MNL model is how to structure the choice set in view of that distinction, i.e. whether head and body scanners comprise a single choice set, or rather two separate branches of the decision tree and, if the latter proves to be the case, whether or not the two scanner types actually belong at all to a common decision tree. These questions hinge on the pattern of substitution between scanners, and hence on the compatibility of the various choice

---

refer to the first half of the year only.

[29]Except that in some years the choice set changed during the year (because of entries), and therefore buyers faced different effective choice sets according to the exact date of purchase; more on this below.

[30]We did estimate models including these variables (as well as individual attributes) and, not surprisingly, the fit improves. However, we are not trying here to account as much as possible for observed behavior,but to do so in a manner consistent with welfare analysis. Needless to say, the detailed analysis

structures with the assumption of Independence of Irrelevant Alternatives (IIA) that underlies the MNL. The search for the appropriate specification will proceed as follows: first, we estimate the MNL once for the entire set of head and body scanners, and once for a restricted set of body scanners only, and conduct an appropriate specification test. As the null hypothesis that IIA holds for the larger set (i.e., that head and body scanners form a single choice set) is rejected, we specify next a nested structure with two branches - one for head and one for body scanners - and estimate it sequentially. The results indicate that the elasticities of substitution between the two types of scanners are nil, and therefore that head and body scanners constitute in fact unrelated sets from the viewpoint of the choice process; this in turn will affect the way by which the gains from innovation are to be computed.

a. Estimating the hedonic price functions

We now turn to the first stage of the procedure to be followed, namely, the estimation of the hedonic price functions, and the subsequent computation of residual prices. As is commonly the case in hedonic price studies, we do not have strong priors regarding the functional form of the hedonic equation and, except for plausible arguments favoring convexity, there are virtually no theoretical guidelines to follow (see Griliches [1971]). Thus, the matter is to be decided by comparing the fit of

---

of choice of CT scanners is interesting in itself from a positive point of view and will be discussed in a separate paper.

alternative specifications (the well-known limitations of such criterion not withstanding), and subsidiary considerations of a pragmatic nature. We experimented with three specifications (log stands for natural logarithms):

(i) Double-log $\log p_i = \alpha + \Sigma\beta_j \log z_{ij} + \epsilon_i$

(ii) Semi-log: $\log p_i = \alpha + \Sigma\beta_j z_{ij} + \epsilon_i$

(iii) Linear-log: $p_i = \alpha + \Sigma\beta_j \log z_{ij} + \epsilon_i$

The double-log and the semi-log are the two forms most commonly used in the literature, and constituted therefore natural candidates. The third form has never been used before - as far as we know - in a hedonic price equation, presumably for the converse of what made it plausible in our case: if the characteristics are defined so that $\delta U/\delta z > 0$, and hence $\delta p/\delta z > 0$ (i.e., so that the more there is of them the better is the products, as is usually the case) then the linear-log is necessarily concave, which is deemed to be rather an implausible feature of these functions. If, on the other hand, $\delta U/\delta z < 0$ as in the case at hand,[31] then (iii) is necessarily convex, as are the other two forms.[32]

The three equations were estimated both for the joint set of head and body scanners (including a dummy variable for head scanners), and for body scanners only, for every year in the period 1976-1981 (1975 will receive

---

[31]Recall that the shorter is scan time (SPEED) and reconstruction time (RTIME), the better is the scanner, and likewise, the smaller the value of RESOL - in mm. - the better the imaging capabilities of the system.

[32]To be more precise, the semi-log is always convex, whereas the double-log depends upon the values of β_j: If $\beta_j < 0$ (i.e. if $\partial U/\partial z_j < 0$) or if $\beta_j > 1$ then it is convex, whereas if $0 < \beta_j < 1$ it is concave.

separate treatment). We compared the fit of the alternative specifications using the Box-Cox transformation, and the linear-log emerged as the clear winner: it ranked first (i.e., minimum corrected mean-square error) in half the cases, second in one-third, and last only in the remaining one-sixth of the twelve cases considered.[33] This was quite fortunate, since the linear-log happens to be a highly convenient specification for our purposes: first, it will allow us to further simplify the form of the 'net utility' in the MNL model, and second, its residuals are defined in the same units as price (since the dependent variable is indeed price, rather than log price as in the other two forms) and can therefore by incorporated directly in the MNL as an independent variable.

We present in Table 3 the estimated hedonic price equations. Notice that the R^2 hovers about .85, and that the estimated coefficients are, with few exceptions, statistically significant. Thus, the characteristics included are indeed of relevance in the market for CT scanners, accounting for most of the observed price variability. Consequently, the inclusion of both price and attributes in the MNL would have resulted - as suspected - in a serious multicollinearity problem, and therefore the use of residual price instead is amply justified.[34]

---

[33]The 12 cases are made up of two sets of six equations each (one equation for each year, 1976-1981): the first set for head and body scanners, and the second for body scanners only. The second best specification was the double-log: it ranked first in 1/4 of the cases, second in 7/12, and third in 1/6. For a detailed presentation of the estimated equations and of the comparisons see Trajtenberg [1983], Tables V.3-V.5.

[34]Since our intention is just to generate residual prices for their subsequent use in the MNL, we do not pursue here a systematic analysis of the hedonic price regressions, even though the results are interesting in themselves (e.g., the pattern followed over time by the estimated coefficients, the reasons for the observed changes in the goodness-of-fit, etc.)

TABLE 3: Hedonic Price Regressions, Linear-Log Form by Year
(a) Head and Body CT Scanners

	1976	1977	1978	1979	1980	1981
Intercept	657.21 (5.3)	931.46 (20.1)	880.05 (13.8)	904.96 (18.9)	861.34 (9.6)	1,062.7 (7.2)
Head Dummy	-87.84 (-1.9)	-75.21 (-2.3)	-121.15 (-2.1)	-123.78 (-1.8)	31.48 (0.3)	209.67 (1.0)
$\ln$ SPEED	-37.33 (-1.7)	-46.82 (-4.2)	-66.86 (-3.7)	-82.83 (-4.6)	-110.93 (-3.8)	-193.68 (-3.2)
$\ln$ RESOL	-67.74 (-0.6)	-226.54 (-3.3)	-147.16 (-1.5)	-189.19 (-3.0)	-380.15 (-4.6)	-335.0 (-2.5)
$\ln$ RTIME	-4.65 (-0.2)	-33.33 (-3.7)	-42.01 (-2.7)	-39.69 (-3.0)	-18.35 (-0.8)	(-2.5) (-0.7)
R^2	.653	.895	.855	.937	.912	.916
d.f.	9	16	19	15	13	9

(b) Body CT Scanners Only

	1976	1977	1978	1979	1980	1981
Intercept	790.08 (20.7)	808.37 (20.6)	833.96 (13.8)	914.17 (19.9)	873.76 (8.5)	1,070.9 (6.2)
$\ln$ SPEED	-75.22 (-11.0)	-62.19 (-5.7)	-56.84 (-3.2)	-82.14 (-4.8)	-109.94 (-3.5)	-190.87 (-2.8)
$\ln$ RESOL	-12.58 (-0.4)	-161.79 (-2.7)	-150.02 (-1.8)	-198.65 (-3.4)	-382.30 (-4.2)	-337.77 (-2.2)
$\ln$ RTIME	-14.62 (-2.5)	-19.93 (-1.9)	-32.25 (-2.0)	-43.04 (-3.6)	-22.70 (-0.8)	-25.44 (-0.6)
R^2	.981	.878	.656	.847	.811	.783
d.f.	4	11	12	12	11	7

t-values in parenthesis

b. The MNL: Preliminary remarks

We argued in III above that the net utility function $V^n(z) = \phi(z) - \alpha p(z)$ can be closely approximated by $V^n(z) \cong z'\beta + z'Gz$, relying for that purpose just on general properties of the underlying utility and hedonic functions. However, if we further assume that the utility branch as the form $\phi(z_i) = \sum_{j=1}^{m} f^j(z_{ij})$, where $f^j(\cdot)$ is a quasi-concave function of the j-th attribute,[35] and substitute the linear-log hedonic price equation for p(z) in $V^n(z)$, then the approximation to the net utility function simplifies to the simple - rather than the 'generalized' quadratic form,

$$(16)\quad V^n(z_i) = \sum_{j=1}^{m}[f^j(z_{ij})-\beta_j \log(z_{ij})] \cong \sum_{j=1}^{m} (a_j z_{ij} - b_j z^2_{ij})$$

Simply put, if both the utility function and the hedonic price equations are additive separable in - some transformations of - the z_{ij}'s, then the interaction terms drop out of the quadratic form, leaving only the linear and squared terms. Quite clearly, the only reason to prefer (16) over the original quadratic approximation is that we thus gain m(m-1)/2 degrees of freedom in the estimation of the MNL; this may turn out to be a significant advantage in the implementation of the model, particularly so if the individual attributes are not included (as in our case).[36] Thus, we shall use the specification laid out in (16) for the estimation of the MNL model.

A note on the problem of changes in the choice set within periods. As

---

[35]In words, the assumption is that utility is additive separable in the characteristics, these being defined either in their original units, or in some quasi-concave transformation of them. This is the working assumption in virtually all aplications of the MNL model.

[36]If the independent variables included in the MNL (such as z) are 'generic', that is, they vary across models (or 'alternatives') but not across individuals, then the maximum number of variables allowed is n-1, n being the number of alternatives in the choice set. This constraint does not hold when

mentioned in IV above, most new models of CT scanners are introduced at the annual meetings of the RSNA that takes place during the month of November, setting the stage for what the market for imaging equipment is going to be like in the course of the year; thus, for the purpose of the empirical analysis we defined a year to be the period November 1 - October 31. Still, the choice sets did not remain constant within these 'radiological years': some entries occurred after November, and the timing of exits was, of course, unrelated to the RSNA meetings.[37] In principle, the MNL allows the choice set to vary across individuals in the sample and, since we have the precise date (month/year) at which each buyer <u>ordered</u> a scanner (denote it by t_ℓ), we could specify a choice set for each individual ℓ, namely, the collection of scanners offered in the market at t_ℓ, net of entries and exits. However, the software used here to estimate the MNL has an option that allows us to easily incorporate expansions in the choice set, but not contractions;[38] moreover, in many cases we were not certain of the precise data of withdrawal of scanners from the market, whereas announcement dates

---

the variables do vary across individuals, or what is exactly equivalent, when the marginal utilities are individual-specific, these 'individual effects' being functions of personal attributes h. Thus, the saving of m(m-1)/2 variables can be very important, if not decisive, in the former case. Still, it is important to emphazsize that what allowed us to use (16) is the finding that the linear-log form provided the best fit for the hedonic price functions: had it been the log-linear or the double log instead, the approximation in (16) would not be a legitimate one.

[37]As a matter of semantic convenience, by 'entries' and 'exits' we refer here to the introduction of new CT scanners to the market and the withdrawal of existing ones, and not necessarily to the stepping in or out of firms, although at times the two sorts of events do coincide.

[38]We made use of the package called MLOGIT, originally written by Charles Manski, later expanded by Bronwyn Hall, and interfaced with SAS by Danny Smith at the NBER in Cambridge. For details on the procedure used to take care of entries, see Trajtenberg [1983], ch.V.

are for the most part well known. Thus, within-period changes in the choice sets due to entries have been duly taken into account, but that is not the case for exits, i.e. if a scanner was offered with certainty at the beginning of the year, we assumed that it remained on the market thorughout the year even though it may have been withdrawn earlier on. It is conceivable that this assymetry may have had a slight impact on the results, but it is virtually impossible to be more precise.

Finally, a comment on goodness-of-fit measures. As opposed to the widespread reliance on the R^2 in regression analysis, there is no universally accepted scalar criterion of fit for discrete choice models. We shall report here two measures: The popular ρ^2 (or 'McFadden's R^2'), and the correlation between predicted (π_i^*) and actual probabilities (π_i), $Cor(\pi_i^*, \pi_i)$. The first is defined as $\rho^2 = 1 - [L(\beta^*)/L(0)]$, where $L(\beta^*)$ is the maximized value of the log-likelihood function, and L(0) the value of the ML function when all the coefficients are assumed to be zero.[39] Even though $0 \leqslant \rho^2 \leqslant 1$, values of the order of .20 are considered to represent good fits (so at least Hensher and Johnson [1981] claim), but more experience with these models is needed in order to establish well-grounded benchmarks. The second measure $Cor(\pi_i^*, \pi_i)$, although less precise and much rarely reported than ρ^2, can still be fairly informative of the performance of the model in the aggregate, that is, in predicting market shares. The correlations to be reported below have been adjusted so as to take into account within-period changes (entries) in the choice sets.

---

[39]That is, $L(0) = \sum_{j=1}^{m} n_j \ell n(1/m)$. where m is the number of alternatives in the choice set, and n_j the number of individuals choosing alternative j. In other words, if we knew nothing about the determinants of choice we would assign equal probabilities to each alternative (i.e. $\pi_j = 1/m$), and hence L(0)

c. A specification test

The most distinctive feature of the MNL, accounting for both its advantages (computational and otherwise) and its limitations, is the assumption of Independence of Irrelevant Alternatives (IIA). One way of stating it is that, for any pair of choice alternatives, i and j in S, the probability ratio π_i/π_j is invariant with respect to the inclusion or exclusion of other alternatives. The important point is that IIA precludes the possibility of having a flexible pattern of substitution between alternatives of, to put it differently, it rules out differential proximity between alternatives in the space of unobserved quality dimensions. For example, one would expect the cross price elasticities within the sub-set of head (or body) scanners to be higher than those between scanners of different types, but IIA constrains them to be equal (i.e. $\partial \ell n\ \pi_i/\partial \ell n p_j$ is independent of i). The question is how to test for the IIA property and, if necessary, how to bypass it (at least partially), that is, how to introduce more flexibility to the model.

Houseman and McFadden [1981] put forward a specification test based upon the comparison between the estimates obtained when applying the MNL to the full choice set, β_f, and those arrived at when estimating a restricted set, β_r: if IIA holds, both should be statistically equivalent. Formally, the statistic

---

represents a sort of natural baseline built on the assumption of complete ignorance.

(17) $$S = (\beta_r-\beta_f)'(Cov_r-Cov_f)^{-1}(\beta_r-\beta_f) \sim \chi^2_{(m)}$$

(m is the rank of the cov matrix) can be used to test the null hypothesis $(\beta_r-\beta_f) = 0$. The problem is that this is in fact an overall test that may fail because of misspecifications other than IIA and, as McFadden [1982] points out, it is not powerful unless deviations from the MNL structure are substantial. Still, it is the only specification test that does not require a departure from MNL: more powerful tests call for higher-level models in which IIA obtains as a special case, thus allowing for nested hypothesis testing. Such is the case, for example, with Houseman and Wise [1978] 'Covariance Probit'; however, and as is typically the case with discrete choice models other than the MNL, the covariance probit can handle only a handful of alternatives, making it impractical in our case.

In order to implement the Houseman-McFadden test, we estimate the MNL for the full set of head and body CT scanners, and for the restricted set of body scanners only. In view of our strong priors (favoring the alternative hypotheses), and in order to avoid excessive computations, we do that only for two years, 1977 and 1981.[40] In both cases the estimated coefficients of the restricted and unrestricted models differ substantially, and indeed, the null hypothesis is rejected: the actual χ^2 value at the .99 significance level is 18.5, whereas the values of the statistic in (17) are S(1977) = 21.83 and S(1981) = 27.77. Thus, we conclude that the data do not support the assumption of IIA for the full choice set, i.e. the choice behavior of buyers does not allow for the symmetric treatment of head and body scanners.

---

[40]We chose those years according to two criteria: first, that they be far apart so that the choice sets are substantially different, and hence the results can be presumed to hold also in the intervening periods, when the sets do

d. The Nested MNL

In light of the above results, we formulate next a nested MNL (NMNL) having two branches, one for the head (H) and the other for body (B) scanners.[41] The probablity of choosing the ith scanner of type h, h = H, S, can then be written as

(18) $\pi(i,h) = \pi(i|h)\pi(h)$

(19) $\pi(i,h) = e^{V_i}/\sum_{j=1}^{n_h} e^{V_j}$

(20) $\pi(h) = e^{\lambda W_h}/[e^{\lambda W_H} + e^{\lambda W_B}]$

where n_h is the number of brands in cluster h, and W_h is the 'inclusive value' of that cluster, i.e.

(21) $W_h = \ln[\sum_{j=1}^{n_h} \exp(V_j)]$, h = H,B.

The key parameter in this model is λ, which is to be interpreted as a measure of substitutability - or 'proximity' - of alternatives across clusters.[42] Thus, and taking the limit cases (λ has to lie in the unit interval): when λ = 1 the IIA property holds for the entire choice set, and therefore cross-elasticities do not depend upon the location of alternatives. Consequently, the

---

overlap; second, that the ratio of the number of head to body scanners be approximately the same, so that the results will not depend upon the relative number of alternatives deleted when going from the full to the resticted model.

[41]Needless to say, the structure of the choice set can be made more complex, with several stages - or levels - of decision (rather than just two as in here), and a finer division into more than two clusters at the lowest stage.

[42]The parameter λ has been incorrectly interpreted in the literature as a measure of the similarity (or 'independence') of alternatives within clusters: see, for example, McFadden [1982], p. 49. In Trajtenberg [1983] we formally prove the role of λ in determining the pattern of inter- and intra- cluster

grouping of alternatives into clusters is altogether inconsequential, and the simple MNL applies to the whole set. On the other hand, $\lambda = 0$ means that the cross-elasticities between alternatives belonging to different clusters is zero, and therefore the decision three should be 'sliced down the middle, i.e. each cluster should be regarded as a separate analytical unit - or'market' - from the point of view of demand behavior. Closely related, the value of λ determines also the form of the $W(\cdot)$ function: if $0 < \lambda \leqslant 1$, then the form of the surplus function is

$$(22) \qquad W = \ln[\textstyle\sum_h(\sum_i \exp V_{i,h})^{\lambda}]$$

whereas if $\lambda = 0$ then

$$(23) \qquad W = \textstyle\sum_h W_h \pi(h) \ , \ W_h = \ln(\sum_i \exp V_{i,h})$$

where the probabilities $\pi(h)$ no longer depend upon the attributes and prices of the alternatives as captured by the inclusive values in (20), but are instead exogeneous to the model,[43], and can therefore be taken to be simply the observed frequencies in the population.

We proceed now to estimate the NMNL model sequentially, that is, first we estimate (19) for head and body scanners separately, then use the estimated coefficients to compute the inclusive values in (21), and lastly we incorporate them in (20) in order to estimate λ;[44] we do that for every year from 1976 to

---

substitution, and explore some of its implications.

[43]In a more general model the $\pi(h)$'s could be made endogeneous, i.e. they could be estimated as functions of variables pertaining to the different clusters as such (other than their inclusive values) and of individual attributes relevant to the choice of clusters.

[44]This procedure is not fully efficient, and hence the second-stage standard

1981 - 1975 will receive separate treatment. The small number of different brands of head scanners offered in the market every year imposed several restrictions on the first-stage estimates of the head branch (recall that the maximum number of 'generic' independent variables that can be included in the MNL is n-1, n being the number of alternatives in the choice set). First, only four years (1976-1979) could be estimated: in 1980 and 1981 the choice set contained too few scanners (only three). Second we had to specify V_i as linear in the z's (with price), rather than quadratic with residual price.[45] Third, only three variables could be included in 1979, and hence RTIME (the least important characteristic) was omitted.

Table 4 presents the first-stage MNL estimates for head scanners, and Table 5 those for body scanners. We shall not analyze here these results in any detail, but just note the following: first, the model fits fairly well - particularly so in terms of COR(π_i^*, π_i) - even though we have used only z variables, ommitting individual and firm-specific attributes. Second, the estimated coefficients change substantially from year to year, suggesting that 'preferences'[46] for attributes have evolved over time, as did the CT technology

---

errors are not entirely reliable. An alternative is to estimate the whole system using the full information maximum likelihood method, but this is difficult to implement in the case of the NMNL, and the experience with the software available at the time for that purpose was not very satisfactory.

[45]Simple correlations between price and characteristics of head scanners were found to be systematically lower than those for body scanners, suggesting that the multicollinearity problems (that motivated the use of the quadratic) might be less severe for head scanners; if so, the linear form may in fact be the appropriate specification for head scanners.

[46]Since the underlying utility function is not linear, 'preferences' - or 'tastes' - are not uniquely defined in terms of the estimated coefficients, but depend upon the level of the attributes. We have chosen to measure preferences for an attribute as the derivative of the choice probabilities with respect to that attribute, averaged over the choice set. Thus, whenever we talk of

Table 4

MNL Estimates for Head CT Scanners - Linear Form

	1976		1977		1978		1979	
PRICE	-0.748	(-1.5)	-0.709	(-1.9)	-0.893	(-5.5)	-0.818	(-2.4)
SPEED	0.018	(0.2)	-0.238	(-1.4)	0.303	(1.9)	-0.318	(-0.9)
RESOL	-4.706	(-3.6)	-5.565	(-2.9)	-7.756	(-4.8)	-10.307	(-4.7)
RTIME	-0.619	(-2.2)	-0.366	(-1.2)	1.119	(2.7)		
$\rho^2 = 1 - \frac{L(\beta^*)}{L(\beta^0)}$	.131		.116		.42		.455	
Corr (π^*, π)	.999 (.0001)		.910 (.012)		.993 (.0001)		.998 (.0016)	
# Scanners	6		6		8		4	
# Obs.	89		56		69		80	

Asymptotic t-values in parenthesis

Table 5

MNL Estimates for Body CT Scanners - Quadratic form with Residual Price

	1976	1977	1978	1979	1980	1981
RFPRICE	11.252 (6.4)	0.993 (4.8)	1.020 (4.8)	0.485 (1.8)	0.695 (2.4)	-0.277 (-2.5)
SPEED	-2.292 (-7.3)	2.138 (2.8)	4.624 (1.0)	-8.669 (-1.5)	11.347 (2.0)	-7.504 (-0.5)
$SPEED^2$	0.236 (4.0)	-1.264 (-3.4)	-8.283 (-0.6)	31.292 (1.9)	-34.838 (-1.6)	74.161 (1.4)
RESOL	69.107 (7.3)	9.113 (2.4)	-34.126 (-6.3)	-15.283 (-5.0)	-18.129 (-3.6)	32.877 (3.9)
$RESOL^2$	-23.360 (-7.6)	-2.533 (-1.5)	15.096 (5.8)	6.291 (3.8)	7.738 (2.7)	-24.028 (-4.2)
RTIME	-3.931 (-5.3)	5.082 (7.0)	2.385 (2.0)	3.288 (3.3)	3.161 (2.8)	-2.591 (-2.8)
$RTIME^2$	1.054 (4.5)	-2.370 (-6.7)	-1.511 (-2.0)	-1.401 (-2.1)	-2.093 (-2.2)	5.560 (3.9)
$\rho^2 = 1 - \frac{L(\beta^*)}{L(\beta^0)}$	.29	.12	.16	.16	.20	.14
$Corr(\pi^*, \pi)$	.999 (.0001)	.877 (.0001)	.900 (.0001)	.870 (.0001)	.722 (.0024)	.547 (.082)
# Scanners	8	15	16	16	15	11
# Observations	285	324	164	177	193	153

Asymptotic t-values in parenthesis

itself. In fact, we have shown elsewhere that a 'dual-inducement' process has been at work (see Trajtenberg [1983], VI.2 and VIII.3-5): relative preferences for attributes in one year influence the direction of technological change (in attributes' space) in the following year, and conversely, as the set of products offered in one period incorporates innovations that emphasize some attributes more than others, preferences towards the enhanced characteristics will weaken next period relative to other attributes. In other words, strong preferences for a given attribute today will induce a relatively large improvement in that attribute tomorrow, which in turn will bring about a relative drop in its marginal desirability afterwards. As we shall see below, this inducement mechanism will be of relevance in choosing the 'base preferences' to be used in computing the yearly gains from innovation. Lastly, it is worth noting that, contrary to what is normally to be expected, the coefficients of RPRICE for body scanners are positive, except for 1981. On the other hand, those for head scanners are negative and fairly stable; we shall elaborate on these two findings later on.

Now to the second stage of the NMNL: Table 6 presents the computed inclusive values and the estimates of λ , for the four years in which the first-stage MNL for head scanners could be estimated. The key result is that, except for 1978, the estimates of λ are very small, implying that the cross elasticities between head and body scanners are nil, and hence that the two types of scanners do not belong to a common decision tree. Lacking the

---

'relative preferences', we refer to the relative magnitudes of these weighted derivatives.

TABLE 6

Nested MNL: Second Stage

(a) Inclusive Values of Body and Head Scanners

	Body Scanners		Head Scanners	
Year	Σe^{V_i}	W_B	Σe^{V_i}	W_H
76	5.22E+20	47.704	1.114E-4	-9.102
77	161,040	11.989	3.012E-5	-10.410
78	5.42E-7	-14.428	3.231E-5	-10.340
79	0.00885	- 4.728	1.366E-7	-15.806

$W = \ln (\Sigma \exp V_i)$

(b) Second-stage estimates

Year	Coef. of W (λ)	Standard Errors	ρ^2
76	0.0205	0.00214	.21
77	0.0784	0.00646	.40
78	-0.2120	0.03510	.12
79	0.0706	0.01211	.10

corrected standard errors[47] we could not test formally the hypothesis that $\lambda = 0$, using either a Wald or a Lagrangian test. However, the stated conclusion is not contingent upon the acceptance of this hypothesis: even if the λ's prove to be statistically different from zero, their small magnitude make the 'between' cross-elasticities negligible, and that is all that is required. As to 1978, the negative value of λ is symptomatic of a local failure of the conditions underlying generalized extreme value models in general, and the NMNL in particular. Although it is difficult to draw any firm conclusion from such a finding, we take it as further evidence that head and body scanners are not to be regarded as substitutes (if anything, $\lambda < 0$ would suggest complementarity, since the cross-elasticities turn out to be positive).

The inference arrived at with the NMNL is no surprise, and only corroborates our priors: the evidence regarding the evolution over time of the CT technology, relative prices and capabilities, and patterns fo acquisition and use, clearly indicates that head and body scanners rapidly diverged from each other, forming two highly segmented sub-markets (see Trajtenberg [1983], V.3, for an extensive and detailed discussion of the issue). At a more general level, we want to stress the fact - largely overlooked in the literature - that the procedure followed above, centered around the estimation of λ, can help solve the all-pervasive problem of drawing market boundaries in empirical micro studies.

We now turn to 1975, the first year in which there was a choice of scanners in the market, and hence the first that could be estimated, albeit in a somewhat special way. Even though announcements of new scanners were made in late 1973

---

[47]These could be obtained by further iterations using the Berndt-Hausman-Hall [1974] procedure; however, the necessary software was not readily available at the time, and hence we were unable to do so.

and throughout 1974, in fact the only scanner sold during that period was the EMI head scanner. In 1975 the first two body scanners were successfully introduced to the market, as well as an additional head scanner. It is clear that at that early stage head and body scanners were indeed close substitutes (they were still very similar in terms of characteristics and prices), and hence the effective choice set comprised both types, totalling four alternatives. That is a very small number for estimation purposes, and moreover, since preferences of buyers seem to have been ill-defined back then due to the newness of the technology, we had a great deal of trouble estimating the MNL for that year. After experimenting with several specifications, the model converged and rendered sensible results with a 'hybrid' specification, shown in Table 7: Notice that, since there was room for only three generic variables, we omitted the quadratic terms and RTIME, and interacted resolution with 'RAD', an attribute of hospitals.[48] All in all the model performs quite well and, in view of the relatively narrow range of values of the included variables, the results are probably not very different from those that would have emerged had the quadratic terms been included as well.

VI Computing the Gains from Innovation

The MNL estimates obtained in V above provide us with the parameters of the utility functions needed to compute the social gains from innovation in CT scanners. To recall, these gains are defined as $\Delta W_t = W(S_t) - W(S_{t-1})$, where

$$(24) \quad W(S_t) = \ln [\sum \exp V_{it}]/\alpha$$

---

[48]RAD is defined as (R1 + R2 + R3)/3, where the R_i's are dummy variables for the availability of X-ray, cobalt and radium therapy respectively. The presumption is that hospitals with a wider range of radiation therapy services will tend to value more resolution in CT scanners. Since RAD is defined only

Table 7

MNL Estimates for 1975, Head and Body Scanners

(Hospitals Only)

RPRICE	10.183	(4.3)
HEAD × PRICE	-2.875	(-6.0)
SPEED	-0.165	(-0.8)
RESOL × RAD	-6.338	(-4.8)

(RESOL at mean RAD -4.183)

ρ^2	.24
No. of scanners	4
No. of observations	181

Thus, all we need is to compute the functions V_{it} using the estimated MNL coefficients and the observed characteristics and prices of scanners in adjacent years, aggregate them as in (24), and take differences. There are, however, two important issues to be considered beforehand: first, the fact that the coefficients of RPRICE for body scanners were found to be positive (except for 1981), and second, the fact that virtually all the estimated coefficients change significantly from year to year, and hence ΔW_t is not uniquely defined.

a. Upward-sloping demand curves, signaling, and welfare analysis

Contrary to textbook dogma, positive price coefficients - and hence upward-sloping demand functions - are by no means an aberration: various types of fairly prevalent price, quantity and quality interrelationships may easily bring them about, particularly when some of the variables are imperfectly observed, thus making room for signaling effects. However, positive price coefficients do pose a serious problem when it comes to welfare analysis, since they can no longer be regarded as estimates of the marginal utility of income (MUI), which plays the usual role of conversion factor - from utility to money terms - in (24) (notice that a positive price coefficient means, on face value, a negative MUI).[49] Likewise, it is no longer clear whether the term $\tilde{\alpha}p_i$ should be included as such in V_{it} when computing (24): that will depend upon the kind of phenomena that give rise to a positive price coefficient.

---

for hospitals, we excluded private clinics in the estimation of 1975. To recall, interacting a generic variable with an individual attribute makes the former vary across individuals, thus bypassing the limitation on the number of variables imposed by the number of alternatives in the set.

[49]More generally, one can no longer obtain a measure of consumer suplus by integrating under the observed, upward-sloping demand curve.

In order to address these issues, consider first the framework put forward by Spence [1973] to analyze price, quality and quantity interdependencies: demand is modeled as a function of price, and a broadly defined 'quality' - or 'desirability' - perameter q, which in turn depends upon price and/or quantity, i.e.,

(25) $x = x(p,q)$

(26) $q = Q(p,x)$

In equilibrium,

(27) $x = x(p, Q(p,x))$

which defines the observed - or 'actual' - demand function, as opposed to the 'virtual' demand function defined in (28) for a given q. Note that whereas (25) behaves as a regular demand function in that $dx/dp = x_p<0$, the price response of (27) is far more complex: totally differentiating (27) and rearranging terms,

(28) $$\frac{dx}{dp} = \frac{x_p + x_q Q_p}{1 - x_q Q_x}$$

Thus, the slope of the observed demand functions depends not only upon the 'pure' price effect x_p, but also on the feedback effects of price and quantity via their impact on perceived quality. Of the many possible cases[50], consider the one where $Q_x=0$ and $Q_p>0$, i.e. where Q(p) stands for what amounts to an

---

[50]For example, $Q_p>0$ corresponds to the 'bandwagon effect', $Q_x<0$ to congestion, etc.

inverse hedonic price function. From (28) it is clear that in such cases the slope of the observed demand function will always be steeper than that of the virtual demand and, if $x_q Q_p > -x_p$, it will be positive. In econometric terms, this is equivalent of having left out quality from the demand equation - when price and quality are positively correlated - thus inducing an upward bias in the price coefficient that may easily overwhelm the pure price effect.

Presumably, however, we have explicitly taken care of 'quality' in the MNL equations - in that we included the vector of characteristics z - and therefore the positive price coefficients are still in want of an explanation. This we have to seek in the nature of the variables affecting choice: although the included attributes are certainly of prime importance in describing the quality of a CT scanner, some are only proxis to the 'true'performance dimensions of the systems, and they surely do not exhaust the relevant quality space (recall the discussion in IV above). It is important to emphasize that this is not just a problem for outside observers - such as ourselves - but it is rather a prime concern for the buyers of the systems themselves: observable attributes are only partially informative of the expected performance of a CT scanner (as is very often the case with durable goods), and the remaining uncertainty leaves room for prices to play a role as signals of quality[51] (see, for example, Alcaly and Klevorick [1970], and Pollack [1977]). In terms of the model outlined above, the demand function now becomes $x = x(p,z,q^e)$, where $q^e = Q(p,z)$ stands for 'expected performance', and z is the vector of observed characteristics. The price derivative is now

$$(29)\quad \frac{dx}{dp} = x_p + x_q Q_p \gtrless 0 \qquad \text{as} \qquad x_q Q_p \lessgtr -x_p$$

[51]Prospective buyers 'shopping' for CT scanners form expectations regarding attainable image quality on the basis of data provided by the manufacturers (such as spatial resolution), the visual inspection of sample pictures, etc.

where $Q_p(p,z)$ can be thought of as the 'signaling effect.' Note that q^e is assumed to depend also upon z, that is, consumers form their expectations regarding performance on the basis of prices, conditional on observed characteristics. In particular, we postulate,

$$(30) \quad q^e = Q[p-p(z)] = Q(\tilde{p})$$

where $\tilde{p}$ is, to recall, the residual from the hedonic price regression. In other words, the informative component of price resides in its deviations from the value that can be predicted on the basis of the data available to consumers (i.e., the vector z).

The key question is whether a signaling equalibrium can be assumed to hold over time, in the sense of actual - or ex-post - quality coinciding on average with expected performance (q^e), thus implying that residual prices convey the right information. Formal empirical tests for the existence of signaling equilibrium have yet to be designed; however, some inferences can be drawn by considering the market conditions required for either an equilibrium or a disequilibirum to emerge.[52]

In some circumstances, firms may have an incentive to violate the commit-

---

This constitutes, however, partial and imperfect information, and uncertainty remains as to the imaging performance of the systems under different operational conditions, as well as over time. Similarly, as ever faster scanners were introducted, users could not anticipate exactly the extent of expansion in the range of applications, or whether image quality could be maintained at high scan speeds. The firms themselves confronted users with additional sources of uncertainty (particularly in the late seventies),related to their upgrade, product line and servicing policies and, more importantly, to their survivability in the field.

[52]See Farrell [1979] and Wolinsky [1981] for models where prices do serve as signals of quality in equilibrium. For a qualified signaling equilibrium involving advertising levels rather than prices, see Schmalensee [1978].

ment implicit in a signaling equilbirum and make a 'quick kill', that is, price their goods higher than warrented by their true quality, realize extra profits, and exit the market once the bluff is called. Quite clearly, the likelihood of such behavior depends crucially on the extent and importance of intertemporal and intermarket links; in particular, it will be inversely related to the importance of reputation, both within the market of interest, and across related markets if the firms are diversified. Similarly, persistently high fixed costs - particularly in R&D and marketing - will strongly discourage the 'hit and run' behavior associated with a signaling disequilibrium. As suggested in IV above, those conditions are undoubtedly met in the market for CT scanners: reputation plays indeed a crucial role in medical instrumentation, most firms operate not just in CT but in a wide range of interrelated markets, sustained R&D and marketing efforts are key to commercial performance, etc. Thus, we can safely conclude that a systematic relationship between residual prices and unobserved quality dimensions was indeed maintained all along in this market.

The implications for welfare analysis are immediate: a signaling equilibrium implies that $q^e = Q(\tilde{p})$ should be regarded as an additional quality dimension, and be incorporated as such in the indirect utility function, i.e.,

(31) $V_i = V^1(z_i) + V^2(q_i{}^e) - \alpha p_i$

Assuming for simplicity that $V^2(q_i{}^e) = \beta q_i{}^e$, and $q_i{}^e = \theta \tilde{p} i$, and recalling that $p_i = p(zi) + \tilde{p}_i$,

$$V_i = V^1(Z_i) + \beta\theta\tilde{p}_i - \alpha[p(z_i)+\tilde{p}_i] \Rightarrow$$

(32) $V_i = [V^1(z_i) - \alpha p(z_i)] + (\beta\theta - \alpha)\tilde{p}_i$

Recalling that the bracketed term in (32) - defined as the 'net utility' - was approximated by a simple quadratic form on z, and denoting $\delta = (\beta\theta - \alpha)$,

$$(33) \quad V_i = \sum_j (a_j z_{ij} + b_j z_{ij}^2) + \delta\tilde{p}_i$$

which is the model actually estimated, except that now the price coefficient no longer stands for the marginal utility of income, α, but incorporates also the marginal utility of the signaling effect, $\beta\theta$ (thus, $\delta \gtrless 0 \iff \beta\theta \gtrless \alpha$). To insist, accepting the hypothesis of a signaling equilibrium implies that the estimated δ should indeed be used in computing V_i for purposes of welfare analysis (i.e., in order to compute $W(S_t)$ in (24)), whereas its rejection would have required that we delete $\beta\theta\tilde{p}_i$ from (33). Still, we need to know the true price coefficient in order to be able to integrate under the 'virtual' rather than under the observed demand function, the practical difference in this case being that it is α that appears in the denominator of (24), whereas using δ instead would be meaningless. Fortunately, the striking differences in the evolution over time of head and body scanners will allow us to associate α with the estimated price coefficients of <u>head</u> scanners.

Consider the following straightforward proposition: uncertainty with respect to product quality, and hence the extent to which prices may play a signaling role, will be greater the more technologically complex a product is, the less experience users have with it, and the faster is the pace of technological advance. Conversely, as the basic configuration and range of applications of a product stabilize, and as experience with it accumulates, the signaling effect will tend to dissappear. Recalling the description in Section

IV, it is immediately clear how this proposition applies to the two types of scanners: head scanners were introduced first, their applications remained unchanged (i.e. for brain studies), and, after an initial stage of improvements, the dominant trend was towards less expensive and simpler systems; on the other hand, the trend in body scanners was all along towards increased sophistication, with quantum technological jumps in the first couple of years, and a slow-down in the pace of change afterwards. Thus, we would expect (a) that the signaling effect vanished early on for head scanners, and hence that, after the first few years following their introduction, the estimated price coefficients would be negative and fairly stable; (b) that the price coefficients of body scanners would be systematically higher than those of head scanners, but that they would tend to converge towards the latter as the pace of innovation subsided. Figure 2, displaying the estimated price coefficients of the two types of scanners, strongly suports these conjectures. Thus, we can safely regard the price coefficients of head scanners for 1976-1979 as unbiased estimates of α, the marginal utility of income, and hence as the appropriate parameters to be used in computing $W(S_t)$[53]. Likewise, the difference between the price coefficients of body and head scanners can be taken as estimates of the magnitude of the signaling effect. Although the foregoing discussion centers on a particular market, we believe that this is just one instance of a fairly widespread phenomenon, and that the line of reasoning developed here could be applied elsewhere; in particular, we suggest that a lot can be learned about signaling effects and related phenomena by studying demand behavior in vertically segmented markets.

---

[53]To recall, head scanners were introduced in 1973, and hence we are allowing 3 years for the signaling effects to vanish; note than in 1975 the price coefficient is still positive and large, reflecting the uncertainty created by the introduction of the first head scanners to compete with the EMI scanner.

b. Alternative measures of welfare gains

We noted in V.d that the coefficients estimated in the MNL model change substantially from year to year, although the quadratic specification makes it difficult to perform systematic comparisons just with the coefficients themselves (see Trajtenberg [1983], Ch. VIII, for convenient ways of summarizing those estimates). However, it is virtually impossible to determine - from this evidence alone - whether the underlying, 'true' preference structure is indeed shifting over time or, instead, that it is stable but the coefficients change because the utility function is somehow misspecified. Be it as it may, the pertinent fact in the present context is that the preference structure as captured by the estimated choice model does change, and therefore the welfare measures are not unique but depend upon the choice of a reference - or base - year. Thus, there are two alternative ways of assessing the value of a change in the choice set from S_t to S_{t+1}:

(i) ex-ante: $\Delta W^a = W_t(S_{t+1}) - W_t(S_t)$

(ii) ex-post: $\Delta W^p = W_{t+1}(S_{t+1}) - W_{t+1}(S_t)$

where W_t stands for the surplus function using the coefficients estimated for year t. That is, the ex-ante, or forward-looking measure answers the following question: how much would the consumer be willng to pay for the option of facing

---

Since we were able to estimate these 'pure' price coefficients only for 1976-79, and in view of the fact that they are fairly stable, we shall use the 1976 coefficient as an estimate of α for earlier years, and that for 1979 as the parameter for subsequent years.

next year's choice set rather than the present one, given his/her preferences today? On the other hand, the question posed by the ex-post criterion is: how much income could be taken away from the consumer, so as to leave him/her indifferent between facing today's and yesterday's choice sets, in light of his/her present tastes. Note that even though this resembles the distinction between compensating and equivalent variations (or the Laspeyres - Paasche dichotomy), ours is in fact entirely different: in the traditional context tases are held fixed and the dilemma resides in choosing the reference utility level or consumption bundle, whereas here it is the taste perameters themselves that change.[54]

In general, the ex-ante and ex-post measures will provide different quantitative answers, and a priori it is not clear whether they would differ in a systematic way (e.g., setting upper and lower bounds as the Laspeyres and Paasche indices do) or, for that matter, which should be deemed to be more 'relevant.' However, the 'dual-inducement mechanism' mentioned in V.d would lead us to expect that the ex-ante measure be systematically higher than the ex-post. That is, 'strong' preferences for a given attribute today will induce a relatively large improvement in that attribute, which in turn will bring about a reduction in its marginal desirability next period; thus, the value of the improvement will necessarily be larger if judged according to the original preferences. Likewise, the same inducement mechanism seems to indicate that the ex-ante measure is somewhat more 'appropriate' or 'relevant', at least for policy considerations, since it could be argued that, for the sake of con-

---

[54]To put it differently, the only source of ambiguity in traditional welfare analysis lies in the existence of income effects: if these were absent then the two situations would be directly and uniquely comperable. In our case, however, the hypothetical measure $\Delta W = W_{t+1}(S_{t+1})-W_t(S_t)$ is meaningless regardless of

sistency, any kind of change should be judged according to the preferences that gave rise to it, rather than by hinsight. Nevertheless, and given that the issue is inherently inconclusive,[55] we shall compute both measures wherever possible; fortunately, the qualitative results hold for both measures equally well.

A word about Divisia-like measures, i.e., estimating each pair of adjacent years (or more) jointly, and using the resulting 'average' preferences to evaluate changes from one to the next. This is a legitimate and doable procedure, and it has the extra advantage of circumventing the problem altogether (i.e., it results in a unique measure). However, this very advantage is its weakness, for if tastes are indeed substantially different from year to year, it is hard to see how suppressing these differences (i.e., imposing a common set of coefficients) can render a 'better' measure. On the other hand, if preferences do not vary much then the ex-ante and ex-post measures would be very similar, and the problem would not be there to begin with.

c. Computations and results

First, note that the residual prices in (33) have to be obtained from the hedonic price function of the reference year, since p(z) (i.e., the non-stochastic component of the hedonic regression) has to be added back to the quadratic approximation in order to retrieve the original utility function. That is, in computing ΔW^a the hedonic price function of year t is used to generate the $\hat{\beta}$'s

---

income effects, since it is tantamount to making interpersonal comparisons.

[55] A similar problem arises in the context of endogenous tastes, and there too the choice of criteria for welfare analysis is not a closed issue; see, for example, von Weizsacker [1971].

of both years (t and t+1), and likewise, when computing ΔW^P the residuals are obtained from the hedonic price regression estimated for year t+1. Second, since we rejected the hypothesis that head and body scanners belong to a common preference tree (from 1976 on), the ΔW's are computed separately for each type of scanners in the period of 1976-1982. As for the initial years, the gains correspond to the joint set of head and body scanners, since the parameters used are those of 1975 (i.e., we use the 1975 W function to compute the 1973-74 and 1974-75 ex-post gains, and the 1975-76 ex-ante gains; note, however, that the 1973 and 1974 sets include only head scanners, one each year). For completeness, we impute the 1975-76 ex-ante joint gains to each type of scanners according to their respective ex-post shares: $s_H = \Delta W_H^P/(\Delta W_H^P + \Delta W_B^P)$, $s_B=1-s_H$.

We present in Table 8 the results for body scanners, performing the calculations in three stages : first, we compute

$$\Delta\tilde{W}_{t,t+1} = \ln (\sum_i \exp V_{t+1,i}) - \ln(\sum_i \exp V_{t,i})$$

as shown in column 1. Second, we divide $\Delta\tilde{W}$ by the true price coefficient, α, as listed in column 2; since α was estimated with prices defined in \$100K, we compute $\Delta W = (\Delta\tilde{W}/\alpha)100$ so as to present the gains in \$K. Finally, we deflate ΔW by the producer price index for capital equipment (the one most closely related to CT Scanners) in order to obtain the 'real' gains shown in column 5. We computed the gains for head scanners (1976-82) and for the joint set of head and body scanners (1973-76), in a similar fashion. Table 9 brings all these together, allowing us to compute the overall gains,

$$\Delta W = \pi(H)\Delta W_H + \pi(B)\Delta W_B$$

Table 8

Welfare Gains from Innovation in Body CT Scanners: 1976-1982

Period	(1) ΔW_B		(2) Price Coefficient (abs. value)	(3) ΔW_B=(1)/(2)x100		(4) Price Index (1982=100)	(5) 'Real' ΔW_B=(3)/(4) (in $K)	
	ex-ante	ex-post		ex-ante	ex-post		ex-ante	ex-post
1975-76	14.715[a]	2.930	0.748	1,967.2	391.7	.620	3,173.0	631.6
1976-77	5.416	0.764	0.748	724.1	107.8	.660	1,097.1	163.3
1977-78	0.110	-0.419	0.709	15.5	-46.9	.712	21.8	-65.9
1978-79	1.413	0.349	0.893	158.2	42.7	.774	204.4	55.1
1979-80	0.681	0.377	0.818	83.3	46.1	.858	97.0	53.7
1980-81	1.551	-0.073	0.818	189.6	-8.9	.945	200.6	-9.4
1981-82	1.709	-	0.818	208.9	-	1.00	208.9	-

[a] Imputed value

Table 9

<u>Welfare Gains by Scanner Type, and Incremental and Cumulative Overall Gains</u>

(all figures in $M)

	Head Scanners			Body Scanners			Overall Gains			
							Incremental		Cumulative	
Period	ΔW_H^a	ΔW_H^p	$\pi(H)$	ΔW_B^a	ΔW_B^p	$\pi(B)$	ex-ante	ex-post	ex-ante	ex-post
1973-74	-	1.2038	1.00	-	1.2038	0	8.7126*	1.2038	8.7126	1.2038
1974-75	-	.2085	.552	-	.2085	.448	1.5090*	.2085	10.2216	1.4123
1975-76	4.7758	.3191	.238	4.7758	.6318	.762	4.7758	.5574	14.9974	1.9697
1976-77	.0559	.0169	.151	1.0971	.1633	.849	.9399	.1412	15.9373	2.1109
1977-78	.3609	.2248	.276	.0218	-.0659	.724	.1154	.0143	16.0527	2.1252
1978-79	-.0127	-.0469	.298	.2044	.0551	.702	.1397	.0247	16.1924	2.1499
1979-80	-.0187	-	.196	.0970	.0537	.804	.0743	.0103*	16.2667	2.1602
1980-81	.0072	-	.084	.2006	-.0094	.916	.1844	.0255*	16.4511	2.1857
1981-82	-.0028	-	.065	.2089	-	.935	.1951	.0270*	16.6462	2.2127

* Computed using the mean ratio $\Delta W^a/\Delta W^p$ for 1976-79

where $\pi(h)$ is the percentage of buyers that purchased scanners of type h. Note that an overlapping series of ex-ante and ex-post pairs could be computed only for four years: 1976-1979. As expected, ΔW^a was found in those years to be systematically higher than ΔW^p by a relatively constant factor:

$$\text{Mean } [\Delta W_t^a/\Delta W_t^p] = 7.24, \qquad \text{S.D.} = 1.34.$$

Thus, we use this proportionality factor to obtain the 1974 and 1975 ex-ante gains (i.e., $\Delta W_t^a = 7.24 \times \Delta W_t^p$, t=74,75), and likewise for the 1980-82 ex-post measures. As a matter of terminology, we refer to ΔW_t as 'incremental' gains, to be distinguished from the cumulative gains $\sum_{\tau=1}^{t} \Delta W_\tau$, displayed in the last two columns of table 9 (the latter will be used only when assessing the impact of innovation on diffusion).

d. <u>On the Poor Health of Welfare Measures in the Medical Sector</u>

Although we have referred all along to $W(\cdot)$ as a 'surplus' or 'welfare' function, and likewise to ΔW as welfare - or social - gains, these notions need to be reassessed and qualified in view of the peculiarities of the case at hand. First, notice that the buyers of CT scanners are not the final consumers, but health-care providers that purchase the systems as a capital input and sell their services to patients. Now, this fact by itself does not invalidate, nor does it cast a shadow on, the welfare measures developed above (or any other measure of consumer surplus): it is easy to show that, if markets are perfectly competitive, then it is exactly equivalent to do welfare analysis either with reference to the demand for the final product (as is usually done), or to the derived demand for inputs (as we do here).[56] Likewise, government regulation in either market can only affect the magnitude of the surpluses actually realized, but the proposed measures will still capture them well (see for example White [1972]).

The problem resides entirely with the behavior of hospitals, both in itself and in relation to the utility of ill-informed patients. In order to pinpoint the source of the difficulties, let us consider in a very schematic way two 'ideal types' of hospital behavior: in the first the hospital is a profit maximizer and a price taker, and in the second it maximizes a utility function having as an argument the quality of CT scanners, z (we assume that the quality of medical services provided by the hospital is indeed a positive function of the quality of the CT scanner owned). In order to focus exclusively on quality choice, we assume that the quantity of medical services, m, is exogeneously determined. Since CT scanners are a capital input and we consider here the flow of services, there is a cost function determining the operating costs of a scanner, which we assume to be multiplicative, i.e., $C = C[z,p(z)]m$; thus,[57]

$$\frac{dC}{dz} = m\left(\frac{\partial C}{\partial z} + \frac{\partial C}{\partial p}\frac{\partial p}{\partial z}\right) \equiv mC_z > 0$$

Denoting by q(z) the price of a CT procedure performed with a scanner of qualty z, and assuming without loss of generality that m = 1, the behavior of the profit-maximizing hospital is simply

---

[56]All that this says is that in a perfectly competitive environment, the 'firm' is, in a sense, a redundant entity, and one could conduct any type of welfare analysis using only utility and costs functions, i.e., one could proceed as if consumers themselves purchased inputs and produced the final goods.

[57]It is assumed that more advanced scanners are more expensive to operate ($\partial C/\partial z>0$), and that they cost more ($\partial p/\partial z>0$). The term $\partial C/\partial p> 0$ involves interest - or leasing - payments, depreciation, insurance premiums, etc.

$$\max_z \Pi = q(z) - C[z,p(z)], \qquad \text{FOC: } q_z = C_z$$

In the case of a utility-maximizer, q(z) is replaced in the objective function of the hospital by the branch of its indirect utility function related to diagnostic medicine, $V^h(z)$ (we assume that the underlying utility function is additive separable), rendering FOC: $V_z^h = C_z$. Likewise, the behavior of the final consumer (the patient) is characterized by $\max_z V^c = V^c(z) + y^c - q(z)$, with FOC: $V_z^c = q_z$. If the market were perfectly competitive, then the following equalities would obtain,

$$V_z^c = q_z = C_z = V_z^h \tag{34}$$

Noting that the FOC for hospitals (i.e., $q_z = C_z$ and $V_z^h = C_z$) define the demand functions for CT scanners, it is clear that if (34) were to hold, then we would be indeed fully capturing the welfare of final consumers in our $W(\cdot)$ functions. Unfortunately, none of those equalities can be taken for granted; in particular, it is rather doubtful that $V_z^h = V_z^c$, i.e., that the interests and perceptions of the medical profession fully coincide with those of patients, as physicians choose diagnostic technologies. If anything, the presumption is that physicians may tend to display an upward bias in their valuation of those technologies, because of 'extraneous' motives such as rivalry between hospitals, long-range scientific goals that have little to do with the immediate well-being of patients, status and prestige, etc. If so, our measures ΔW will somewhat overstate the 'true' social gains, the difference being gains accruing to the medical community today, that probably will not be entirely passed-on in the long-run to society at large.[58]

---

[58]Notice that if the divergence between V_z^h and V_z^c was due primarily to long-

A more fundamental problem, though, is that V_z^c is by no means a well-defined construct to begin with, simply because patients are - for the most part - not in a position to evaluate independently the quality and medical value of diagnostic procedures, not to speak of innovations in them. Thus, we lack the baseline needed to assess the magnitude of the 'principal-agent' problem that may occur in the choice of medical technologies. In other words, we cannot really gauge the extent to which doctors may deviate from the choices that consumers would have made, if the latter had the same medical knowledge that their 'agents' command. However, that is a problem with medicine, not with economics: all <u>we</u> can do is to estimate those ΔW's, on the assumption that what was referred to as 'extraneous' motives are not the main determinants of hospitals' choices of technologies, and therefore that the - presumably upward - bias in the resulting measures are not too substantial.

VII. <u>Diffusion, Social Returns, and the Pattern of Innovation</u>

The ΔW's obtained in the previous section stand for the yearly <u>incremental</u> gains accruing to the 'representative' buyer of CT scanners. We want now to compute the yearly flow of <u>total</u> gains, relate them to R&D expenditures so as to obtain a rate of return and, more importantly, examine in detail the time profile of those benefits and costs.[59] As it will shortly become clear,

---

term scientific interests influencing the choice of technologies by hospitals, it is not clear which criteria should be used in evaluating the social gains from innovation in those technologies: since using V_z^c - if it were possible - would be too myopic, the issue boils down to the choice of an appropriate 'social rate of discount', to be applied to a measure based on V_z^h (such as our ΔW).

[59]The computations will be carried out using the ex-ante gains only: since those were found to be systematically higher than the ex-post measures, the total gains corresponding to the later would just be a - relatively constant - fraction of the former. The qualitative results to be discussed below are thus

obtaining those total gains is by no means a mere computational issue, but has to do instead with an important phenomenon in the realm of technical change, namely, the dynamic interaction between innovation and diffusion.

a. Total Gains and Diffusion

As a first step, consider the problem of delimiting - in time and in technology space - the benefits from innovation accruing to a consumer buying, say, a personal computer today. Are these benefits to be identified with the cumulative gains stemming from the long sequence of innovations in computers from the ENIAC on? Or perhaps just from the first Apple onwards? Alternatively, should we rather consider him/her as benefitting just from the last incremental gains, gauged by contrasting the 1985 versus the 1984 sets of personal computers in the market? Looking at it from a different angle, the same conundrum can also be phrased as follows: are we to compute the total gains generated by the innovations embedded in the 1985 set of available PC's simply by multiplying the incremental gains - ΔW_{85} - times the number of buyers of PC's in 1985? Or rather times the projected number of buyers from 1985 onwards? And what about those buying replacement units versus first-time buyers (i.e., new adopters)?

Aside from the issue of choosing an appropriate baseline,[60] it is clear that

---

unaffected by the type of measure used; on the other hand, the rate of return to R&D will obviously be much smaller if the ex-post gains were to be used instead of the ex-ante. As previously suggested the ex-ante gains seem to be a more meaningful measure, at least for normative analysis, but the issue is inherently inconclusive.

[60]By 'baseline' we mean the starting point for the computation of benefits, e.g., the first mainframe computer versus the first PC in the hypothetical case just mentioned (incidentally, why not the abacus?). If the goal is to analyse the welfare impact of a specific, well defined innovation(s) (as is the case here with CT), then the baseline is simply identified with the first appearance

the key to the problem lies in the dynamics of demand and the impact of successive innovations upon it. The proper way to address it would therefore be to formulate and estimate a dynamic discrete choice model, relying for example on Heckman [1981].[61] If that were available, then the total gains sought here would be obtained simply by integrating the ensuing intertemporal demand function, over individuals and over time. However, both conceptual and technical difficulties prevented us so far from estimating such a model. Thus, we shall limit ourselves here to a highly simplified version, that amounts essentially to the reduced form of a full-fledged dynamic model.

In order to grasp the nature of the reduced form model, consider the following polar versions of the diffusion process: in the first, diffusion is due entirely to the workings of the traditional 'demonstration' - or 'contagion' - effects, e.g., learning, emulation, rivalry, etc. In other words, the process has to do only with dynamic phenomena occurring within the population of potential adopters, and is not affected by external forces. Thus, for example, if technological change would have ceased immediately after the introduction of the first CT scanner, then under the assumptions of this model the pattern of diffusion would have been the same as it was in actuality. In the opposite extreme version the diffusion path is nothing but a temporal demand curve having only extensive margins, that is, it traces the distribution of some version of

---

of the new technology. The problem might arise when the question is posed the other way (as we did at first), i.e. when trying to assess the benefits accruing to a consumer at a point in time, since it is not clear then how far back into the past should one pursue those benefits.

[61]Such a model would comprise the decision of whether or not to buy in each period, as well as the conditional choice of what product to buy. The former, upper-level choice has to do primarily with the process of diffusion (i.e., when to adopt), although traditional investment motives, accounting for replacements and additions to capacity, should play an increasingly important role over time.

'reservation prices' in the population of potential adopters. Thus, successive innovations that result in what can be thought of as reductions in 'real prices' trigger immediate adoption by inframarginal individuals. Consequently, if the process of technological advance were to come to a halt diffusion would stop as well, all future purchases would be for replacement or capacity additions only, and hence the innovations that took place up to that point would cease to generate additional benefits.

The implications of these alternative scenarios for the computation of total gains are immediate: if diffusion corresponds to the 'distribution of reservation prices' case, then the total gains generated by innovations occurring at time t are just

$$(34) \qquad TW_t = \Delta W_t n_t,$$

where n_t stands for the number of buyers and TW_t for total gains. On the hand, if diffusion is due entirely to demonstration affects, then

$$(35) \qquad TW_t = \Delta W_t \; N \int_t^{\infty} f(\tau) e^{-r(\tau - t)} d\tau$$

where r is an appropriate discount rate, N is the size of the population of potential adopters, and $f(\cdot)$ the marginal distribution of adoption times, corresponding to the cumulative distribution $F(\cdot)$. To make it clearer, let us ignore discounting for a moment and rewrite (35) as,

---

These choice probabilities would be a function, inter alia, of the evolution of the technology over time, as well as of expectations in that regard.

$$TW_t = \Delta W_t\, N[1 - F(t)] = \Delta W_t[n_t + N(1 - F(t + 1))]$$

or, defining the number of future adopters as $n_t^f = N[1 - F(t + 1)]$,

$$(36) \qquad TW_t = \Delta W_t(n_t + n_t^f)$$

Contrasting (36) and (34), we see that the total gains would be larger in the 'demonstration effects' model, since they include also the benefits bestowed by current innovations to future buyers (with discounting n_t^f will be somewhat smaller).

Actual diffusion processes may correspond to either model or, most likely, to a combination of both, and it is of course an empirical matter to uncover the appropriate characterization. For that purpose we shall estimate the aggregate diffusion process as a function of time, and of the cumulative gains from innovation, $CW_t = \sum_{\tau=1}^{t} \Delta W_\tau$. As in traditional diffusion models, time is meant to capture the forces associated with the 'demonstration effects', whereas CW_t can be thought of as tracing the cumulative changes in a quality-adjusted index,[62] thus bringing-in the scenario associated with the 'distribution of reservation prices'. In particular, we assume that the diffusion path corresponds to a logistic distribution,[63] and that innovation impacts the process by raising the ceiling K, i.e., $K = K(CW_t)$. As to the functional form of $K(\cdot)$, we tried both a linear and a concave specification, corresponding to an underlying uniform and exponential distribution of 'reservation prices' respectively. Since the two yielded very similar results, we shall present here

[62]For a preliminary discussion of the construction of real price indices on the basis of our ΔW's, see Trajtenberg [1983], ch. X.

[63]The assumption of a logistic distribution is by no means an innocent one, particularly when the goal is to estimate correctly the speed of diffusion - see Trajtenberg and Yitzhaki [1982]. For our purposes here, however, the choice of

only those obtained with the linear form, $K(CW_t) = K_o + k\ CW_t$. The estimated equation is thus,[64]

$$(37) \qquad F(t) = (K_o + k\ CW_t)/[1 + \exp(\alpha - \beta t)]$$

Note that (37) cannot be linearized by taking the log of the odds ratio (as is usually done in diffusion studies), but has to be estimated instead with non-linear methods. We applied (37) to the adoption of CT scanners, with t defined in months and covering the period Nov. 1972 - July 1981, and the total population of potential adopters - hospitals and clinics - being N = 3,457.[65] The dependent variables is thus $F(t) = n_t/N$, where n_t stands for the number of first-time buyers in month t; the figures for CW_t are taken from table 9 (under 'overall cumulative gains -ex-ante'). As a benchmark we estimate also a logistic equation without $k \cdot CW_t$ but with a free ceiling.

The results are shown in table 10: first, note that diffusion was strongly influenced both by 'demonstration effects' (embedded in t) and by technological advance, as manifested in the fact that the estimates of β and of k are both highly significant; moreover, the fit improves greatly when going from (i) to (ii) (in terms of the RSS), implying that the traditional diffusion model, ignoring innovation, is widely off-mark. To put it in quantitative terms, k = 0.025 means that for every million $ worth of improvements in CT, the number of

---

a probability distribution does not make much of a difference.

[64]We experimented also with ΔW_t effecting the rate of diffusion, β, i.e., we estimated $F(t) = K(CW_t)/[1 + \exp(\alpha - \beta t - \beta' t\Delta W_t)]$; the coefficient β' turned out to be significant, but very small, and hence we omit those results here.

[65]This includes 3,078 community hospitals with more than 100 beds, plus 379 private clinics that actually purchased CT scanners by July 1981. For more details on the diffusion of CT scanners, see Trajtenberg and Yitzhaki [1982].

TABLE 10

Non-linear Estimates of the Diffusion of CT Scanners

$$F(t) = (K_o + k\ CW_t)/[1 + \exp(\alpha - \beta t)]$$

	K_o	k	α	β	RSS
(i)	0.459		4.053	0.071	0.013
	(.004)		(.076)	(.002)	
(ii)	0.074	0.025	3.443	0.06	0.006
	(.003)	(.002)	(.07)	(.001)	

Asymptotic standard errors in parentheses adopters increased by 2.5%; thus, if technological change would have ceased just after the introduction of the first CT scanner, only 7.4% of the total population would have adopted throughout the period (since $K_o = 0.074$). In reality, the ceiling had climbed to 49% by 1982, as a result of the flow of innovations from 1974 on.[66]

The estimated equation can also be put into use in order to obtain, albeit in an indirect way, a measure of the <u>initial</u> gains from innovation, that is, of the gains associated with the introduction of the first CT scanner (recall the discussion in II.a above). The question can be formulated as follows: what would those initial gains (ΔW_{73}) have to be, in order to give rise to the initial ceiling K_o, given the estimated (sub)function $K(CW_t) = K_o + k\ CW_t$? The answer is simply $\Delta W_{73} = K_o/k = 0.074/0.025 = 2.99$, that is, and to put it carefully, (ii) in table 10 is equivalent to an equation in which K_o is deleted, and $\Delta W_{73} = 2.99$ rather than zero. In other words, if the behavior underlying equation (ii) is stable, then the introduction of the first CT scanner had to be worth 3 million $ so as to induce 7.4% of hospitals and clinics to adopt the innovation. Although we shall use this figure along with the other measures in the coming sections, it should be born in mind that ΔW_{73} was computed in a very different, and probably less precise fashion.

A word about assigning benefits to purchases for replacement and additions to capacity: again, these can be properly dealt with only in the context of a full-fledged dynamic model, incorporating a capital accumulation process.

---

[66]Note that $CW_{82} = 16.65$, and hence $K_{82} = 0.074+0.025\times16.65 = 0.49$.

Short of that, and preferring to understate rather than overstate the total gains, we proceed on the assumption that the benefits accruing to repeat purchases at time t are just the incremental gains ΔW_t.[67]

We can now compute the yearly total gains as (note that n_t includes also second scanners and replacements),

$$(38) \qquad TW_t = \Delta W \, [n_t + N(K_o + k \, CW_t) \int_{t+1}^{\infty} f(\tau)e^{-r(\tau-t-1)}d\tau]$$

where the value of the parameters K_o,k and those of $f(\cdot)$ are taken from the diffusion equation as estimated in table 10, and the yearly discount rate is assumed to be .05 (since $f(\tau)$ is defined in months, the rate is actually .0041). The integral in (38) does not have a closed-form solution (because of discounting), and hence we had to resort to numerical integration. The computations are presented in table 11, which is largely self-explanatory. As suggested above, these figures are likely to be biased downwards, and should therefore be regarded as a lower bound: besides our assigning only the incremental gains to additional and replacement scanners, we have ignored the gains stemming from <u>upgrades</u>, i.e., from the retrofitting of older units, usually at a fairly low cost to users.[68] The only possible source of upward bias lies in the implicit assumption that the underlying distribution of 'reservation prices' is a step function, i.e., that inframarginal users are

---

[67]Note that this is the lower bound, the upper bound being the sum of the incremental gains from the time of the last purchase to the present one; the latter requires separate calculations for each indivdiual, making it too costly for us to undertake here.

[68]This was a fairly prevalent practice among the main manufacturers of CT scanners, and it was aimed primarily at overcoming fears of 'premature technological obsolescence' in this rapidly advancing field. Nearly 25% of all scanners installed up to mid-1981 have been upgraded, and hence ignoring them

TABLE 11

Computation of Total, Ex-ante Gains (in $M)

	(1) ΔW_t	(2) $K(CW_t)$	(3) 1-F(t)	(4) 1-F(t) 'Dis-counted'	(5) n_t^f	(6) $\hat{n}_t^f$	(7) n_t	(8) $n_t+\hat{n}_t^f$	(9) Total Gains (in $M)
1973	2.9900	.0740	.9383	.7720	240	197	16	213	638.3
1974	8.7126	.2898	.8776	.7506	879	752	43	795	6,926.3
1975	1.5090	.3271	.7769	.6853	879	775	221	996	1,502.9
1976	4.7758	.4453	.6285	.5676	967	874	374	1,248	5,959.1
1977	.9399	.4686	.4512	.4141	731	671	390	1,061	997.1
1978	.1154	.4714	.2854	.2646	465	431	250	681	78.6
1979	.1397	.4749	.1625	.1515	267	249	275	524	73.2
1980	.0743	.4767	.0862	.0806	142	133	271	404	30.0
1981	.1844	.4813	.0245	.0230	41	38	392	430	79.3
1982	.1951	.4861	.0121	.0113	20	19	428	447	87.2

See 'Notes to Table 11' on the next page

Notes to TABLE 11

(1) ΔW_t = incremental gains

(2) $K(CW_t) = K_o + kCW_t = 0.074 + 0.025CW_t$

(3) $1-F(t) = \sum_{t+1}^{\infty} f(\tau)d\tau$

(4) $1-F(t)$ 'discounted' $= \sum_{t+1}^{\infty} f(\tau)e^{-r(\tau-t-1)} d\tau$

(5) $n_t^f = (2)\times(3)\times N, \quad N=3,457$

(6) $\tilde{n}_t^f = (2)\times(4)\times N$

(7) n_t = number of scanners sold in year t

(8) $n_t+\tilde{n}_t^f$ = number of current and future 'discounted' beneficiaries from the incremental gains at t

(9) $TW_t = (1)\times(8)$ (total gains)

indeed just at the margin prior to purchase so that, as the latest innovations trigger adoption, they receive ΔW_t in full; otherwise only a fraction of the incremental gains would actually be realized.[69] On the other hand, the restraining influence of 'technological expectations' may significantly reduce, and even reverse this potential bias.[70]

b. R&D expenditures and social returns

Having thus obtained the stream of social benefits from innovations in CT, we want now to relate them to the costs of bringing them about, i.e., to R&D outlays on CT, and compute some version of a rate of return. First, a brief note about the data: one of the most pervasive problems in empirical micro studies is the difficulty in obtaining reasonably good data at the level of disaggregation prescribed by theory, i.e., specific, well-defined 'industries' or 'markets'. R&D is particularly troublesome in that respect, since the allocation of R&D expenditures across projects within firms is usually regarded as highly confidential information, and there is little - if any - external evidence of such allocation. Furthermore, even if that kind of information were available, it is not clear how one should distribute overheads among the firm's various research fields, not to mention the problem of accounting for externalities across R&D projects. Not surprisingly, studies

---

might result in a significant undercounting of benefits.

[69]Obviously, the problem arises only because we are working with discrete time periods: this would not be the case if we could compute dW_t for sufficiently small time intervals.

[70]That is, users may delay adoption even though they have already passed their threshold, expecting further technological improvements (see, for example, Balcer and Lippman [1982]); in that case the gains at the time of adoption will exceed the latest incremental gains.

using product-specific R&D are extremely rare.

In spite of those difficulties, we have succeeded in gathering fairly complete data on the R&D devoted specifically to CT scanners by almost all firms that have - or had - been active in this field. Even though the sources used vary substantially in their reliability and extent of coverage, there is plenty of independent evidence indicating that at least the orders of magnitude are correct(see Trajtenberg [1983], section VII.3). The yearly figures for total and US only R&D expenditures in CT are shown in table 12 (the breakdown by firm cannot be exhibited because of a pledge of confidentiality).

The question now is whose R&D should be used in computing a social rate of return, that is, are we to count the R&D performed both in the US and abroad, or just the R&D done by US firms? The source of ambiguity lies in the fact that we have estimated the benefits accruing to US users only, whereas the market for CT is global, both in terms of the origin of innovations and in the spread of their benefits. Thus, for example, if we were to use just the R&D performed by US firms, the resulting rate of return could be seen as an overstatement, since the measured gains are due in part to foreign innovations whose costs are not counted. Given the inconclusive nature of the issue, we shall compute rates of return both to total and to US-only R&D, and loosely interpret them as lower and upper bounds to the 'true' rate. As to the computational procedure, we have chosen the capitalized benefit/cost ratio,[71] (see, for example, Griliches [1958]), i.e.,

[71]An alternative would have been to use the internal rate of return; however, and aside from the fact that it may not render a unique result, this method is too sensitive to the benefits and costs in the initial period. Consequently, small changes - or errors - in the magnitude or timing of those initial figures (that usually are the least trustworthy), can drastically alter the results.

TABLE 12

R&D Expenditures in CT Scanners

(in constant 1982$-millions)[a]

	(1) R&D by US firms	(2) Total R&D[b]	(3) Total No. of Firms	(4) R&D per Firm (2)/(3)
1968-71	-	6.22	1	1.55[c]
1972	.22	5.42	2	2.71
1973	.82	9.00	3	3.00
1974	7.79	22.62	8	2.83
1975	28.63	59.68	12	4.97
1976	58.18	96.08	13	7.39
1977	46.64	79.68	14	5.69
1978	37.03	64.33	11	5.85
1979	33.70	56.05	9	6.23
1980	29.58	46.40	8	5.80
1981	22.58	37.94	8	4.74
Total	265.17	483.42		
Mean (1974-81)	33.02	57.85	10	5.44

[a] The R&D deflator used is taken from Cummins and Hall [1982]

[b] CGR, Hitachi and Philips are not included

[c] Average for the 1968-71 period

$$i = rB/C = r[\sum_{\tau=73}^{82} TW\ \gamma^{(\tau-73)} / \sum_{\tau=68}^{81} R\&D_\tau \gamma^{(\tau-73)}]$$

where $\gamma = 1/(1 + r)$, with the interest rate assumed to be $r = 0.05$. Applying this formula to the two alternative R&D series.

(i) R&D in US only: $i = .05(14,813/214) = 3.46$

(ii) total R&D: $i = .05(14,813/397) = 1.87$

that is, a dollar of R&D expenditures in CT scanners resulted in $3.5 of annual returns in perpetuity when only local R&D is considered, and in $1.9 when foreign R&D is included as well. Taking their average as a summary figure (recall that (i) and (ii) can be interpreted as upper and lower bounds), we conclude that the social rate of return to R&D in this field was about 270%. To recall, we have not included profits in the measure of benefits, because of lack of accurate data. However, even if profits amounted to, say, a staggering 50% of revenues over the whole period,[72] their impact on the computed rate of return would be negligible, simply because the TW's are larger by a few orders of magnitude. Thus, the above figure can indeed be regarded as a social rate of return, subject only to the qualifications voiced in VI.d.

What can be learned from this result? First, a rate of 270% reaffirms the view that R&D is indeed a highly rewarding activity for society as a whole, its returns greatly exceeding those generated by more conventional forms of investment. This is hardly news, but the sparsity of solid quantitative results

---

[72]This exceeds what any firm had made in any particular period, not to mention the fact that a few actually incurred substantial losses.

of this kind, and the traumatic impression left by the productivity slowdown, seem to have fostered the (mistaken) perception that we may be exhausting the 'technology frontier'. True, our ability to produce ever larger quantities of existing goods may not be increasing as much as in the good old days, but this very same economy can and does generate new and better products with a vengeance, only that we are not used, or have not yet learned, to substitute dollar figures for 'new' and 'better'.

Unfortunately, there are very few studies where rates of return to R&D for individual innovations have been computed, and hence it is difficult to assess the significance of the particualr figure arrived at here. Although referring to a process rather than a product innovation, the main study in this context is still Griliches [1958][73]: he estimated the social returns to R&D in hybrid corn to be about 700%, much higher than our estimates for CT scanners. Although there is no reason whatever to expect that social, ex-post rates of return for unrelated innovations will be of a similar order of magnitude, a closer look reveals that in this case a methodological difference accounts for most of the disparity in results. This has to do with the treatment of future benefits: in Griliches' study the flow of benefits continues ad infinitum at the level determined by the ceiling of the diffusion curve, i.e., whenever hybrid corn is planted, now as well as in the future, society gets the benefits of its superior yield vis à vis conventional corn. By contrast, in our case the benefits cease as the diffusion process dies off, since the purchase of additional - or replacement - scanners in the future does not confer further gains. If we

---

[73]The other important study of the kind is Mansfield et al [1977]: they found that the median social rate of return in a set of 17 innovations was 56%, the highest being 307%. Unfortunately, their results are not directly comparable to ours, since they used internal rates of return rather than benefit/cost ratios

recalculate the rate of return on hybrid corn omitting 'future' benefits (i.e., those accruing after 1955), using the figures for cumulated social net returns and cumulated research expenditures from table 2 in Griliches [1958], we get $i = .05(4{,}405/63) = 3.5$, almost the same as our upper bound of 346%. This similarity of results is somewhat suggestive and mildly reassuring, but in order to be able to draw solid inferences one would need further empirical studies of a similar nature, and a much better understanding of the determinants of social returns to R&D.

c. The time profile of benefits and costs

Moving beyond the summary view provided by rates of return, we proceed now to examine in detail the evolution over time of both social gains and R&D expenditures: in so doing we hope to shed some light on the dynamics of the innovative process itself. To that end we present in Figure 2 the time path of incremental gains, both the actual figures and a 3-year moving average,[74] and in figure 3 the profiles of - the log of - total gains and total R&D expenditures.[75] To recall, the incremental gains reflect advances in the technology itself (that is, the social valuation of those advances), whereas total gains incorporate also the effect of a changing market size. However,

---

(recall footnote 68).

[74]There is room to believe that the timing of specific innovations (e.g., the date when the improved EMI head scanner was introduced in the US market) is effected by random events, and, given the arbitrariness of any discrete partition of the time dimension, a moving average may better capture the essence of the underlying process.

[75]We had to use logs (natural logarithms) rather than the actual numbers in plotting figure 3, simply because of the enormous differences in the order of magnitude of the figures: the total gains in the 1973-1977 period are in the billions of dollars range, whereas later on they dip below 100 millions $, and the R&D expenditures are well under that mark throughout.

Figure 2

The Time Profile of Incremental Gains

(Yearly figures and 3-year moving averages)

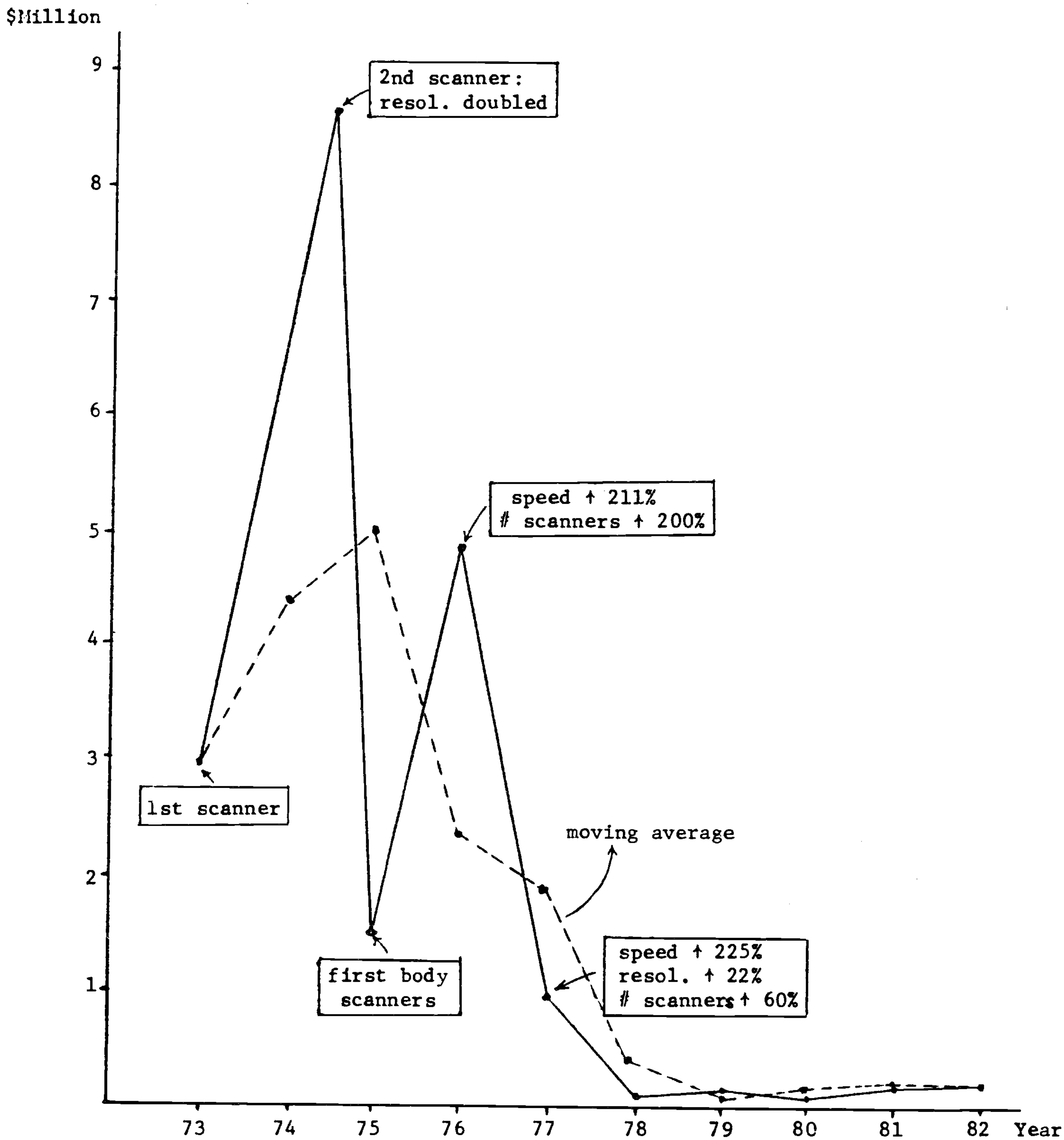
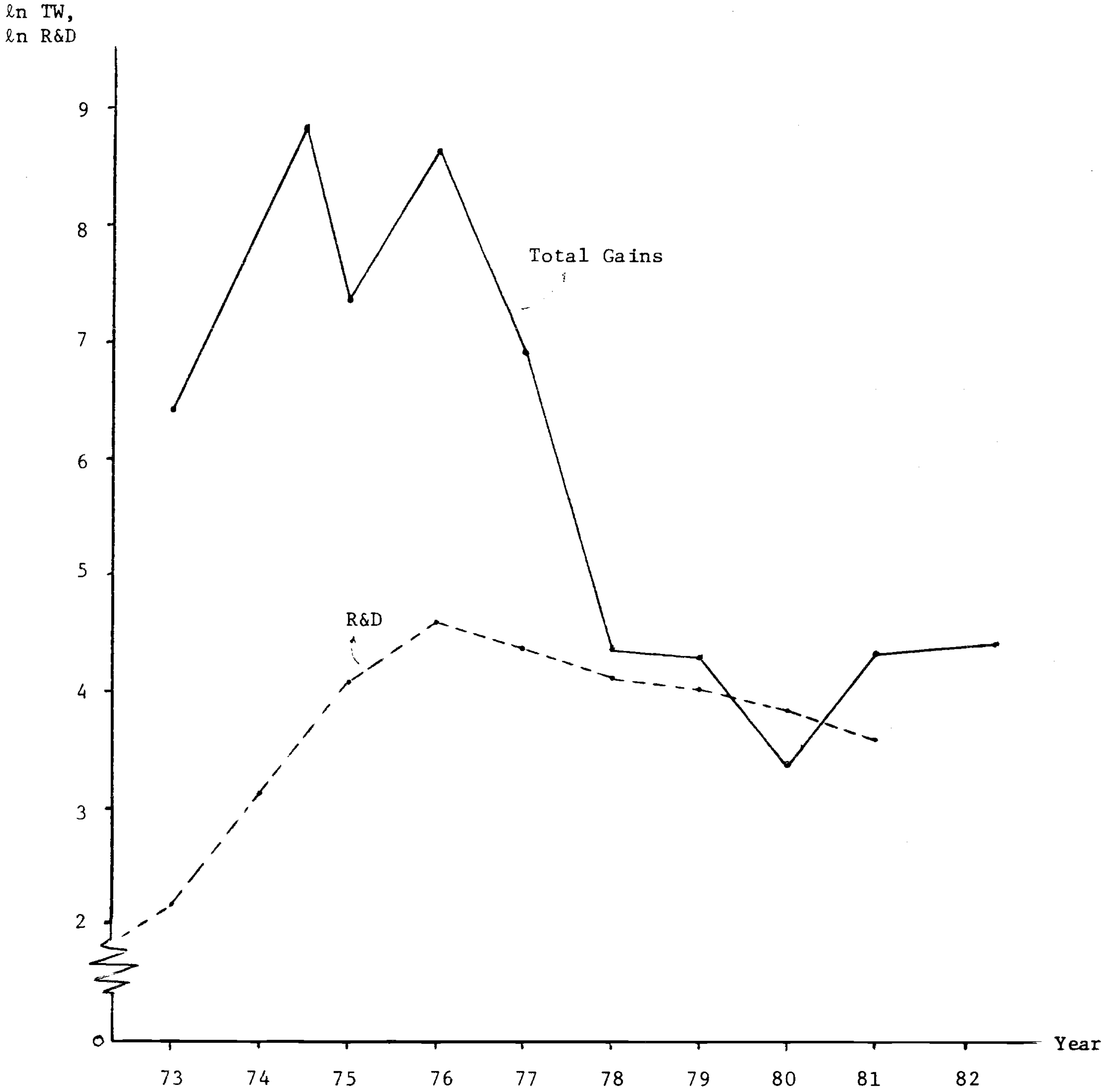


Figure 3

Total Gains and Total R&D Expenditures

(logarithmic scale)



Note: Original figures for total gains and R&D: in constant 1982 million $

since successive improvements in the technology were found to strongly influence the size of the market (by raising the ceiling of the diffusion process), the time profiles of both incremental and total gains look very much alike, except of course for differences in scale.

The first feature to note is that the gains generated in the first half of the period are far larger than those in the second half. In fact, the smoothed-out - or stylized - profile resembles a log-normal distribution (or a normal density truncated at about the first quartile), that is, it starts high, rises still further during the initial period, and then declines rapidly, carrying a long, low-level tail into the future.[76] Such a pattern is, on reflection, highly plausible and appealing, and it may be possible to account for it by a - generalization of a - fairly common feature of economic processes, namely, an initial, short-lived phase of scale economics, promptly followed by the setting-in of sharply diminishing returns. What is peculiar in the context of innovation is that these increasing and decreasing returns appear to 'take place' in three different dimensions: in the 'production' of innovations (i.e., in R&D/technology), in the utility or value generated by them, and in the determination of market size (that is, the size of the population of adopters).

With respect to R&D, it is usually relatively 'easy', both from the viewpoint of the resources needed and of the underlying scientific and technological principles, to improve the performance of a technology during its initial stages. Later on, however, as the obvious advances are no longer there to be made and the technology is pushed to its limits, the marginal cost of

---

[76]Although we have measured the gains only up to 1982, we know that in the period 1982-1985 no major technological advances have taken place; still, some minor improvements were introduced, similar in character to those occurring in 1980-82. Thus, it is reasonable to assume that the tail will indeed extend into

further improvements rises, and thus the rate of innovation tends to abate. As to the second dimension, it seems that there is often a 'threshold effect' in the utility derived from the characteristics of new products, in the sense that below a certain level of performance the product is pretty much worthless (that was, for instance, the case with minimal resolution in CT scanners). That would account for an initial phase of 'increasing returns' in the valuation of technical improvements, but soon after diminishing marginal utility prevails: thus, for example, increasing the speed of CT scanners from, say, 12 to 2 seconds was not nearly as valuable as going from 60 to 10 seconds. As to the third dimension, market size, consider how n_t^f evolves over time (recall that n_t^f stands for the number of users that benefit from the incremental gain ΔW_t as the diffusion process unfolds):[77] from (38), and assuming for simplicity $K_o = 0$,

$$\frac{dn^f}{dt} = k \frac{CW_t}{\partial t}[1 - F(t)] - k\ CW_t f(t)$$

and hence,

$$\frac{d\ \ell n\ n^f}{dt} = \frac{d\ \ell n\ CW_t}{dt} - h(t) \tag{39}$$

where h(t) stands for the 'hazard rate', i.e., $h(t)=f(t)/[1-F(t)]$. In words, the behavior of n_t^f over time depends upon the rate of change of cumulative gains vis à vis the hazard rate. As argued above, we would expect the former to be non-increasing, whereas if diffusion follows a logistic pattern, $h(t)=\beta F(t)$ and hence it increase monotonically. Thus, n_t^f will in general be a concave function

the future.

[77]This is obviously relevant only for the time profile of total gains.

and have a maximum, i.e. even if innovation proceeds at a constant pace, total gains will eventually decrease, due to the fact that the diffusion process exhausts itself faster than the rate of expansion of the market brought about by technological change.

As a convenient illustration of those effects, consider the innovation that generated the largest benefits in the history of CT, namely, the introduction of the second CT scanner model (the EMI CT 1000) in 1974: the basic design of the system was virtually identical to the original EMI Mark I, except that its resolution was almost doubled, a change that improved dramatically the ability to visualize brain pathologies.[78] Thus, if the first scanner proved the <u>feasibility</u> of CT, the second transformed it into a useful diagnostic tool that could be widely applied. This seems to be a fairly general phenomenon in product innovations: the first commercial models of a new product are rarely more than just the embodiment of a potentially useful idea; the real leap-forward (and the concomitant benefits) comes with the advent of a model in which some key quality dimensions are greatly improved, turning it into a <u>practical</u> - rather than just an ingenious - device that can command wide appeal. Examples about: the Ford-Model T in cars, the DC-3 in commercial aircraft, the UNIVAC I in mainframe computers, etc.

This early emergence of a 'big winner' that brings about the most gains (let us refer to it as the 'Ford-T effect') can thus be seen as the lumpy realization of 'increasing returns', that operate <u>simultaneously</u> along the three

---

[78]It also had a much shorter reconstruction time, but that was not a crucial characteristic at the time.

dimensions mentioned above, i.e. technology, utility and market size. Likewise, the rather dramatic drop in the flow of gains from innovation occuring later on, can be attributed to diminishing returns setting-in at the same time in those same dimensions.

The idea that the process of technological advance may follow a cyclical pattern, and perhaps even be the driving force behind wider economic fluctuations, has been for a long time a favorite theme for some, as much as a dubious subject of speculation for others.[79] Once again, the lack of appropriate measures of innovation has precluded so far the uncovering of solid empirical evidence that could have 'set the record straight'. Taking a step in that direction, our study shows that in one particular instance innovation did follow indeed a wave-like pattern over time. Needless to say, much more is needed in order to establish whether or not technological progress at large conforms to a cyclical pattern; in particular, one would need to show: (a) that most technologically progressive fields exhibit a similar innovative time profile, and (b) that the innovation processes in different fields are not synchronized or contiguous timewise, but rather sparce and lumpy (by synchronized we mean that as one field starts experiencing a decline the next flourishes, so that even if each fulfills (a), we would still observe in the aggregate a smooth and approximately constant rate of innovation). To insist, there is nothing in our study to warrant such far-reaching generalizations (no individual case study can do that); however, since casual observation and common experience (those much abused providers of 'informed guesses') seem to indicate that (a) and (b) might

---

[79]Schumpeter [1939] is still the most important and comprehensive statement on this subject. As of late there has been, however, a resurgence of interest in the possible existence of 'technological cycles": see, for example, Sahal [1981], Judd [1985], and Shleifer [1985].

indeed be quite prevalent, the uncovering of the evidence presented here, even if it refers just to one case, does enhance the plausibility of 'technological cycles'. Of course, only further empirical studies of a similar sort can bring us closer to a more definitive statement on this much debated matter.

A few remarks about the time profile of R&D expenditures: as figure 3 reveals, the flow of R&D outlays closely traces, with one year lead, the level of activity - or 'action' - in the field, as manifested in, say, numbers of entrants and of new models, sales to new adopters (see table 1), etc. On the other hand, it is not highly correlated with either incremental or total gains, reflecting the above-mentioned increasing/decreasing returns sequence in the technology dimension. To further highlight this notion, consider that the ratio of social gains to R&D for the period 1968-1977 was a staggering 80 to 1 (taking R&D up to 1976, on the assumption of a minimal lead-time of one year), whereas the rate for 1978-1982 was a bare 1.4, and in 1980 the R&D outlays actually exceeded the benefits. A straightforward implication is that average social rates of return to R&D are not informative enough to serve as guides for policy: the question is not so much whether public support to R&D is warranted (it is not too difficult to make a case for it), but rather until when along the innovation cycle should such support be provided. In the case of CT, for example, it is clear that whereas it was socially desirable to promote research during the inital stages, that was certainly not the case after 1976.[80] The problem is that if one were to wait for hard evidence on returns to R&D over time in order to decide what fields to support, the result might well be a

---

[80]It is worth noting that the British Department of Health supported Hounsfield's construction of the first prototype CT scanner at EMI in 1970-72, and it seems that such support was crucial for the carrying out of the project. Conversely, the American neurologist W.H. Oldendorf had independently developed

stand-off, i.e. once the evidence is available the support may no longer be necessary. Thus, learning more about typical time profiles of innovations may substantially contribute to the design of public policy in this area.

d. Further Applications

The bulk of this paper has been devoted to methodogical and implementation issues in the measurement of product innovation, and hence there is little room left here for applications, other than the most immediate ones as reported in the previous section. To be sure, there are plenty of interesting ways in which those measures can be brought into play: some we have already explored in Trajtenberg [1983] and elaborated in forthcoming papers, but a lot remains to be done. Here we shall limit ourselves to mentioning the following: first, the ΔW's can be easily related to the dynamics of market structure (e.g, changes in concentration or in monopoly power), thus providing the means to address one of the key issues in this area. Preliminary results show a strong positive correlation between innovation and competition, and suggest simultaneity rather than simple-minded one-way causal links. Second, the ΔW's can be used to construct quality -adjusted price indices (or, more precisely, cost of living indices), far superior to those that can be computed just from hedonic price regressions. Third, these measures can help shed light on the usefulness of patent counts for the study of innovation: so far we have found that whereas simple patent counts correlate well with R&D outlays, patents weighted by number

earlier on a tomographic device based on principles similar to the EMI scanner, but he was unable to pursue the project for lack of support in the US.

of citations follow fairly closely the time profile of total gains. This seems to be enough for now as a stimuli to the research agenda on technological change.

References

Alcaly, Roger E. and Klevorick, Alvin K. (1970). "Judging Quality by Price, Snob Appeal, and the New Consumer Theory". Zeitschrift für Nationalokonomie, 30(1-2), pp.53-64.

Balcer, Yves and Lippman, Steven A. (1982). "Technological Expectations and Adoption of Improved Technology". Social Systems Research Institute, University of Wisconsin-Madison, workshop series #8226.

Becker, Gary S. (1965). "A Theory of the Allocation of Time". Economic Journal, 75, pp.493-517.

Berndt, E., J. Hausman, B. Hall, and R. Hall (1974). "Estimation and Inference in Non-Linear Structural Models". Annals of Economic and Social Measurement, 3(4), pp.653-665.

Cowling, K., and Rayner, A.J. (1970). "Price, Quality, and Market Share". Journal of Political Economy, 78(6), pp.1292-1309.

Cummins, Clint and Hall, Bronwyn (1982). "The R&D Master File Documentation". National Bureau of Economic Research. Unpublished manuscript.

Farrell, Joseph (1979). "Prices as Signals of Quality". Brasenose College, Oxford. Unpublished manuscript.

Griliches, Zvi (1958). "Research Costs and Social Returns: Hybrid Corn and Related Innovations". Journal of Political Economy, 66, pp.419-31.

__________, ed., (1971). Price Indexes and Quality Change. Cambridge, Massachusetts: Harvard University Press.

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

Hausman, J., and Wise, D.A. (1978). "A Conditional Probit Model of Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences". Econometrica, 46, pp.403-426.

Hausman, Jerry, and McFadden, Daniel (1981), "Specification Tests for the Multinomial Logit Model". M.I.T. Working Paper #292.

Heckman, J. (1981). "Statistical Models for the Analysis of Discrete Panel Data". In Structural Analysis of Discrete Data, edited by C. Manski and D. McFadden. Cambridge: M.I.T. Press.

Hensher, David A., and Johnson, Lester W. (1981). Applied Discrete Choice Modelling. London: Croom Helm; New York: Wiley.

Judd, Kenneth (1985). "On the Performance of Patents". Econometrica, Vol. 53, no. 3, pp.567-586.

Krantz, D., Luce, D., Suppes, P. and Tversky, A. (1971). Foundations of Measurement, Vol. 1. New York: Academic Press.

Kuznets, Simon (1962). "Inventive Activity: Problems of Definition and Measurement". In the Rate and Direction of Inventive Activity: Economic and Social Factors, Universities-National Bureau Conference Series No. 13. Princeton, N.J.: Princeton University Press.

Lancaster, Kelvin J. (1966). "A New Approach to Consumer Theory". Journal of Political Economy, 74, pp.132-57.

__________ (1971). Consumer Demand: A New Approach. New York and London: Columbia University Press.

__________ (1979). Variety, Equity, and Efficiency. New York: Columbia University Press.

McFadden, Daniel (1981). "Econometric Models of Probabilistic Choice". In Manski, C. and McFadden, D. (eds.), Structural Analysis of Discrete Data with Econometric Applications. Cambridge, Massachusetts: The M.I.T. Press.

__________ (1982). "Econometric Analysis of Qualitative Response Models". M.I.T. Unpublished manuscript.

Mansfield, Edwin (1968). Industrial Research and Technological Innovation - An Econometric Analysis. New York: Norton.

Mansfield. E., Rapoport, J., Romeo, E., Wagner, S., and Beardsley, G. (1977). "Social and Private Rates of Return from Industrial Innovation". Quarterly Journal of Economics, 91(2), pp.221-40.

Muth, Richard R. (1966). "Household Production and Consumer Demand Functions". Econometrica, 34, pp.699-708.

Pollak, Robert A. (1977). "Price Dependent Preferences". American Economic Review, 67(2), pp.64-75.

Rosen, Sherwin (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition". Journal of Political Economy, 82(1), pp.34-55.

Sahal, D. (1981). Patterns of Technological Innovation. Reading, Massachusetts: Addison-Wesley.

Scherer, F.M. (1980). Industrial Market Structure and Economic Performance. Chicago: Rand McNally College Publishing Company.

Schmalensee, R. (1978). "A Model of Advertising and Product Quality". Journal of Political Economy, 86(3), pp.485-503.

Schumpeter, Joseph (1939). Business Cycles. New York: Harper and Row.

Shleifer, Andrei (1985). "Implementation Cycles". M.I.T. Unpublished manuscript.

Small, Kenneth A. and Rosen, Harvey S. (1981). "Applied Welfare Economics with Discrete Choice Models". Econometrica, 49(1), pp.105-130.

Spence, Michael A. (1973). "Price, Quality, and Quantity Interdependencies". Center for Research in Economic Growth. Stanford University, Memorandum #161.

__________ (1975). "Monopoly, Quanlity, and Regulation". The Bell Journal of Economics, 6(2), pp.417-29.

Trajtenberg, Manuel (1979). "Quantity Is All Very Well But Two Fiddles Don't Make a Stradivarius. Aspects of Consumer Demand for Characteristics". The Maurice Falk Institute for Economic Research in Israel, Discussion Paper #7910.

__________ (1983). "Dynamics and Welfare Analysis of Product Innovations". Harvard University. Unpublished Ph.D. dissertation.

__________ (1984). "The Use of Multivariate Regression Analysis in Contrast-Detail Studies of CT Scanners". Medical Physics, 11(4), July/August 1984, pp.456-64.

Trajtenberg, Manuel, and Yitzhaki, Shlomo (1982). "The Diffusion of Innovations: A Methodological Reappraisal". National Bureau of Economic Research, Working Paper No. 1008.

Von Weizsäcker, Carl C. (1971). "Notes on Endogenous Change of Tastes". Journal of Economic Theory, 3(4), pp.345-72.

White, L. (1972). "Quality Variation When Prices Are Regulated". Bell Journal of Economics, Autumn, pp.425-36.

Wolinsky, Asher (1981). "Prices as Signals of Product Quality". Bell Laboratories, Economic Discussion Paper #232.