present and future of the important science of econometric model building. His is a perspective that has been shaped by his manifold contributions to the field and by his own personal approach to economic research. I interpret this approach or style as a thrust for the widest possible explanation of economic phenomena—breathtakingly wide at times—combined with a very high level of theoretical and econometric sophistication. But at all times it has also meant keeping close to the data—always giving the data the chance to refute or modify our hypotheses, but also doing such mundane tasks as keeping meticulous records of forecasts and trying to improve the database. His approach has done much to make economics a science.

NOTES

1. This statement, of course, applies only to the application of simultaneous estimation methods to large econometric models.
2. See e.g. Green, 1967.
3. For a fairly detailed discussion of aggregation conditions as applied to aggregation across countries, see Stevens et al. 1984, chapter 2.
4. The Fed's Multicountry Model in 1984 had five countries and an aggregate "rest of the world." The model was limited to one composite good per country, a very limited treatment of oil and raw materials, and no breakdown of fixed investment into any of its components. See Stevens, ibid.

REFERENCES


The Economics of Technical Progress

Ryuzo Sato and Thomas Mitchell

0. INTRODUCTION

Technical progress is a dynamic phenomenon of great importance to all economies. One simple-minded description of the phenomenon would be that it is a transition to a new technological state in which greater output rates are producible from a given stock of inputs. Alternatively, technical progress reduces the input stock requirements for producing given rates of outputs. The significant implication of ongoing technical progress is the productivity growth which is a prerequisite for raising the standard of living in an economy.

The concept of technical progress and productivity growth is as old as Adam Smith's Wealth of Nations (Smith, 1937). In the first pages of the first chapter, Smith described how a simple division of labor can lead to an increase in the rate of output of pins without any changes in the quantities, or qualities, of the productive inputs. This is but one example of what we would generally consider to be "technical progress." A modern characterization of technical change—which includes technical regress or productivity decline, as well as technical progress or productivity growth—is the oft-quoted description presented by Solow (1957, p. 312): "technical change [is] a short-hand expression for any kind of shift in the production function. Thus slowdowns, speedups, improvements in the education of the labor force, and all sorts of things will appear as 'technical change.'"

More recently, Ball and Griffin (1988, p. 23) observed that Solow's definition was of a sufficiently general nature to allow technical change to "reflect the effects of short-run disequilibria, as well as the long-term effects of the diffusion of new processes associated with technological change," even though the latter case probably comes to mind more readily as an example of technological change. This general view of technical change is the one to which we subscribe. For a mathematical formality, define in a traditional way a scalar production function \( f(x) \), where \( x \) is a real, nonnegative n-vector of inputs (Shephard, 1970):

\[
f(x) = \max \{ y : x \text{ can produce } y, y \geq 0 \},
\]

where \( y \) is a rate of output. Technical change can then be viewed exactly as Solow saw it, as anything which would affect the solution of the above maximization problem.

Hoping that the above suffices to define and delimit the scope of what we mean by "technical change" (at least for the present), what we would like to do in what follows is present a brief summary of the vast literature on productivity and technical change which has sprung up since Solow's seminal paper in 1957. As a disclaimer of sorts, we note that our focus in Part I of the paper will be on the theoretical and empirical developments in the parametric approaches to exogenous technical progress. This is referred to generally as the production (or cost) function

*Center for Japan-U.S. Business & Economic Studies, New York University, New York, NY 10006 and Department of Economics, Southern Illinois University at Carbondale, Carbondale, IL 62901-6519, respectively. This paper was prepared for presentation at the annual meeting of the Eastern Economic Association, Baltimore, Maryland, March 3-5, 1989. The authors are grateful to Nancy Maltz, who cheerfully typed this manuscript with a very high productivity index.

This paper was prepared for presentation at the annual meeting of the Eastern Economic Association, Baltimore, Maryland, March 3-5, 1989. The authors are grateful to Nancy Maltz, who cheerfully typed this manuscript with a very high productivity index.
approach. In the second part of the paper we will consider developments in the analysis of endogenous technical change, and finally Part III presents a differential game model of technological competition between countries. For topics outside the scope of this survey, we refer the curious reader to Bilsawanger and Ruttan (1978) for the study of "induced innovation," and for the area of nonparametric measurement of productive efficiency see Färe, Grosskopf and Lovell (1983) for a starting point. For other issues in the analysis of technical progress, see Nelson's (1981) survey.

I. THE ECONOMIC ANALYSIS OF EXOGENOUS TECHNICAL CHANGE

A. Early Aggregate Studies

Although earlier studies had been done—see, for example, Schmoekel (1952) and Valavanis-Vall (1955)—it was Robert Solow's (1957) study of aggregate U.S. data for the period 1909–1949 that first attached significant empirical importance to the contribution of technology changes in the growth process. Solow based his study on a aggregate production function

\[ Q = A(t) \cdot f(K, L) \]

where, it is interesting to note, Solow let "K and L represent capital and labor inputs in 'physical' units," A(t) is the time dependent shift factor, and Q is output. It is assumed that factors receive their marginal product as a real factor price, then total differentiation of (1.1) with respect to time and division by Q yields the familiar growth accounting relation

\[ A = \dot{A} = \dot{\alpha}K + \dot{\alpha}L, \]

in which \( \alpha_k \) and \( \alpha_l \) are the factor shares of capital and labor, respectively, and the "hat" indicates a time rate of growth: \( \dot{\alpha} = d\ln x/dt. \) A relation which is equivalent to (1.2) is

\[ \dot{A} = Q - \dot{\alpha}K + \dot{\alpha}L. \]

This formalizes the "total factor productivity" view of technical change as the difference between the growth rate of output and the growth rate in a Divisia index of the two inputs, i.e., the "total factor input,"

It is clear from the assumption of competitive input and output markets that \( \alpha_k \) and \( \alpha_l \) sum to one. As a result, Solow defined new variables to transform (1.2) to

\[ \dot{Q} = \dot{\alpha}K + \dot{\alpha}L. \]

where \( Q/L \) is output per man hour and \( K/L \) is capital per man hour. From time series data on \( \dot{K} \) and \( \dot{L} \) the discrete approximations of \( A \) were computed. Solow's conclusion from his computations was an astounding one: for the period 1909–1949 "gross output per man hour doubled . . . with 87½ percent of the increase attributable to technical change and the remaining 12½ percent to increased use of capital" (Solow, 1957, p. 320). As for the average annual rate of technical change, which he assessed as "neutral" overall, Solow's computations revealed estimates of roughly 1½% per year for the first half of his sample period and 2½% per year for the second half.2

With data collection procedures becoming ever more sophisticated, as well as data being disaggregated down at least as far as the 2-digit SIC level,3 results such as Solow's which ascribe such a large portion of the growth in some measures of real output to the effects of technical progress are no longer the norm. Even Solow's general method for measuring technical change slipped into the background until it was "dusted off" by Baltagi and Griffin (1988). However, the 1957 Solow study did lead to numerous empirical investigations of technical change at the aggregate level over the succeeding 30–40 years. These papers attempted to identify technical change with one of three or four alternative definitions of "neutral technical change," to be discussed below. At the same time, researchers were giving attention to an emerging problem: the separate estimation of technical change and scale effects.

B. Notions of "Neutral" Technical Change and Bias

Over fifty years, Hicks provided both a policy-oriented rationale for an interest in "economic progress" (including the "theory of invention"), as well as an initial taxonomy for technical progress upon which subsequent writers have built. The substantive question concerning distribution which was posed by Hicks (1963, p. 113) was: "Is economic progress likely to raise or lower the proportion of the National Dividend which goes to labour?" Hicks showed that the answer to the question depends on whether labor's "proportion of the National Dividend" is interpreted as an absolute share (i.e. the real income of labor) or a relative share (i.e. the fraction of aggregate output accruing to labor).

With regard to the "absolute" share, the answer is obvious. If technical progress increases the supply of a factor, then the factor's absolute share will increase if the demand for the factor is elastic. Regarding relative shares, however, the answer to Hicks' question was not so obvious, at least at that time. The proposition which provided the answer to the question about relative shares made use of a new concept, the elasticity of [factor] substitution. From Hicks (1963, p. 117): "An increase in the supply of any factor will increase its relative share ... if its 'elasticity of substitution' is greater than unity." This significant result has been derived numerous times over the years. Many comparative statics terms which are used to measure the effects of technical progress are usually expressed in a form containing the quantity \( 1 - \sigma \) as a factor, where \( \sigma \) is the elasticity of substitution.4 For early examples of this, see Kendrick and Soto (1963, p. 990), Nadir (1970, p. 1143). Also, see Link (1987, p. 19).

In addition to providing a real motivation for the study of technical progress, Hicks also introduced the beginnings of a classification scheme for technical progress. "If we concentrate on two groups of factors, 'labor' and 'capital,' and suppose them to exhaust the list, then we can classify inventions according as their initial effects are to increase, leave unchanged, or diminish the ratio of the marginal product of capital to that of labor. We may call these inventions 'labor-saving,' 'neutral,' and 'capital-saving' respectively" (1963, p. 121). This defined a standard question for most future empirical studies of technical progress: Was technical progress neutral (in the sense of Hicks)? It was the notion of neutrality considered by Solow (1957, p. 316): "I have defined neutrality to mean that the shifts were pure scale changes, leaving marginal rates of substitution unchanged at given capital/labor ratios." Hicks' notion of neutrality has not been the only one considered over the years, however. Several others have been defined and tested empirically, and it is these that we turn to.

The three most popular types of "neutral technical progress" are Hicks', Harrod-, and Solow-neutral technical progress. These are defined more rigorously elsewhere, but the following characterizations will suffice for the purposes of the present paper. We have already seen that technical progress is considered Hicks-neutral if the marginal rate of substitution is left unchanged at a constant capital/labor ratio, i.e. the marginal rate of substitution is
"independent of technical progress" (Sato and Beckmann, 1965). Technical progress is 
Harrod-neutral if the output/capital ratio is left unchanged at a constant rate of return 
to capital; i.e. the rate of return to capital depends only on the output/capital ratio. The third type 
of neutral technical progress which has been noted with great frequency is Solow-neutral 
technical progress, which is defined as the case when the output/labor ratio is left unchanged at 
a constant wage rate; i.e. the wage rate for labor depends only on the output/labor ratio. 

Let us restrict our attention for the present to a production process with only two factors of 
production, capital K and labor L, used to produce a single output Y according to a production 
function F which is linearly homogeneous in (K, L); Y = F(K, L), where t is a variable 
representing the state of technology. (Changes in t are equivalent to technical change 
occurring.) Then a well-known set of results is that Hicks-neutral technical progress can be 
represented by the following form of the production function,

\[ F(K, L, t) = A(t)K^aL^b, \]

where A is a positive "augmenting" function, Harrod-neutral technical progress is equivalent to 
a "labor-augmenting" form of the production function,

\[ F(K, L) = F(K, B(t)L), \]

and Solow-neutral technical progress is equivalent to a "capital-augmenting" form of the production function,

\[ F(K, L) = F(A(t)K, B(t)L). \]

Clearly (1.5), (1.6), and (1.7) are special cases of the general representation of "factor 
augmenting technical change":

\[ F(K, L, t) = f(A(t)K, B(t)L), \]

in which we need not restrict f to be linearly homogeneous. Obviously (1.5) results when \( A(t) = B(t) \) and f is linearly homogeneous; (1.6) results when \( A(t) = 1 \) for all t; and (1.7) results when \( B(t) = 1 \) for all t. Let us remark here that Eichhorn (1978, p. 25) calls technical change "output 
augmenting" if it can be represented by the general form in (1.5), even if f is not linearly 
homogeneous. If f is linearly homogeneous, however, then technical change is said to be 
Hicks-neutral. Similarly, technical change is called "labor augmenting" if it can be represented 
by the general form in (1.6), even if f is not linearly homogeneous. When f is linearly 
homogeneous, then technical change is said to be Harrod-neutral. Finally, the form of technical 
change in (1.7) is generally "capital augmenting," but when f is linearly homogeneous, then 
technical change is said to be Solow-neutral.1

Other kinds of neutrality have been proposed and tested, and in some cases they have had 
greater explanatory power for actual time series data than Hicks', Harrod-, and Solow- 
neutrality. For these other notions of neutrality as well as some empirical tests, the reader 
is referred to Sato and Beckmann (1968), Beckmann and Sato (1969), and Chambers (1988).

While neutrality of technical change is a useful point of reference theoretically, 
we encounter it empirically as a statistical improbability. Whether we consider marginal rates of 
substitution, output/labor ratios, or something else in its relation to the capital/labor ratio, the 
unit price of labor, or some other relevant variable, technical progress is not likely to leave the 
relation completely invariant. For this reason, the notion of the "bias of technical change" has 
been derived. Hicks termed technical change labor-saving if it decreased the marginal rate of 
substitution at a given capital/labor ratio. (Equivalently we could call this capital-using 
technical progress if there are only two factors.) This non-neutral technical progress is therefore 
"biased" toward, or away from, the usage of one of the factors.

The "bias" of technical change has been defined in many ways. "The Hicksian definition 
measures the bias along a constant capital-labor ratio; the Harrodian definition measures the bias 
along a constant capital-output ratio; and Solow's definition measures the bias along a 
constant labor-output ratio" (Nadiri, 1970, p. 1142). As an example, consider the empirical 
result of Yajima (1974). He found—using the same data as Sato (1970)—that technical 
progress in the U.S. for the period 1960-1960 had neither Hicks- nor Harrod-neutral. His 
estimate suggested "Solow labor-saving" technical progress (equivalently "Solow capital-
using"), meaning that at a fixed wage the output/labor ratio would rise. Alternatively, 
technical progress would be Solow capital-saving (Solow labor-using) if at a fixed wage rate the 
output/labor ratio fell. Technical progress would be Harrod labor-saving (Harrod capital-
using) if at a fixed rate of return to capital the output/capital ratio were to fall, whereas Harrod 
capital-saving (Harrod labor-using) technical progress is the case when the output/capital ratio 
rises with a fixed real rate of return to capital.

The preceding paragraphs have described the most frequently considered notions of 
technical bias, and these are relevant for the many studies which have estimated A and B in 
(1.8), frequently along with the elasticity of substitution. Since these types of neutrality tend 
to be well-defined only in the analysis of production with two factors, for 40 years following 
Hicks' initial classification scheme this was sufficient. Empirical researchers only distinguished 
between capital and labor inputs. Since Binswanger's (1974a, 1974b) studies, however, 
empirical analyses of production and technical change have considered models based on more 
than two factors of production. In this setting the traditional concepts of technical bias are 
more generally relevant, and a new definition of bias has been needed. If we let \( a_i \) be the share 
of factor i \( (i = 1, 2, \ldots, n) \), then there is a measure of technical bias, \( B_i \) for each factor. From 
Binswanger (1974a, 1974b) or Chambers (1988):

\[ B_i = \frac{a_i}{n}, \quad i = 1, 2, \ldots, n. \]

The shares may be cost shares to allow for production under non-constant returns to scale 
(Chambers, 1988), or the cost shares may be allowed to collapse to output (revenue) shares 
under assumptions of competitive markets (constant returns to scale in a long run equilibrium). 
The variable \( t \) is generally interpreted as time, but more generally it can be seen as an index of the 
state of technology. The classification scheme for technical change based on (1.9) is quite 
straightforward: technical change is "factor i-saving" if \( B_i < 0 \), "factor i-neutral" if \( B_i = 0 \), and 
"factor i-using" if \( B_i > 0 \). Since their definition forces the shares to sum to one whether returns 
to scale are assumed constant or not, the biases over all factors must sum to zero. Hence all 
bias cannot be positive, nor can they all be negative.2 Again calling on a result from the 
empirical literature, Jorgenson, Gollop, and Fraumeni (1987, p. 25) found that "the patterns of 
productivity growth that occurs most frequently among the forty-five industries is labor-saving in 
combination with capital-using and intermediate-input-using productivity growth." (Their 
study period was 1948-1979, and the data were U.S. sectoral-level.) This means that the 
shares of output accruing to capital and intermediate-input have been rising "at the expense of 
labor," one might say. While the finding of widespread labor-saving technical change supports 
Hicks' hypothesis, the Jorgenson, Gollop and Fraumeni study and numerous others like it 
in motivation are suggestive of a "rediscovery" of the distributional issues which concerned Hicks. 
For "at the root of most definitions of "biasness" is an interest in the consequences of technical
change for different sectors of society." (Chambers, 1988, p. 203). The directions (one for each factor) of the biases of technical change are potentially quite important for making economic policy, and so also is the source of the bias(s). In the next section we consider how the source of a technological bias may be misidentified. This will be demonstrated as a particular manifestation of a larger problem, namely the separate estimation of scale effects and technical change effects.

C. Measuring Technical Bases, Substitution Parameters, and Scale Effects

In all empirical studies of technical change, researchers have attempted to estimate some or all of: technological biases, i.e. deviations from a particular kind of neutrality; parameters which attempt to measure the substitution possibilities between pairs of inputs; and finally the effects of scale. In recent years these studies have been carried out in a dual framework of prices. This avoids the simultaneity problem which plagues quantity-based econometric models such as those carried out in the 1960s and early 1970s. In order to highlight the issues here, consider a typical single-output framework under constant returns to scale. A unit output-price function, \( p(w, t) \), is given a translog structure in positive factor prices \( w_1, w_2, \ldots, w_n \) and time \( t \):

\[
\ln p (w, t) = a_0 + \sum_{i} a_i \ln w_i + \alpha_1 + \frac{1}{2} \sum_{i} \beta_{ii} \ln w_i \ln w_i
\]

\( + \sum_{i} \beta_{it} \ln w_i + \frac{1}{2} \beta_{tt} t^2, \tag{1.10} \)

where the parameters satisfy the usual conditions for symmetry, convexity and homogeneity. (See, e.g. Jorgenson, Gollop and Fraumeni, 1987.) From an assumption of competitive input markets, Shephard’s Lemma is employed to derive the cost share (\( \tau_i \)) equations:

\[
\tau_i (w, t) = -\frac{\partial \ln p}{\partial \ln w_i} = -a_i + \sum_{j} \beta_{ij} \ln w_j + \beta_{it}, \quad i = 1, 2, \ldots, n. \tag{1.11} \]

Letting \( \tau_i \) denote the rate of technical change, from (1.10) we derive the additional equation:

\[-\tau_i (w, t) = -\frac{\partial \ln p}{\partial t} = -a_t + \sum_{j} \beta_{ij} \ln w_j + \beta_{it}. \tag{1.12} \]

This indicates that the negative of the rate of technical change is given by the growth rate of the unit output price when all factor prices remain fixed.\(^{20}\) In the translog model, the biases of technical change are measured by the constant parameters \( \beta_{ii} = -\tau_i / \alpha_i (i = 1, 2, \ldots, n) \). From joint estimation of an equation system including \( n - 1 \) of the cost shares and the equation for \(-\tau_i\), one concludes that technical change is factor i-using, neutral, or saving as the estimate of \( \beta_{ii} \) is positive, zero, or negative, respectively. (See Jorgenson et al., 1987.)

This approach is enlightening as far as it goes, but suppose technical change actually affects the productivity of each input. This alters the "effective price" of each factor so that productivity growth (decline) is captured by a fall (rise) in the appropriate factor’s effective unit price.\(^{21}\) As a simple example, consider the case in which all effective factor prices change at a constant exponential rate over time: if \( \omega \) is the effective unit price of factor t, then \( w_t = e^{\omega t}w_0 \) where \( \omega \) is a real constant and \( w_0 \) is the observed unit factor price (i = 1, 2, \ldots, n). We would term this "price-augmenting" technical change, the dual analogue of factor-augmenting technical change.\(^{11}\) If we replace \( w_t \) in (1.10) with \( w_t = e^{\omega}w_0 \), and rearrange, we get:

\[
\ln p (w, t) = a_0 + \sum_{i} a_i \ln w_i + (\alpha_1 + \sum_{i} a_i) t
\]

\[
+ \frac{1}{2} \sum_{i} \beta_{ii} \ln w_i \ln w_i + \sum_{i} (\beta_{it} + \sum_{j} \beta_{ij}) t \ln w_i
\]

\[
+ \frac{1}{2} \beta_{tt} t^2 + \sum_{i} \beta_{it} \ln w_i + \frac{1}{2} \beta_{tt} t^2, \tag{1.13} \]

from which the biases of technical change are found to be \( \beta_{ii} / \alpha_i = -\omega^2 \ln p / \alpha_i \ln w_i - \beta_{it} + \sum_{j} \beta_{ij} \) (i = 1, 2, \ldots, n). That is, the biases of technical change are given by \( \beta_{ii} \) plus a weighted sum of the constant growth rates of the effective factor prices (\( \omega_i \)) where the weights are the \( \beta_{ij} \).

If technical change is factor i-saving, say, the declining share \( \omega_i \) can result from declining productivity of factor i (\( \eta_i > 0 \), i.e. a rising effective price \( w_i \)) causing the term \( \beta_{ii} \) to be negative (since \( \beta_{ii} \) must be nonpositive for concavity of \( p \) in \( w_i \)). However, factor i’s share will be declining even with improving productivity (\( \eta_i < 0 \)) if \( \beta_{ii} + \sum_{j} \beta_{ij} \) is negative and greater in absolute value than the nonnegative term \( \beta_{ii} \). Hence from a policy-making perspective, the cause of a particular technological bias is of great importance, for a factor’s declining share does not necessarily indicate declining productivity of the factor.

While the example here is quite simple, it also convincingly supports Stevenson’s (1980, p. 162) observation that the direction of bias is easier to determine than to explain, for although the signs of \( \beta_{ii} / \alpha_i \) (i = 1, 2, \ldots, n) can be determined, separate estimation of all the \( \alpha_i \) in (1.13) is not possible.

With the benefit of a perspective using the theory of Lie transformation groups (Sato, 1980, 1981), it is now easier to understand why there have been so few investigations into the nature of technical change, rather than a simple identification and interpretation of its effects. The separation of technical change effects from the effects of scale is a major theoretical problem, and the notion of "heterolocality" is a way to describe this "identifiability problem." The "heterolocality of a technology under a particular type of technical change" is the situation in which the effect of technical change, operating on the productivities of the inputs, is simply a relabeling of the isoquant map or of the price-space contours of the cost function. In the terminology of group theory, the underlying technology, represented by the production or cost function, is invariant under the transformation represented by technical change. In such a situation technical change effects are completely indistinguishable from the effects of scale because they are exactly the same as a scale effect. In the econometric model of Eqs. (1.10)-(1.12), heterolocality will result if all of the \( \alpha_i \) are equal: \( \alpha_1 = \alpha_2 = \ldots = \alpha_n = \frac{1}{n} \). Then in p collapses to:

\[
\ln p (w, t) = a_0 + \sum_{i} a_i \ln w_i + (\alpha + \frac{1}{n}) t
\]

\[
+ \sum_{i} \beta_{it} \ln w_i + \frac{1}{2} \beta_{tt} t^2. \tag{1.14} \]

i.e., the form of \( \ln p \) in (1.10) plus \( \lambda \), since \( \lambda \) can be moved to the outside of the summations and \( \sum_{i} a_i = 1 \) and \( \sum_{i} \beta_{ii} = \sum_{i} \beta_{it} = \sum_{i} \beta_{ij} = 0 \) from the homogeneity of \( p(w, t) \) in \( w_i \). Hence the technical change effect is indistinguishable from a scale factor \( \omega \) in the price function \( p(w, t) \).

This identifiability problem is not a recently recognized phenomenon; it was known over a
quarter century ago to Solow (1961, p. 67): "The problem of measuring economies of scale and distinguishing their effects from those of technical progress is an econometric puzzle worthy of anybody's talents." One of the contributions of group theory toward solving this puzzle is seen in a formal theoretical model of exogenous technical change (Sato, 1980, 1981) in which we can clearly see the conditions present when scale effects and the effects of technical change are indistinguishable. In addition, it is apparent that holothecity has been a problem historically because researchers have primarily used only factor-augmenting types of technical change, and the factor-augmenting type is a "bad" hypothesis if scale effects are to be distinguished from technical change effects when the production function is homogeneous, as is often assumed."

In recent years empirical researchers have utilized duality theory and cost functions to analyze producer behavior and technical change. In this context "price-augmenting" technical change is a poor hypothesis because of the homogeneity of the rational firm's cost function in factor prices; this can be seen in the impossibility of estimating the $a$s in (1.13). Then it is evident that another contribution from group theory is the presentation of new types of technical change functions which will not lead to the extreme identifiability problem which the factor- and price-augmenting types produce. The structure given to these new types is the structure of Lie groups of transformations.

Indeed a recent result (Mitchell, 1989) gives a functional representation to an effective factor price vector of arbitrary dimension $n$, satisfying the group properties and three additional regularity conditions. The reader is referred to Mitchell (1989) for the specific conditions in the theorem, but for the purposes of this survey we will present only the sufficient conditions. Choose an arbitrary function $\psi: R^n \to R$ under the condition that $\psi(z) - \psi(0)$ can be solved uniquely for $z = \psi'(y) e R^n$. Choose also an arbitrary, scalar-valued, strictly monotonic function, $\Gamma$, of a scalar argument. Further suppose that $\Gamma'$ is defined at the origin, $-\Gamma'(0)$ for all positive $z$, and has a range containing the range of $\psi$. Then define functions $H(z)$ and $h(z)$, where $z \in R^n$, in order to partition $\psi$ into the components $H$ (homogeneous) and $h$ (heterogeneous), the dimensions of $H$ and $h$ are one and $n - 1$, respectively. Form $\phi: R^n \times R \to R$, according to

$$
\phi(w, t) = \psi^{-1}(H(w) + \Gamma^{-1}(t))
$$

Then the definition of $\phi$ in (1.15) defines a transformation which satisfies the intuitively appealing properties of a group of transformations (Sato, 1980, 1981; Mitchell, 1984, 1987, 1989), and these functions may be used to define effective factor prices $w$ as functions of nominal (observable) factor prices $w$ and the technology index $\Gamma$ (possibly time), and substituted into an econometric model of producer behavior and technical change such as Eqs. (1.10)-(1.12). Examples derived from (1.15) have been presented (Mitchell, 1989), but we will reproduce one here to illustrate how useful the above result is.

Let $n = 4$; one might think of identifying the four factors of capital, labor, energy, and intermediate goods (cf. Berndt and Wood, 1975; Jorgenson and Fraumani, 1981). Let $F(t) = 1$, $t \in R$, and

$$
\phi(z) = \left( \begin{array}{c}
\ln z_1 \\
3z_2 - \frac{a_2}{a_1} \ln z_1 \\
\frac{a_3}{a_1} z_2 - \frac{a_3}{a_1} \ln z_1 \\
z_4
\end{array} \right), \quad z_i > 0
$$

from which we find

$$
\phi(w, t) = \psi^{-1}(H(w) + \Gamma^{-1}(t)) = \left( \begin{array}{c}
d_1 \\
w_1 + a_2 d_2 \\
w_2 + a_3 w_3 d_3 \\
w_4
\end{array} \right)
$$

Here the technical change is of the (exponential) price-augmenting type for the first price, the additive type for the second; the "ratio additive" type for the third (see Sato, 1981); and there is no productivity change for the fourth factor. Now it can be seen that there are many different types of transformations, $\phi(w, t)$, where one combines the basic types—augmenting, additive, ratio additive, "none," etc.—in various ways through the price vector $w$ using (1.15).

One obvious advantage to technical change functions such as (1.17), of course, is that by avoiding multiplicative forms, $w_i = A_i w_i$, for all $i$, one avoids the circumstances surrounding an impossibility theorem such as those due to Diamond et al. (1978). As a corollary to their "non-identification theorem," Diamond, McFadden, and Rodriguez (DMR) (1978, p. 135) noted that "one cannot obtain a definitive acceptance of the factor augmentation hypothesis [in the multiplicative case of (1.8)] from aggregate data: these data could have been generated by a neoclassical production function with an arbitrary elasticity of substitution on the observed path, and this elasticity could always be chosen so that the resulting production function is not factor augmenting." Hence, the DMR results indicate that it will not always be possible to estimate $A, B$, and $c$ in a model of production described by (1.8). DMR pointed out (p. 136) that the multiplicative factor augmentation hypothesis does imply certain restrictions that can be interpreted as "necessary conditions for the consistency of observations with the factor augmentation hypothesis." And although these restrictions can be rejected by the data, the fact remains that one may have data consistent with these necessary conditions even though the data are not generated by a process described by (1.8).

While the DMR Impossibility theorem has been circumvented in some studies (cf. Berndt and Khaled, 1979; Baltagi and Griffin, 1988), an alternative approach which has not been fully explored is the substitution of technical change functions, such as may be generated by using (1.15), into an econometric model like (1.10), (1.11), and (1.12). While this contribution, like that of David and van de Klundert (1965, p. 359), "will be confined simply to establishing the form taken by factor-efficiency growth . . . rather than identifying the sources from which it flows," we feel that it is a logical "next step" after identifying the effects of the efficiency growth, as measured by movements over time in relative factor shares.
II. THE ECONOMIC ANALYSIS OF ENDOGENOUS TECHNICAL CHANGE

A. The Basic Model

Technical progress is not necessarily the only reason for apparent productivity gains in any society, but it is generally accepted as a significant source of productivity gains. Practically all types of technical change are determined within the economic system as the result of meaningful rational behavior of economic agents; that is to say, technical change is "endogenous."

An important dimension of technical change is that it is the outcome of investment-type activities. For example, the firm reinvests a fraction of present profits in research and development (R&D) for the purpose of enjoying future cost-saving benefits via improvements in the efficiency of its production. Studies done by industrial organization specialists have shown that the R&D cost relation to output follows the same (U-shaped) pattern as the cost relation of other factor inputs. This pattern indicates that R&D can be viewed as one of the factors of production, and that we may apply the production function approach to the theory of endogenous technical progress.

A pioneering attempt to deal with the problem of endogenous technical progress is the work of Kamien and Schwartz (1969). They proved that under certain conditions the technical change of the firm will asymptotically approach Hicks-neutral if the elasticity of factor substitution is less than one. A model developed by Sato and Ramamurthy (1974, 1987) showed that technical progress within a firm can be optimally generated within the framework of increasing output demand and of factor prices increasing at a constant rate. Since this model is the forerunner of the other dynamic models developed later, let us briefly review the Sato-Ramamurthy Model. (A list of symbols used in this part is given in the Appendix).

The model assumes a linearly homogeneous production function and Hicks neutral type of technical change:11

\[ Y(t) = F(T(t), K(t), L(t)) = T(t)F(K(t), L(t)), \]

where \( Y(t) \) = output, \( K(t) \) = capital input, \( L(t) \) = labor input, and \( T(t) \) = index of productivity improvement, or uniform factor augmentation (see Eq. (1.5)). The firm's total cost of production is expressed as

\[ C(t) = Y(t) \cdot \frac{G(t)}{T(t)} \frac{w(t)}{w(t)} \frac{r(t)}{r(t)}, \]

where \( w(t) \) and \( r(t) \) represent the unit prices of labor and capital, respectively. Factor prices are considered to be increasing at percentage rate \( g \); that is

\[ w(t) = \omega e^{g}, \quad r(t) = r e^{g}. \]

Since the cost function is homogeneous, the cost function can be alternatively expressed as

\[ C(t) = Y(t) \cdot \frac{G(t)}{G(t)} \frac{w(t)}{w(t)} \frac{r(t)}{r(t)}, \]

where \( g(t) = T(t)e^{g} \) and \( G(t) = G(t, w), G(t, r) \) constant. The firm's demand curve shifts outward continuously over time. The demand price is a function of the firm's output and a parameter \( Z \), which represents the aggregate output of the economy and serves as a shift parameter in the

\[ P(t) - P(Y(t), Z) = P[Y(t)], \]

where \( Y(t) = Y/Z \).

The firm is assumed to allocate a certain percentage, \( c \), of sales to R&D, that is

\[ R&D \, \text{expenditure} = dP(Y(t)), \]

where \( d \) is a decision variable. The firm's technical progress function is represented by

\[ \frac{\text{d}P(Y(t))}{Z} = h(c), \]

The technical progress function or \( h(\cdot) \) function, is analogous to a conventional production function, and in that respect, it manifests the standard neoclassical properties of positive "marginal product" (\( b > 0 \)) and "diminishing returns" to R&D expenditures (\( b < 0 \)). This conventional "concavity condition" is illustrated in Figure 1.

From an economic perspective, the firm's problem is to determine the optimal pattern of R&D expenditures and output levels over time so as to maximize its discounted long-run profits. This situation is expressed as a dynamic variational problem of the following sort:

\[
\max_{\text{t}, e^{g}} \int_{0}^{\infty} \left[ (1 - \delta) P(Y(t) - h(c) e^{g} Y(t) \frac{G(t)}{Z}) dt \right.
\]

subject to:

\[ \dot{h} = g(\text{d}) \left( \frac{P(Y(t))}{Z} - g \right). \]

The term in the brackets of the objective function (the expression to be maximized) is the firm's expected net-profit function at any given moment in time. The expected future stream of net profits is discounted at the rate \( \delta \) to account for time preference on the part of the owners of the firm. A complete analysis of this dynamic optimization problem will not be presented here. However, a brief description of a stable solution to the problem is given in Figure 2.

Figure 2 illustrates the combinations of productivity index values and Lagrange multiplier

\[ \lambda(c) \]

\[ \text{R&D Expenditures (output)} \]

Figure 1. Technical Progress Function
knowledge via basic and applied research. Basic knowledge is considered to be an "intermediate product" in the production of new applied, or technical, knowledge. We formally introduce the behavioral equations

\begin{align}
\dot{A} &= h(a, b, \beta) \cdot A \\
\dot{B} &= f(a, b, \mu) - \mu B
\end{align}

where \( \mu > 0 \) is a depreciation factor in basic knowledge. It takes account of the fact that a part of the production effort of knowledge is aimed at renewing and transferring knowledge. One may think of the beginning of the Middle Ages as the time when \( f(a, b) \) was near zero, and so the level of basic knowledge deteriorated. An increase in the stock of basic knowledge or applied knowledge depends on the "technical progress" production functions, \( f \) and \( h \). Thus the rate of change (flow) of basic knowledge \( (dA/dt - \dot{A}) \) is positively related to the specialized research workers, \( a \), employed to produce basic research, and to the research capital, \( b \), but that rate is negatively related to the depreciation rate, \( \mu \). In the same way the rate of change (flow) of applied knowledge \( (dA/dt - \dot{A}) \) depends on the specialized researchers of workers, \( a \), and research capital, \( b \). It is assumed here that the "technical progress" production functions satisfy the concavity condition and other usual regularity conditions (illustrated in Figure 1).

**Basic Knowledge**

The main feature of the model is that basic knowledge is assumed to be essential for the production of technical knowledge. Looking at the extreme case in which the stock of basic knowledge is zero (a case that may exist in principle but not in practice), we would find that the following conditions would hold:

\begin{align}
(i) \quad h(a, b, 0) &= 0 \\
(ii) \quad \frac{\partial h}{\partial B}(a, b, 0) &= +\infty
\end{align}

Equation (II.9) states that (i) production of applied knowledge is impossible without some amount of basic knowledge and (ii) the marginal productivity of basic knowledge is extremely high as the stock of basic knowledge decreases to zero.

The first expression states that without basic knowledge the production of applied knowledge is impossible. The assumption is quite realistic. Elementary language and numerical skills fall under the category of basic knowledge. It is clear, for example, that a student with no knowledge of Russian could not possibly formulate an intelligent interpretation of the original works of Tolstoy. Similarly, a researcher who has no basic knowledge of a certain computer language could not possibly produce useful programs written in that language.

The second expression states that the marginal productivity of the first unit of basic knowledge in the production of technical knowledge is extremely large. As an approximation, one may envisage the change in applied knowledge that is brought about by mastering simple arithmetic.

The firm is faced with a discount rate, \( \rho \); time series of expected prices of \( x \) and \( b \); \( P_0(t) \) and \( P_1(t) \); time series of expected prices of capital and labor, \( r(t) \) and \( w(t) \); and the demand function, \( P(X) \), which is also subject to dynamic change. The firm may naturally expect the increasing scarcity of resources to be reflected in knowledge, thus increasing its productivity and reducing its cost. We will see below that this indeed happens.
Optimization Behavior

We assume that the firm produces the output, \( Y \), from \( K \) and \( L \), using its neoclassical production function

\[
Y(t) = F(A(t)K(t), A(t)L(t)),
\]

where \( A(t) \) is the level of factor augmentation at time \( t \), and it is the measure of technical progress of the firm.

From the production function we can derive the total cost function

\[
TC(Y, w, r, A) = \frac{1}{A(t)} \overline{TC}(Y, w, r),
\]

where \( \overline{TC}(Y, w, r) \) is the minimum cost of producing \( Y \), given the production function with no factor augmentation. This property is valid for a general production function. We do not assume any homogeneity (see Stato, 1981, Chapter 6). The firm’s objective is to choose the appropriate amounts of output \( Y \) and appropriate flows of basic and applied knowledge in such a way that the long-run profit is maximized. One extreme case would be that in which the firm makes no additional investment in the creation of technical progress but produces only output \( Y \). Generally, this policy is not optimal, because by investing in basic and applied knowledge, the firm will further reduce the cost of producing output \( Y \). As long as the cost saving exceeds the revenue increase due to more production of \( Y \), the firm also invests in technical progress ventures. This trade-off relationship can be more precisely studied by the following formulation.

The firm allocates resources in order to:

\[
\text{maximize} \quad \text{[Long-run profit] = Revenue - Production cost - Research cost}
\]

under the dynamic constraints. More specifically, the firm’s objective is to solve the maximization problem of the discounted long-run net profit function:

\[
\max_{A(t), \theta(t), Y(t)} \int_{t_0}^{\infty} e^{-\theta t} \left[ Y(t)P(t) - \frac{TC(Y(t), w(t), \theta(t))}{A(t)} \right] dt
\]

subject to the technical progress functions of applied and basic research respectively, that is

\[
\dot{A} = b(a_A, a_B, B) - A
\]

\[
\dot{B} = \nu(a_A, a_B, \nu) - \alpha B
\]

and the initial conditions.

In making decisions about the choice of physical output and research outputs, the firm must know the future course of the price of output, \( Y \), and input prices. Here we assume that the firm’s vision for the future course of those prices is based on the so-called rational expectation hypotheses. More specifically, we assume that the commodity price and the factor input prices increase at certain constant rates:

\[
\begin{align*}
(1) \quad P_t(Y(t)) & = e^\rho P(Y(t)) \\
(2) \quad \tau(t) & = e^{\beta t} \\
(3) \quad w(t) & = e^{\alpha t} \\
(4) \quad P_t(L) & = e^{\rho}P(L) \\
(5) \quad P_t(K) & = e^{\rho}P(K)
\end{align*}
\]

Here we assume that the commodity price, \( P(Y) \), is increasing at the same rate, \( \alpha \), as the prices of inputs in the R & D sector, whereas the wage rate and the return to capital both grow at rate \( \beta \).

Using these assumptions, we now consider the technical-progression index expressed in real (net) terms. We call \( g(t) \) the technical progress index in real terms, defined by

\[
g(t) = \frac{\theta(t)}{e^{\beta t}} \cdot A(t)
\]

Then \( g(t) \) is the nominal index of technical progress, \( A(t) \), divided by the difference between the index of the regular factor input prices, \( e^{\alpha t} \), and the index of the output price, \( e^{\beta t} \). The term \( g(t) \) measures the real effect of technical change—real in the sense that the nominal rate is divided by the net inflation rate \( (\beta - \alpha) \). We shall call \( g(t) \) the “real net index of technical progress.”

Using the real net index of technical progress, we can express the total cost function as

\[
TC = \frac{TTC(Y, w, r)}{A(t)} - \frac{1}{A(t)} \overline{TC}(Y, w, r) \quad \text{(by homogeneity)}
\]

\[
\begin{align*}
\cdot \quad TC & = \frac{1}{\frac{A}{\overline{TC}(Y, w, r)}} \\
& = e^{\frac{\theta(t)}{\beta - \alpha}} C(Y)
\end{align*}
\]

where \( C(Y) = \overline{TC}(Y, w, r) \).

Also using the expectation hypotheses, we can express the total revenue function as

\[
\begin{align*}
(1) \quad Y(t)[P(Y(t)) & = e^{\rho}Y(t)\overline{P}(Y(t)) \\
(2) \quad Y(t)[\tau(t) & = e^{\beta}Y(t) \\
(3) \quad Y(t)[w(t) & = e^{\alpha}Y(t) \\
(4) \quad Y(t)[P_t(L) & = e^{\rho}Y(t) \\
(5) \quad Y(t)[P_t(K) & = e^{\rho}Y(t)]
\end{align*}
\]

where \( R[Y(t)] = Y(t)[\overline{P}(Y(t)] \).

Equation (1.14) obviously implies that the percentage change in the nominal index of technical change is equal to the sum of the percentage change in the real net index and the net inflation rate, that is,

\[
(1.14a) \quad \frac{\Delta A}{A} = \frac{\Delta \frac{A}{\overline{TC}(Y, w, r)}}{\overline{TC}(Y, w, r)} + (\beta - \alpha)
\]

Using these new variables, we can reformulate our maximization problem of optimal endoge-


and the real (gross) rate of interest \( \rho + \alpha - \alpha \). Here the value of the real index of (applied) technical progress is measured in terms of the marginal product of basic knowledge. Equation (II.18) states that the marginal revenue of producing the real index of technical change, \( g \), must in the long-run be equal to the value of its supply cost multiplied by the real rate of interest \( \rho + \alpha \). This is the direct interpretation of (II.18) using (II.20), which states that the percentage change in the current stock of the applied knowledge, \( \beta - \alpha \), is constant and equal to the percentage increase in the real wage of the labor force employed in the production of output, \( \beta - \alpha \). Equation (II.19) is the direct consequence of the fact that the additional flow of basic research just offsets the depreciation of the basic knowledge.

We also note from equation (II.20) that the current cost of maintaining a fixed positive amount of \( g \) is independent of \( g \) itself. The cost is the solution of

\[
\min_{P(a, b)} \ P(a, b) + \beta - \alpha
\]

subject to

\[
h(a, b, B) = \beta - \alpha
\]

and also nonnegativity of variables.

Policy Implications

The dynamic differentiated R&D model developed in this section gives us a complete conceptual and technical answer to the question: What is the value of additions to our stock of knowledge, basic or applied? The answer is generated in the process of solving: \( q_1 \) is the "implicit price" or "supply price" or "opportunity cost" of the stock of basic knowledge, \( B \), and \( q_2 \) is the "implicit price" of applied knowledge, \( g \). Through \( q_2 \) we account for the effect of present increases in \( g \) on the discounted stream of profits in the future. Through \( q_1 \), we account for the effect of the change in \( g \) on the production of \( q_2 \) in all future periods, and from there on the forthcoming income. The relation in equilibrium between the two effects is given by (II.17):

\[
q_1 = \frac{q_2}{\rho + \mu - \alpha} \cdot \frac{d\theta}{d\theta}
\]

Other things being equal, if \( q_1 \) is proportional to \( g \), the higher the stock of technical knowledge, \( g \), the more basic knowledge is worth. The accumulated effect, expressed by \( q_2 \), and \( q_3 \), will be higher than the current effect of R&D on productivity, so in studies measuring only the current effect there is downward bias of the returns to R&D.

The distinction between basic and applied research is important for public policy decisions. Government agencies must often choose among projects with varying proportions of basic and applied research. Conceptually, the optimal ranking of competing projects must be made on the basis of such values as \( q_1 \) and \( q_2 \). Note how the parameters \( \rho, \mu, \) and \( \alpha \) affect this optimal ranking. The lower is \( \rho \) (i.e., the less firms discount the future), the higher is the implied price of basic knowledge. The same result obtains for \( \mu \), the rate of decay of basic knowledge. Note that the lower is the growth rate of output prices, the lower is this implied price of basic knowledge relative to that of applied knowledge. Thus, in sectors of the economy in which output prices are growing relatively slowly, applied research is relatively more important to the firm. These parameters are likely to differ from one industry to the next, and equation (II.17) is therefore a
valuable policy guide in making decisions about financing "basic knowledge" intensive research versus "applied" intensive research.

The model captures an aspect much discussed in the R&D literature, which is the returns-to-scale element. We may suppose that in our model only big firms will have sufficiently large markets to make a profitable feasible solution, $g$, possible. Once they have the stock of research, $g$, they will not stop accumulating technological knowledge until the marginal profit of $g$ is zero, which occurs at $g^*$. The model shows that a firm may protect itself successfully against a rise in factor price that is faster than the rise in the product's demand function by engaging in R&D. We may transfer the analysis and conclusion to the economy as a whole. By allocation of resources to R&D, we may deny the Malthusian hypothesis and prevent the conclusion of the Doomsday models. The developments in agriculture indeed follow this trend: Part of the growing demand for cereal was supplied by technological improvements.

C. "Optimal" Biased Technical Change

More than fifty years ago, Hicks succinctly stated the problem of induced biased technical progress in another oft-quoted passage:

"A change in the relative prices of the factors of production is itself a spur to invention, and to invention of a particular kind—directed to economizing the use of a factor which has become relatively expensive" (1963, p. 124).

We now assume that factor prices are increasing at different rates and that the monopolistic firm, through allocation of resources of research, can determine the "optimal" rate of augmentation of each factor.

Assumptions and Model

1. We shall assume that the firm is a monopolist with an infinite time horizon and perfect foresight. There is a vast literature on the impact of market structure on research efforts; in this paper we consider only the limiting case where the firm can fully internalize the benefits from its research.

2. The production function is linear homogeneous and can be written as

$$Y = A(t)X_1, B(t)X_2$$

where $X_1$ and $X_2$ are two inputs. For example, they can be taken to be the conventional inputs, labor and capital, or $X_1$ can be considered as a composite non-energy input and $X_2$ as an energy input. $A(t)$ and $B(t)$ are the levels of factor augmentation at time $t$.

3. Let $P_1$ and $P_2$ be the prices of the inputs $X_1$ and $X_2$. We assume that they increase at constant proportionate rates $\delta_1$ and $\delta_2$:

$$P_1 = P_1e^{\delta_1t}, P_2 = P_2e^{\delta_2t}.$$ (11.24)

4. The price of the product $P$ is determined by the demand curve $P = P(Y)$.

5. We shall assume that the firm is in existence and that it finances all its research internally. Thus we assume away the problem of entry-and-exit. As we are confining ourselves to the study of an internal steady-state, this assumption does not affect the basic result of the model.

6. The expenditure on research determines the rate of increase of $A(t)$ and $B(t)$ according to two technical progress functions. In empirical and industrial organization literature, there is a tradition of considering the proportion of sales revenue devoted to research as an index of the research intensity of the industry. We shall, therefore, express the research expenditures as proportions of total sales revenue, $PY$. Since $PY > 0$ by assumption 5, this particular formulation does not affect the results.

$$\dot{A} = h_1(\theta_1PY), \dot{B} = h_2(\theta_2PY), \theta_1 > 0 > h_2^*$$

where $\dot{A}$ and $\dot{B}$ are the rates of growth and $h_1^*$ and $h_2^*$ are the first and second order derivatives with respect to $\theta PY$.

7. Since the production function is linearly homogeneous, it is well known that the cost function can be written as:

$$C = YQ(\frac{P_1}{A(0)} - \frac{P_2}{B(0)}) = YV(\theta_1, \theta_2)$$

where

$$\frac{\partial V}{\partial \theta_1} = \frac{\partial V}{\partial \theta_2} = 0.$$

Further it can be easily shown that $G$ is linearly homogeneous in its arguments and that $\frac{\partial G}{\partial \theta} = 1$ in $\theta_1$ and $\theta_2$.

$$\frac{\partial G}{\partial \theta_1} + \frac{\partial G}{\partial \theta_2} = -1.$$

The object of the firm is to maximize the discounted value of its profit stream (revenue minus production and research costs):

$$\max_{\theta_1, \theta_2} \int e^{-\eta r}(1 - \theta_1 - \theta_2)PY - YV(\theta_1, \theta_2)dr$$

subject to

$$\dot{\theta}_1 = [h_1(\theta_1PY) - \omega_1] \theta_2,$$

$$\dot{\theta}_2 = [h_2(\theta_2PY) - \omega_2] \theta_2.$$

This problem is easily analyzed using the Maximum Principle of Pontryagin where we reduce it to the study of an autonomous Hamiltonian,

$$H = [(1 - \theta_1 - \theta_2)PY - YV] + \eta_1[h_1(\theta_1PY) - \omega_1] + \eta_2[h_2(\theta_2PY) - \omega_2]E.$$ (11.29)

Here $\dot{\theta}_1, \dot{\theta}_2$ and $Y$ are control variables whose values are set at the level at which they maximize $H$, while $\theta_1$ and $\theta_2$ are state variables whose values at any instant of time are given. The rates at which $\theta_1$ and $\theta_2$ vary over time are given by differential equations, while $\eta_1$ and $\eta_2$ are co-state variables and can be interpreted as shadow prices.

We may simply present the result of the analysis as The Hickian proposition:

Proposition: The rate of bias of endogenous technical progress generated by this firm will, in steady-state, equal the difference in the rates of growth of its input prices.
III. TECHNICAL PROGRESS, INFORMATION ASYMMETRY AND DYNAMIC COMPARATIVE ADVANTAGE

Countries that are leaders in technology, like the United States, Germany, France, and the United Kingdom, have generated scientific breakthroughs and innovations in the past through endogenously determined R&D activities. Latecomers like Japan, South Korea, and most of the developing countries, in contrast, have imported technology from the technological leaders. Technological leaders tend to invest a larger share of R&D funds in basic research. The latecomers, on the other hand, invest relatively more in applied R&D. First by imitation, then by improvements on the processes and products imported, the latecomers may gain a competitive edge in export markets as they learn to produce the same (or similar) goods at a lower cost and export them to the world market. This aspect of international competition (technology game) can be explained in terms of dynamic comparative advantage.

As Lyons (1987, p. 170) noted: “The idea of technology-based trade may be old, but a full investigation of the welfare consequences, and thus policy implications, has had to await progress in the technology of economic theory. We are still a long way from a comprehensive understanding, but at least the bounds of our ignorance have been reduced.” Among the contributions to this new technology one can mention Flaherty (1984), Krugman (1979), Reinganum (1984), Sato (1988), Sato and Tsutsui (1986), and Spencer and Brandt (1983).

Sato (1985, 1988) assumes that there are two monopolistic firms, one in each country. The firm in Country I undertakes both pure and applied research. Pure knowledge is produced by the input of specialized researchers, $L^I$, and research capital, $K^I$. It is taken to be depreciated at a given rate, $\rho$. Production of applied knowledge requires not only the input of skilled workers, $L^I$, and capital, $K^I$, but also a stock of pure knowledge, $B^I$. Applied knowledge also depreciates at a constant rate, given by $\gamma$. The flow of new technology does not depend on current efforts alone but is influenced by past investments due to the lag effects. So the model assumes that the gross increase in technical knowledge is a weighted average of past values of investments in research.

The firm in Country II does not engage in pure research but a part of the new knowledge generated by Firm I diffuses to it. The diffusion coefficient, the percentage of new pure knowledge that is available to Firm II, is a constant, $0 < \gamma < 1$, for which it may have to pay a royalty. Applied knowledge is produced by Firm II using scientists, $L^I$, capital, $K^I$, and the stock of basic knowledge that was diffused to it at time $t$. The production function of applied knowledge of Firm II is taken to differ from the corresponding function of Firm I by a multiplicative factor $\delta$ if $\delta > 1$. Firm II is less efficient.

The price of the product is a function of world output and time as the demand curve shifts over time. For simplicity it is assumed that at any given output, the shift in demand increases the output price at a constant proportionate rate, $\alpha$. Similarly the prices of inputs are taken to increase at a given proportionate rate, $\beta$ ($1 - I$). A higher level of applied knowledge corresponds to lower costs of production. The firm maximizes the discounted value of future profits and play a Cournot-Nash differential game in which each firm does the best it can given the other agent’s actions. Even then one must differentiate between open-loop and closed-loop strategies. In the open-loop strategy, the firm chooses a dynamic path on the assumption that the path of the other firm is given. In closed-loop strategy, the firm takes into consideration current information on rival’s strategies and variables.

Under these assumptions we now present the technology game. To simplify the notation, the “function of t” notation will be suppressed from the equations. The firm in Country I seeks to

$$\max_{I, K^I} \int_0^\infty e^{-\rho t} \left[ P(Y^I) Y^I - \frac{Y^I}{\theta} (P_L L^I + P_K K^I) - (1 - \theta)(P_L L^I + P_K K^I) \right] dt,$$

where $Y^I = Y^I + Y^R$, and $\theta$ is the share of the annual cost of developing basic knowledge financed by Country II ($0 < \theta < 1$), subject to the technological constraints

$$Y^I = \beta Y^I + \int_0^\infty e^{-\rho t} \gamma B^I (L^I, K^I, B^I) W^I(t - r) dr,$$

given the optimal paths of the variables determined by the firm in Country II. In (III.1) and (III.2), $W^I(t - r)$ is Country I’s weighting function at time $t$ (for past values of investments in research $i = A, B$) and $s$ is the social discount rate (assumed to be the same for both countries). Country II’s firms seeks to

$$\max_{I, K^I} \int_0^\infty e^{-\rho t} \left[ P(Y^I) Y^I - \frac{Y^I}{\theta} (P_L L^I + P_K K^I) - (1 - \theta)(P_L L^I + P_K K^I) \right] dt,$$

subject to the technological constraints

$$Y^I = \beta Y^I + \int_0^\infty e^{-\rho t} \gamma B^I (L^I, K^I, B^I) W^I(t - r) dr,$$

given the optimal paths of the variables determined by the firm in Country I. (Note the assumption $W^I(t - r)$.)

There are three crucial parameters in this differential game: $\delta$ — the relative efficiency parameter for applied technology, $\gamma$ — diffusion index of basic technology, and $\theta$ — the index of cost sharing of basic research. Their relative magnitudes will determine how the market evolves in the long run.

Assuming that the firms are in a steady-state, Sato (1988) has examined the effects of various parameters on the market share of each firm. He has done this by drawing an iso-share curve in $\kappa - \gamma$ parameter space where $\kappa = 1/A$. Figures 3 and 4 show the iso-share curve for open-loop and closed-loop strategies. As $\gamma$ increases, the iso-share curve rises showing that, even for smaller values of $\delta$ (higher values of $\kappa$), Firm II can maintain its market share. In the case of closed-loop strategy, in contrast to open-loop strategy, Firm II can have a higher share than Firm I even if $\delta > 1$. In other words, the diffusion of pure knowledge permits Firm II to compensate for the inefficiency in producing applied knowledge.

In Sato and Tsutsui (1986), one firm produces pure knowledge but it diffuses to many
firms (N firms), each producing applied knowledge. One must now define the diffusion coefficient and efficiency parameter for each firm but the paper considers a symmetric case where γ and k are the same for each of the N firms. They considered four cases relating to the information structure: (1) e-c case where all N + 1 firms follow a closed-loop strategy; (2) o-o case where all firms follow an open-loop strategy; (3) e-o case where the firm producing pure knowledge follows a closed-loop strategy and all others follow an open-loop strategy; and (4) o-e case where the producer of pure knowledge follows an open-loop strategy and all others follow a closed-loop strategy.

CONCLUSION

The theory of technological competition is a new branch of economics but, more than its infancy, it suffers from the intellectual incompatibility of its progenitors, the theory of international trade and the theory of innovation. Ever since the time of David Ricardo, international trade theory has been the bastion of lassiz-faire. Even when the need for a tariff or subsidy was conceded, the argument was that such policies should be non-discretionary and non-discriminatory.

But the theory of innovation, in contrast, recognizes the need for market structures that permit the firms to internalize the benefits from their R&D investments. As Caves (1987) pointed out, the industrial organization literature suffers from anglocentrism with the United States acting as a dominant firm. If market structures are to be explicitly considered, then it behooves us to recognize the international differences in the boundaries between firms and the market, variations in the forms of owner-manager control, and sources of financing.

Dagrupa and David (1987) argued that the scientific ethos of pure scientists is for wide dissemination of the results. To say that it benefits others is a tautology. Arrow (1962) showed the need for public funding of research that produces a positive externality. As long as the externalities of basic research are internal to a nation, tax supported research provides an
acceptable solution. When information provides externalities to foreign firms, the allocation of tax dollars for basic research in one country raises valid questions. But one should not fall to the temptation to attribute the success of one nation to its getting a free ride on basic research in another nation. Research in one nation is available to every other nation. If one wants to objectively study international competitiveness in terms of externalities arising from national research, one should address the question why some nations seem to benefit and others do not.

This may require a truly international tradition in industrial organization.

NOTES

1. Actually, Eq. (1.1) could be regarded as a special form of the more general two factor production function \( Q = F(K, L) \), where the time variable \( t \) allows for technical change. In (1.1) \( F(K, L) \) represents a scalar measure of the "total factor input" to the production process. From (1.1) it is clear that \( A(t) = Q/(F(K, L)) \), and in this form it is obvious that \( A(t) \) is a measure of the productivity of the "total factor input.” Eq. (1.3) readily follows from \( A = Q,F \) and the linear homogeneity of \( f \) follows the substitution of the share weighted sum of \( K \) and \( L \) for \( F \). Hence the interpretation of \( A(t) \) as the rate of technical change is equivalent, in this framework, to the percentage change in the total factor productivity index. A. (Link (1967) credits Tinbergen (1959) and Stigler (1947) for introducing the notion of total factor productivity.

2. As a comparison, Vayntruk-Vail's (1955) estimate of the rate of technical progress was 8% per decade or approximately 0.77% per year, for his sample period 1869-1952.


4. For an intuitive interpretation of the effects of technical progress on the supply and demand of a given factor, see Mitchell (1989).

5. Since we refer to "the" elasticity of substitution, we obviously have in mind a two-factor production process. For more on "the" elasticity of substitution, see our earlier work (1984).

6. See, for example, Sato and Beckmann (1968), Nadiri (1970), and Takayama (1974).

7. The representations of factor augmenting technical change in (1.6), (1.7), and (1.8) are of the familiar and popular, multiplicative type. See Rose (1968), Takayama (1974), Eichhorn and Korn (1974), Eichen (1978). While the multiplicative form is more familiar, we note that a broader interpretation is possible. Chambers (1988, p. 210) sees "factor augmenting technical change" as any general process which improves input efficiency, so that effective inputs, say \( K, L \), depend in general on "the state of technology as well as the level of actual input usage." According to this view, the types of technical change represented in (1.6), (1.7), and (1.8) are merely special cases under the general formulation \( K^* = K(1 + \theta(K, L)) \) and \( L^* = L(1 + L(L, K)) \). On this point, see also Sato (1980, 1981) and Mitchell (1984, 1987, 1989).

8. See also Stevenson (1980) or Jorgenson, Gillop, and Fraumeni (1987) who drop the logarithmic nature of the derivative in (E.9): \( b = b\theta/K(1 + \theta) \). Boweswager (1974b, p. 964) noted that "this definition has the advantage that it leads to a single measure of bias for each factor in the n-factor case while Hicks' definition would lead to \( n \cdot 1 \) measures of bias for each factor."

9. Chambers (1988, p. 219) has noted that many individuals refer to these share-based measures of bias as "share-saving," "neutral," and "using respectively. Input-saving, neutral, and using biases are based on the sign of \( dG_1(w) ) \), in which the demand for input 1 is a function of \( x, \theta, \) factor price vector \( w, \) and the technology index \( t. \)

10. This explains the equivalence between "labor-saving" and "capital-using" as alternative descriptions of the same type of technical progress when there are only the two factors \( K \) and \( L \).

11. This is the dual analogue—under constant returns to scale—of the definition of the rate of technical change from a production function: \( r = -d\ln f(K, L)/dt \). See-Ohms (1974).


13. This merely assigns the simple exponential form to Boweswager's (1974b, p. 965) As and Chambers' (1988, p. 223) AS.

APPENDIX

A List of Symbols for Sections II-A and II-B

\[ Y = \text{output} \quad F(K, L) = \text{production function} \]
\[ K = \text{capital} \]
\[ L = \text{labour} \]
\[ T = \text{Hickman neutral type of technical progress} \]

C or TC = cost function

\[ w = \text{wage rate} \]
\[ r = \text{return to capital} \]
\[ \omega = \text{inflation rate of output price} \]
\[ \beta = \text{inflation rate of factor input prices} \]
\[ g = \text{real rate of technical change} \]
\[ \alpha = \theta, \text{demand price of output} \]
\[ P = \text{price of output} \]
\[ Z = \text{parameter of aggregation production} \]
\[ h = \text{technical progress function} \]

A = stock of applied knowledge

B = stock of basic knowledge

\[ \theta_a, \theta_b = \text{research workers} \]
\[ \theta_a = \theta_b = \text{research capital} \]
\[ \theta_a = \text{applied-research expenditure as proportion of total revenue} \]
\[ \theta_b = \text{basic-research expenditure as proportion of total revenue} \]

R = total revenue

\[ q = \text{discount factor} \]
\[ \theta = \text{transformation factor} \]
\[ \theta = \text{depreciation rate of applied knowledge} \]
\[ \theta = \text{depreciation rate of basic knowledge} \]
\[ \theta = \text{price of research workers} \]
\[ \theta = \text{price of research capital} \]
\[ H = \text{Hamiltonian function (dynamic optimization)} \]
\[ * = \text{notion for equilibrium values} \]
REFERENCES


“The Economics of Technical Progress”: A Comment

Barbara M. Fraumeni

The first half of this paper begins with a lovely overview of technical change, being clear and well presented. For someone well versed in the subject, it was necessary to begin to concentrate more on the paper when the discussion of factor or price augmentation began, particularly with regard to the possible confusion between technical change and economics of scale in empirical work. In terms of solving identification problems in economic models, I wonder how common holotheticity, or simple relabelling of isologous with technical change, is. I suspect that holotheticity is a curiosity more than a real issue.

The technical change formulation suggested for estimation seems quite flexible, as it allows for different forms of technical change and apparently only requires current dollar data. I assume that determination of the actual form used, i.e. price augmenting, additive or no technical change etc., is done by experimentation. I would be interested to hear how difficult this turns out to be in practice.

In the second part of the paper at least initially a more restrictive assumption is made about technical change, specifically that it is Hicks neutral in form. To integrate the two parts of the paper it might be advisable to estimate a form for technical change in the first half of the paper and assume that form in the second half. In addition, if the technical change survey included a brief survey on R&D the paper would be more cohesive.

Being a wireless capital stock generator, I enjoyed the comment that their model essentially went back to the Middle Ages and assumed a benchmark of zero for basic knowledge. I wish I could do the same for capital stock.

I am not comfortable with the assumption that commodity prices increase at the same rate as prices of inputs in the R&D sector. R&D inputs produce and intermediate output, an output that is very different from the output of the sector as a whole. If this was done for mathematical convenience, it should be indicated as such.

To rationalize why basic knowledge is worth more the higher the stock of technical knowledge, I thought of an industry in which technology is vital. Is this the idea? The discussion of how the distinction between basic and applied research is important to public policy is a beneficial addition to the paper. It’s intuitively obvious once described, but so often intuition is omitted in a mathematical paper. The idea that R&D can “save” an industry with slowly rising output prices (or sluggish demand) is a general point that applies to technical change in general.

Finally, onto the Cournot-Nash part of the paper. Does firm II have an advantage in a closed loop strategy because it does not have basic research costs? In the N + 1 firm model I was surprised that Koe was below Koe. Is it fair to draw an analogy with the von Stackelberg model in which the leader-follower situation is the best for the leader of all possible situations? It would be best if this result was motivated with some insight into what drives the model.

Overall, the Sato and Mitchell paper is an excellent paper, a pleasure to read and discuss.

*Northeastern University, Boston, MA.