WORKING PAPER SERIES

EDUCATION, INFORMATION, AND EFFICIENCY

Finis Welch

Working Paper No. 1

NATIONAL BUREAU OF ECONOMIC RESEARCH, INC.

NBER WORKING PAPER SERIES

EDUCATION, INFORMATION, AND EFFICIENCY

Finis Welch

Working Paper No. 1

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR AND SOCIAL INSTITUTIONS
National Bureau of Economic Research, Inc.
261 Madison Avenue, New York, N.Y. 10016

June 1973

Preliminary; Not for Quotation

Education, Information and Efficiency

Finis Welch

This represents two chapters of a proposed book co-authored by Bob Evenson and me. The subject is relationships between agricultural productivity, research and information.

The first chapter of this part is concerned with the "theory" of the value of information. Among other things, the Bayesian learning model is used as a vehicle for describing optimal learning from experience. The second chapter presents results for a number of empirical studies concerned with relationships between education and allocative efficiency. Section I, is reprinted from my J.P.E. paper, "Education in Production". Section II is the "Scale Economy" paper of mine which has existed in various unpublished versions for two years now. The final section summarizes recent dissertations by Wallace Huffman (Chicago), Nabil Khaldi (SMU), and Charles Fane (Harvard).

I would appreciate any "constructive" comments.

Information and Efficiency

The purpose of this chapter is to present a framework for determining the value of allocative skills, a point about which the productions literature is surprisingly silent. Of the relatively few discussions concerned with efficiency most focus only on questions of technical efficiency -- of failure to maximize output given input. In fact, in Friedman's definition of entrepreneurial capacity he opts for a purely technological definition: manager A is more efficient than B, if A gets more output from the same inputs, i.e., if A's production functions is superior. A more complete definition that takes account of technical and allocative efficiency is that A is more efficient than B, if when faced with the same input and product prices, A earns more profit or returns to his managerial input. The extra return could arise either from different production functions or from an added ability to optimize, to have smaller ex post errors between marginal factor costs.

In a world of perfect information, there is no room for questions of efficiency. All actors are perfectly informed of the consequences of any given action and each will do whatever is necessary to maximize whatever he maximizes. In this sense, all will be perfectly efficient. If information is imperfect, the actor follows an optimization strategy which takes account of ex post errors. This is to say only that hindsight is better than foresight and ex ante optimizing stratagies are those which minimize the preception (or expectation) of the cost of ex post

---

*I am indebted to John Hause, Jim Heckman, John Koehler, and Lee Lillard for helpful comments. Support for this research is provided by the National Science Foundation, NSF-GS-28030.

mistakes. In the following sections a framework is presented to describe the cost of imperfect information and the implications following from the existence of learning possibilities.

In Section I, the cost of ignorance is defined under the assumption that decision units maximize expected yields. In Section II the Bayes learning model is presented as a basis for determining the value of information or learning. The next section parameterizes the cost of ignorance and the associated information revenue functions for quadratic yield functions and, what amounts to the same thing, for second-order Taylor approximations to more general functions. The idea is that for the quadratic or second-order approximation, cost of imperfect information are functions of error variances in activities or of variances in state parameters which result in activity errors. Accordingly, the value of information is derived from the foreseen reductions in error variances.

The remaining sections describe a number of applications. Section IV deals with learning from experience. Specifically it considers the problem of input selection for a firm that has imperfect knowledge of its production function. Each input trial serves the dual purpose of contributing of short term profit and of contributing information of production parameters. Optimal learning from experience implies that the expected marginal loss of short term profits associated with an experimental input trial is compensated at the margin by the expectation of increased future revenues resulting from more accurate choice of future inputs. Although the initial application allows only sequential experimentation, a discussion follows which takes account of opportunities for spacial experimentation via replication of experiments. One interesting result is that when spacial experimentation is possible, multi-plant firms will not be technically efficient, i.e., they will not maximize aggregate output from given input aggregates. The final part of Section IV discusses problems associated with systems change and the oblesence of knowledge.

Section V presents approximate losses for firms and consumers as functions of anticipated error variances in first-order or marginal optimizing conditions together with a brief discussion of the effects of uncertain quality of goods. Finally, Section VI describes scale economy in the use of information.

I. The Cost of Ignorance

Ignorance is defined as the perception of imperfect information. If one is ignorant (has imperfect information) but ignorant of his ignorance he will act as though he has perfect information and this case is not analyzed.

Actors, (decision making units) are assumed to have yield functions: for the firm these are profit functions and for the consumer they are utility functions. Yield is defined as a function of two classes of variables, those endogenous (determined by the actor) and those that are exogenous (given by the state of the world), the state vector. To the firm, the state vector includes input supply functions, production technology, and product demand functions. The firm's activities, the endogenous variables, consist of inputs. To the consumer, the state vector consists of goods prices, initial wealth, and demand functions for his labor. His activities include the labor he sells and the goods he purchases. A condition of ignorance is defined as a subjective density function of alternative states of the world. The yield function for a given state of the world is presumed to be concave with respect to activities so that corresponding to each state (vector) there is an activity (vector) that maximizes yield. Ignorance has cost only if activities are determined without knowledge of the state of the world. In this case the assumption is added that the actor selects that activity which maximizes expected yield. Given the yield function

(1) $$\pi = \pi\ (x,\ \beta)$$

with x indicating activities and β referring to the state of the world, the activity which maximizes expected yield is

$$(2) \qquad x = \max_x \int_\beta g(\beta)\, \pi(x, \beta)\, d\beta$$

where g(β) is the subjective density function describing the condition of ignorance and within the integral x is independent of β. Define x(β) as the yield maximizing activity for given β so that

$\pi(x(\beta), \beta) - \pi(x,\beta) \geq 0$ which gives the ex post cost of ignorance for any realized state, β. This indicates that although the actor selects his activity to maximize expected yield, he is aware that corresponding to any state vector the activity that he has selected need not be optimal. The cost of ignorance is defined as the expected cost of these ex post mistakes, i.e.,

$$(3) \qquad \text{C.I.} = \int_\beta g(\beta)\, (\pi(x(\beta),\beta) - \pi(x,\beta))\, d\beta.$$

The cost of ignorance is expected maximum less maximum expected yield.

II. The Bayes Learning Model

To learn is to move from one condition of ignorance to another, to modify the subjective density of states of the world. The problem with measuring the value of learning is that one does not know what he will learn until he has learned it, and then, since he has learned what he has learned, it may be difficult to get him to pay for the experience. To avoid problems of enforcing contracts with ex post payment, assume that payment occurs (or that value is imputed) before learning. In this case, the value of a learning experience is the anticipated increase in expected yeild associated with the experience, and learning is itself treated as a (subjectively) random variant. In conjunction with the earlier discussion, let $g(\beta)$ refer to the initial condition of ignorance and let $f(\ell)$ describe the density of learning realizations (i.e., resultant ignorance conditions) with $h(\beta|\ell)$ being the ignorance density for a given ℓ. Bayes' formula is that for all β,

$$(4) \qquad g(\beta) = \int_{\ell} f(\ell)h(\beta|\ell)d\ell \quad ,$$

i.e., the mean of $h(\beta|\ell)$ over $f(\ell)$ is the initial density $g(\beta)$. This is a strong assumption. It is not clear a priori that following a learning experience the probability of a given state of the world, β_o, when averaged over prior learning probabilities, $f(\ell)$, should equal the probability initially assigned to that state. Assume then that for β_o, the posterior mean (over learning possibilities) is $g(\beta_o) + \delta(\beta_o)$ so that the actor expects the learning experience to increase the probabilities he assigns to β_o by $\delta(\beta_o)$. It is clear, however, that the actor will not pay for the $\delta(\beta_o)$ because he can simply augment his initial prior by this amount without being submitted to or having to pay for the learning experience. Thus to the extent that learning has value, that value is subject to the restriction imposed by equation (4).

Corresponding to each ignorance condition foreseen as resulting from the learning experience there is an activity, $x(\ell)$, which maximizes expected yield, i.e.,

$$x(\ell) = \underset{x}{\text{Max}} \int_{\beta} h(\beta|\ell)\pi(x,\beta)d\beta.$$

The cost of ignorance foreseen as following the learning experience, L, is

$$(3.a) \quad CI(L) = \int_{\ell} f(\ell) \int_{\beta} h(\beta|\ell)(\pi(x(\beta),\beta) - \pi(x(\ell),\beta))d\beta d\ell$$

Notice, however, that expected maximum yield for the prior learning experience is the same as expected maximum yield in the initial ignorance condition. That is,

$$\int_{\ell} f(\ell) \int_{\beta} h(\beta|\ell)\pi(x(\beta),\beta)\,d\beta\,d\ell = \int_{\beta}\pi(x(\beta),\beta) \int_{\ell} f(\ell)h(\beta|\ell)\,d\ell\,d\beta$$

$$= \int_{\beta} g(\beta)\pi(x(\beta),\beta)\,d\beta \quad \text{(equation (4))}.$$

Let x(o) represent the activity that maximizes expected yield in the initial contion described by $g(\beta)$ with CI(o) referring to the cost of ignorance in that condition. Expected yield in the initial condition is

$$\int_{\beta} g(\beta)\pi(x(o),\beta)\,d\beta = \int_{\ell} f(\ell)\int_{\beta} h(\beta|\ell)\pi(x(o),\beta)\,d\beta\,d\ell \;.$$

But according to the definition of $x(\ell)$, for each ℓ,

$$\Delta\bar{\pi}\,(\ell) = \int_{\beta} h(\beta|\ell)(\pi(x(\ell),\beta) - \pi(x(o),\beta))\,d\beta \geq 0.$$

and the value of the learning experience,

(5) $VL = \int_{\ell} f(\ell)\Delta\bar{\pi}(\ell)\,d\ell$, is the anticipated increase in expected yield which, insofar as learning is foreseenas possibly changing activities, has positive value. Notice, of course, that the value of learning is also the expected reduction in the cost of ignorance:

(5.a) $VL = CI(o) - CI\,(L).$

III. Parameterization

It is clear that the cost of ignorance arises from subjective dispersion in state parameters, so that it is tempting to model the value of information in terms of its anticipated effects in reducing subjective dispersion. In this section, I present second-order approximations that treat learning or information essentially as the anticipation of reductions in (subjective) dispersion of state parameters. The effects of learning must, however, be intermediated through activities for these approximations to serve as valid indexes of the value of information.

In general, there is no presumption that the yield function is concave with respect to state parameters. The presumption is only that for given states of the world, yield is concave with respect to activities. Assume then that yield is convex with respect to state parameters. In this case if learning is foreseen simply as reducing subjective dispersion (about a given mean), then learning reduces expected maximum yield. Of course, the Bayes formulation does not permit this result. The initial ignorance condition is treated as the mean of the conditional learning distributions, so that the marginal distribution of maximum yields (over learning possibilities) is the initial distribution.

In these models when activities are selected without prior knowledge of states of the world, the cost of ignorance arises from the prior expectation of ex post mistakes, from recognition that after the state of the world is specified activities will not generally be optimal. The value of information arises through partitioning of the initial ignorance distribution among the conditional distributions so that for each of these distributions the activity that maximizes expected yield is preferred to the activity that would have been selected without the information. In a sense, learning is foreseen as the ability to adjust activities to levels that are "closer" to optimal than is otherwise possible.

Quadratic Yield Functions. To illustrate the above discussion assume that the yield function, $\pi(x(\beta),\beta)$, is quadratic with respect to activities. We have then that $\frac{\partial\pi}{\partial x}$ is linear in x and according to the definition of $x(\beta)$ as the activity that maximizes yield,

$$\left.\frac{\partial\pi}{\partial x}\right|_{x = x(\beta)} = 0.$$

It follows that

$$\frac{\partial\pi}{\partial x} = b(\beta)(x - x(\beta))$$

and that

$$\frac{\partial^2\pi}{\partial x^2} = b(\beta)$$

where $b(\beta)$ is negative and is distributed independently of x. The yield corresponding to activity, x, for given β is

(6) $$\pi(x,\beta) = \pi(x(\beta),\beta) + \left.\frac{\partial\pi}{\partial x}\right|_{x=x(\beta)} (x - x(\beta)) + \frac{1}{2}\frac{\partial^2\pi}{\partial x^2} (x - x(\beta))^2$$

Expected yield, given x, is maximized when

$$\frac{d}{dx} \int_\beta g(\beta)\, b(\beta)(x - x(\beta))^2 d\beta = 0.$$

Define

$$g'(\beta) = \frac{1}{\bar{b}} g(\beta) b(\beta)$$

with $$\bar{b} = \int_\beta g(\beta) b(\beta) d\beta .$$

Maximization of expected yield requires minimization with respect to x of

$$\int_\beta g'(\beta)(x - x(\beta))^2 d\beta$$

which gives

$$x = \int_\beta g'(\beta) x(\beta) d\beta$$

as the mean of $x(\beta)$ in the transformed distribution $g'(\beta)$.

The cost of ignorance in the initial condition is then

$$CI(o) = \frac{-\bar{b}}{2} \int_\beta g'(\beta)(x - X(\beta))^2 d\beta = \frac{-\bar{b}}{2} \sigma^2_{x(\beta)}$$

and is proportionate to the error variance in activities associated with the decision, x. Transform the conditional densities, $h(\beta|\ell)$ into $h'(\beta|\ell)$

$$= \frac{b(\beta)}{\bar{b}} h(\beta,\ell)$$

and the learning density into $f'(\ell) = \dfrac{\bar{b}(\ell)}{\bar{b}} f(\ell)$

where $\bar{b}(\ell)$ is the mean of b in $h(\beta|\ell)$ and $\bar{b}$ is the mean of $\bar{b}(\ell)$ in $f(\ell)$ as well as in $g(\beta)$. The value of a learning experience is then

$$(7) \qquad CI(0) - CI(L) = -\frac{\bar{b}}{2} \int_\ell f'(\ell) \int_\beta h'(\beta|\ell)[(x(0)-x(\beta))^2 - (x(\ell)-x(\beta))^2] d\beta$$

$$= -\frac{\bar{b}}{2} \int_\ell f'(\ell)(x(\ell) - x(0))^2 d\ell$$

where $x(\ell)$, the mean of $x(\beta)$ in $h'(\beta|\ell)$, maximizes expected yield in $h(\beta|\ell)$ and $x(0)$, the mean of $x(\beta)$ in $g'(\beta)$, is also the mean of $x(\ell)$ in $f'(\ell)$. Thus in $g'(\beta)$, variance of $x(\beta)$ about $x(0)$ can be partitioned into the sum of the variance of $x(\ell)$ about $x(0)$ between the conditional learning distributions and a $f'(\ell)$ weighted average of variances of $x(\beta)$ about $x(\ell)$ within the conditional learning distributions. The cost of ignorance is proportionate (with the same factor) to the between class variance. Thus, in this sense, the value of information is proportional to the predicted change in activity error variances.

Although the yield function is not necessarily quadratic, in accordance with a second-order approximation following the Taylor expansion I will simply assume that it is. This assumption permits costs of ignorance to be viewed, in what I think is an intuitively appealing description, as functions of anticipated error variances in activities. Or, since ignorance is the perception of dispersion in states of the world, the cost of ignorance is also a function of subjective variance in state parameters when the impact of state dispersion is intermediated into activity errors. Accordingly, the value of information is measured in terms of its effects in reducing error variances in activities via reductions in state dispersion.

The Second-Order Approximations. The Taylor expansion for yields corresponding to a given activity, x, when expanded about the optimal activity, $x(\beta)$, is

(8) $$\pi(x,\beta) = \pi(x(\beta),\beta) + \frac{\partial\pi}{\partial x}(x - x(\beta))$$

$$+ \frac{1}{2}(x-x(\beta))' \frac{\partial^2\pi}{\partial x'\partial x}(x-x(\beta)) + \text{higher order terms.}$$

Evaluated at $x(\beta)$, $\frac{\partial\pi}{\partial x} = 0$ and $\frac{\partial^2\pi}{\partial x'\partial x}$ is negative definite. Notice that the more general notation is used to permit x to be a (k by ℓ) vector. In this notation, $\frac{\partial^2\pi}{\partial x'\partial x}$ is the Hessian with

$$\frac{\partial^2\pi}{\partial x'\partial x} = \{\frac{\partial^2\pi}{\partial x_i \partial x_j}\} = \{\Pi_{ij}\}.$$

Since $\{-\Pi_{ij}\}$ is postive definite it can be written as a matrix whose typical element is

(8.a) $$\{-\Pi_{ij}\} = \{\sum_{\ell}^{k} a_{i\ell} a_{i\ell}\}.$$

The second term in the expansion is then

$$-\frac{1}{2}\sum_{\ell=1}^{k} {}'a_{\ell}\,(x-x(\beta))(x-x(\beta))'a_{\ell}$$

where a_{ℓ} is K by ℓ with typical element $\{a_{i\ell}\}$. Transforming the density, $g(\beta)$, into the loss-weighted density, $g'(\beta)$, as in the quadratic case, gives the cost of ignorance as

(9) $$CI(0) = \frac{1}{2}\sum_{\ell=1}^{k} \bar{a}_{\ell}'V(x)\bar{a}_{\ell}$$

where $\bar{a}_{\ell}$ is the mean of a_{ℓ} in $g(\beta)$ and $V(x)$ is a k by k matrix of loss-weighted activity error variances.

To express the cost of ignorance as a function of state parameter variance, note that as a first-order approximation,

$x-x(\beta) = \frac{\partial x}{\partial \beta}(\tilde{\beta}-\beta)$ where β is r by ℓ. The loss associated with x is

$$\frac{1}{2}(\tilde{\beta}-\beta)'\frac{\partial x'}{\partial \beta}\{-\pi_{ij}\}\frac{\partial x}{\partial \beta}(\tilde{\beta}-\beta) = \frac{1}{2}\sum_{\ell=1}^{k} c_{\ell}'(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'c_{\ell}$$

where c_{ℓ} r by $\ell = \frac{\partial x'}{\partial \beta} a_{\ell}$.

a_{ℓ} is as defined in equation (8.a). The cost of ignorance is then

(10) $$CI(0) = \frac{1}{2}\sum_{\ell=1}^{k} \bar{c}_{\ell}'\,V(\beta)\,\bar{c}_{\ell}$$

with $V(\beta)$ being $E(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'$ for the loss-weighted densities so that $V(\beta)$ is a covariance matrix r by r for state parameters. Recall that from equation (7) the value of information is proportional to the anticipated reduction in activity error-variances associated with learning. Therefore,

(7.a) $$VL = -\frac{\bar{b}}{2} [E_{g'(\beta)} (x-x(\beta))^2 - E_{f'(\ell)} E_{h'(\beta|\ell)} (x(\ell)-x(\beta))^2]$$
$$= -\frac{\bar{b}}{2} (\sigma_x^{\,2} - \sigma^2_{x(\ell)}) = -\frac{\bar{b}}{2} \delta\sigma_x^{\,2}$$

with δ being the anticipated percentage reduction in activity error variance. By analogy, the value of information corresponding to costs depicted in equation (9) where cost refers to activity error variances and equation (10) in which cost is expressed in terms of variance in state parameters is

(9.a) $$VL = \frac{1}{2} \sum_{\ell=1}^{k} \bar{a}_\ell' \Delta V(x) a_\ell$$

and

(10.a) $$VL = \frac{1}{2} \sum_{\ell=1}^{k} \bar{c}_\ell' \Delta V(\beta) \bar{c}_\ell .$$

The value of information is thus approximated by functions that take account only of forseen changes in error variances of either activities or of decision parameters.

IV. Applications:

Uncertain Technology and Learning From Experience

Consider a problem in experimental design. Assume that a profit maximizing competitive firm does not know its production function. The function is estimated and the estimates are revised with each new observation, with each realization of output corresponding to each input trial. Assume, first, that the firm has no other option for acquiring information. Scale replication and learning from experience of others are each ruled out for the moment but are considered later. Here, the firm faces a problem of acquiring knowledge, one of its many forms of capital. Against the contribution of an input set to expected short terms profit, the firm balances the contribution of these inputs to reductions in parameter estimate variances, for reduced estimation error reduces the size and therefore the cost of future mistakes. Let the unknown production function be $y = f(x,v)$. Assume that this function is estimated through ordinary least squares (OLS) regression based on historic observations. Assume also that the production function can be linearized so that

$$(11) \qquad y_* = z'\beta + u$$

describes the function for which y_* and z' represent suitable transformation of y and x' and β represents production parameters. (In the Cobb-Douglas function $y_* = \log y$ and $z_i = \log x_i$.) From equation (10) the cost of ignorance corresponding to a given variance of state parameters is

$$CI(0) = \frac{1}{2} \sum_{\ell=\ell}^{k} \overline{c}_{\ell}' \, V(\beta) \, \overline{c}$$

so that minimization of the cost of ignorance requires minimization of $V(\beta)$. Fortunately, if the production function can be linearized as in equation (11), and if the stochastic terms, u_i, corresponding to the i^{th} (past) observation are independently distributed (of Z and of u_j for $j \neq i$) with constant variance and zero mean, then OLS regression minimizes $V(\beta)$ according to the Gauss-Markov Theorem.

Define β_o as the OLS estimate

$$\beta_o = (Z_o' Z_o)^{-1} Z_o' Y_o$$

based on T previous observations such that the i^{th} row of $Z_o (T \times r)$ is the i^{th} past observation Z_i'. It follows that

$$V(\beta_o) = \sigma_u^2 (Z_o' Z_o)^{-1}.$$

Now consider the effect on $V(\beta)$ of adding a new observation, $Z_1 (n \times r)$, to form β_1, the modified estimate. Since $V(\beta_1)$ depends only on the design matrix $Z = \binom{Z_0}{Z_1'}$ the effects of Z_1 in reducing the variance of $\hat{\beta}$ are known, at least up to σ_u^2, before the experiment is performed, i.e., before the y's associated with Z_1 are observed. In fact,

$$V(\beta_1) = \sigma_u^2 (Z_o' Z_o + Z_1' Z_1)^{-1}$$

and it can be shown that /1

$$(12) \qquad \Delta V(\hat{\beta}) = V(\beta_0) - V(\beta_1) = \sigma_u^2 (Z_o'Z_o)^{-1} Z_1'(Z_1(Z_o'Z_o)^{-1}Z_1' + I_N)^{-1} Z_1 (Z_o'Z_o$$

which with equation (10.a),

$$(10.a) \qquad VL = \frac{1}{2} \sum_{\ell=1}^{K} \bar{C}_\ell' \Delta V(\beta) C_\ell$$

gives the informational value, for one period, of the experiment, Z_1. Let Z_1 refer to a single observation z_1' and let ψ represent a present worth summary statistic (discussed later) for the life expectancy of the information contained in z_1', so that the value of the experiment becomes

$$(17) \qquad VE = \frac{\psi}{2(z_1'(Z_o'Z_o)^{-1}z_1 + 1)} \sum_{\ell=1}^{k} \bar{c}_\ell'(Z_o'Z_o)^{-1} z_1 z_1'(Z_o'Z_o)^{-1}\bar{c}_\ell.$$

Since

$$\bar{c}_\ell'(Z_o'Z_o)^{-1} z_1$$

/1 $Z'Z = Z_o'Z_o + Z_1'Z_1$

Pre-multiply by $(Z'Z)^{-1}$ and post-multiply by $(Z_o'Z_o)^{-1} Z_1'$ which gives

$$(Z_o'Z_o)^{-1}Z_1' = (Z'Z)^{-1}Z_1' + (Z'Z)^{-1}Z_1'Z_1(Z_o'Z_o)^{-1}Z_1'$$

then

$$(Z_o'Z_o)^{-1}Z_1' \; (I_n + Z_1(Z_o'Z_o)^{-1}Z_1')^{-1} = (Z'Z)^{-1}Z_1'$$

post multiply by

$$Z_1(Z_o'Z_o'z_o)^{-1}.$$

$$(Z_o'Z_o)^{-1}Z_1'(I_n+Z_1(Z_o'Z_o)^{-1}Z_1(Z_o'Z_o)^{-1} = (Z'Z)^{-1}Z_1'Z_1(Z_o'Z_o)^{-1}$$

but

$$(Z'Z)^{-1}Z_1'Z(Z_o'Z_o)^{-1} = (Z_o'Z_o)^{-1} - (Z'Z)^{-1}$$

and

$$(Z'Z)^{-1} = (Z_o'Z_o)^{-1} - (Z_o'Z_o)^{-1}Z_1'(I_n+Z_1(Z_o'Z_o)^{-1}Z_1')^{-1}Z_1(Z_o'Z_o)^{-1}$$

is scalar,

$$\bar{c}_\ell'(Z_o'Z_o)^{-1}z_1 = (\bar{c}_\ell'(Z_o'Z_o)^{-1}z_1)' = Z_1'(Z_o'Z_o)^{-1}\bar{c}_\ell$$

and

$$\sum_{\ell=1}^{k} \bar{c}_\ell'(Z_o'Z_o)^{-1}z_1 z_1'(Z_o'Z_o)\bar{c}_\ell$$

$$z_1'(Z_o'Z_o)^{-1} \{\sum_{\ell=1}^{k} \bar{c}_\ell \bar{c}_\ell' \}(Z_o'Z_o)^{-1}z_1 .$$

But

$$\bar{c}_\ell = \frac{\partial x'}{\partial \beta} \bar{a}_\ell$$

and

$$\sum_{\ell=1}^{k} \bar{c}_\ell \bar{c}_\ell' = \frac{\partial x'}{\partial \beta} \{-\bar{\Pi}_{ij}\} \frac{\partial x}{\partial \beta}$$

with

$$\{-\bar{\Pi}_{ij}\} = \sum_{\ell=1}^{k} \bar{a}_{i\ell}\bar{a}_{j\ell} .$$

So that the value of the experiment, z_1, becomes

$$\text{(17.a)} \qquad VE = \frac{\psi}{2} \frac{z_1'Az_1}{z_1'Bz_1 + 1}$$

with

$$B = (Z_o'Z_o)^{-1}$$

and

$$A = B \frac{\partial x'}{\partial \beta} \{-\bar{\Pi}_{ij}\} \frac{\partial x}{\partial \beta} B .$$

In principle, the experiment, z_1, should be designed to maximize its value net of its cost. Before considering cost, compute marginal information revenue for x_1 rather than z_1. Define Q as the function transforming x into z such that

$$z = Q(x) \quad \text{with} \quad \frac{\partial z'}{\partial x} = \{q_{ij}\} = \{\frac{\partial z_i}{\partial x_j}\} .$$

Note that

$$\frac{\partial}{\partial z_1}\left(\frac{z_1'Az_1}{z_1'Bz_1+1}\right) = \frac{2A\, z_1}{(z_1'Bz_1+1)^2}$$

when A and B are symmetric as in this case. Therefore the marginal value of the experiment, $x_1(z_1 = Q(x_1))$, is given as

$$\text{(18)} \qquad MVE = \frac{\partial VE}{\partial x_1} = \psi \frac{Az_1}{(z_1'Bz_1+1)^2} \frac{\partial z'}{\partial x_1}$$

which is k linear (in the z's) equations, each divided by a fourth-order polynomial.

Costs for the experiment, x_1, is defined simply as maximum expected yield less expected yield at activity, x_1. For the second-order approximation, this is

$$\text{(19)} \qquad EC = \frac{1}{2}(x_1-x_*)'\{-\Pi_{ij}\}(x_1-x_*)$$

where x_* is the activity that maximizes expected yield. Marginal experimental cost is then the linear function,

$$\text{(20)} \qquad MEC = \frac{\partial EC}{\partial x_1} = \{-\Pi_{ij}\}(x_1-x_*).$$

The first-order conditions for an optimum experiment equate marginal revenue and marginal cost giving

$$\psi\, Az \frac{\partial z'}{\partial x} = (z_1'Bz_1+1)^2\{-\Pi_{ij}\}(x_1-x_*)$$

which, even if $\frac{\partial z'}{\partial x}$ is constant, i.e., if z is linear in x, gives a fifth-order polynomial. It is not generally possible to derive analytic solutions to these equations, but in application, numeric solutions can be obtained. Turn, for example to the simplest possible case, the quadratic production function,

$$y = ax + bx^2 + ux$$

and assume that this function is estimated by OLS in average product form,

$$\frac{y}{x} = a + bx + u.$$

Based on T historic observations with activity variance σ_x^2 and mean $\bar{x}$, the variance of $(\hat{a}, \hat{\beta})$ is given as

$$V_o\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} = \frac{\sigma_u^2}{T\sigma_x^2}\begin{pmatrix}\sigma_x^2 + \bar{x}^2 & -\bar{x}\\ -\bar{x} & 1\end{pmatrix}.$$

With a new observation, x_1, the change in variance is

$$\Delta V\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} = V_o - V_1 = \frac{\sigma_u^2}{\sigma_x^2((T+1)\sigma_x^2 + (x_1-\bar{x})^2)}\begin{bmatrix}(\sigma_x^2-\bar{x}(x_1-\bar{x}))^2 & (\sigma_x^2-\bar{x}(x_1-\bar{x}))(x_1-\bar{x})\\ (\sigma_x^2-\bar{x}(x_1-\bar{x}))(x_1-\bar{x}) & (x_1-\bar{x})^2\end{bmatrix}$$

Assume unit product price with input price, p. To maximize expected profit, x is selected according to the rule,

$$x_* = \frac{(\hat{a}- p)}{-2\hat{b}}.$$

so that

$$\frac{\partial x}{\partial \hat{a}} = -\frac{1}{2\hat{b}}$$

and

$$\frac{\partial x}{\partial \hat{b}} = \frac{(a-p)}{-2b^2}.$$

In this case, the information value of the experiment, x_1 is

$$\frac{1}{2}\frac{\sigma_u^2\psi}{-4\hat{b}^3T\sigma_x^2}\frac{\{(\hat{a}-p-\hat{b}\bar{x})(x_1-\bar{x}) - \hat{b}\sigma_x^2\}^2}{((T+1)\sigma_x^2 + (x_1-\bar{x})^2)}$$

and experimental cost is

$$-\frac{1}{2}\hat{b}(x_1-x_*)^2.$$

Equation of marginal cost and revenue yields a fifth degree polynomial so that as in the general case analytic solutions cannot be obtained.

As a numeric example, suppose that the firms's history is summarized by

$\bar{x} = 20$

$\sigma_x^2 = 10$

$T = 9.$

Assume, also that $a = \hat{a} = 110$

$b = \hat{b} = -1$

$\sigma_u^2 = S^2 = 16$

with current conditions given by $P = 60$ and $\psi = 9$ (a discount rate of about 11 percent). The activity maximizing expected profit is $x_* = 25$. For this case, marginal information revenue is

$$40 \times \frac{3(100 + (x_1-20)^2)(3(x_1-20)+1)-(3(x_1-20)+1)^2(x_1-20))}{(100 + (x_1-20)^2)^2}$$

and marginal information cost is (x_1-25). These functions are plotted in Figure 1 with the optimum experiment given as $X_1 = 32.1$.

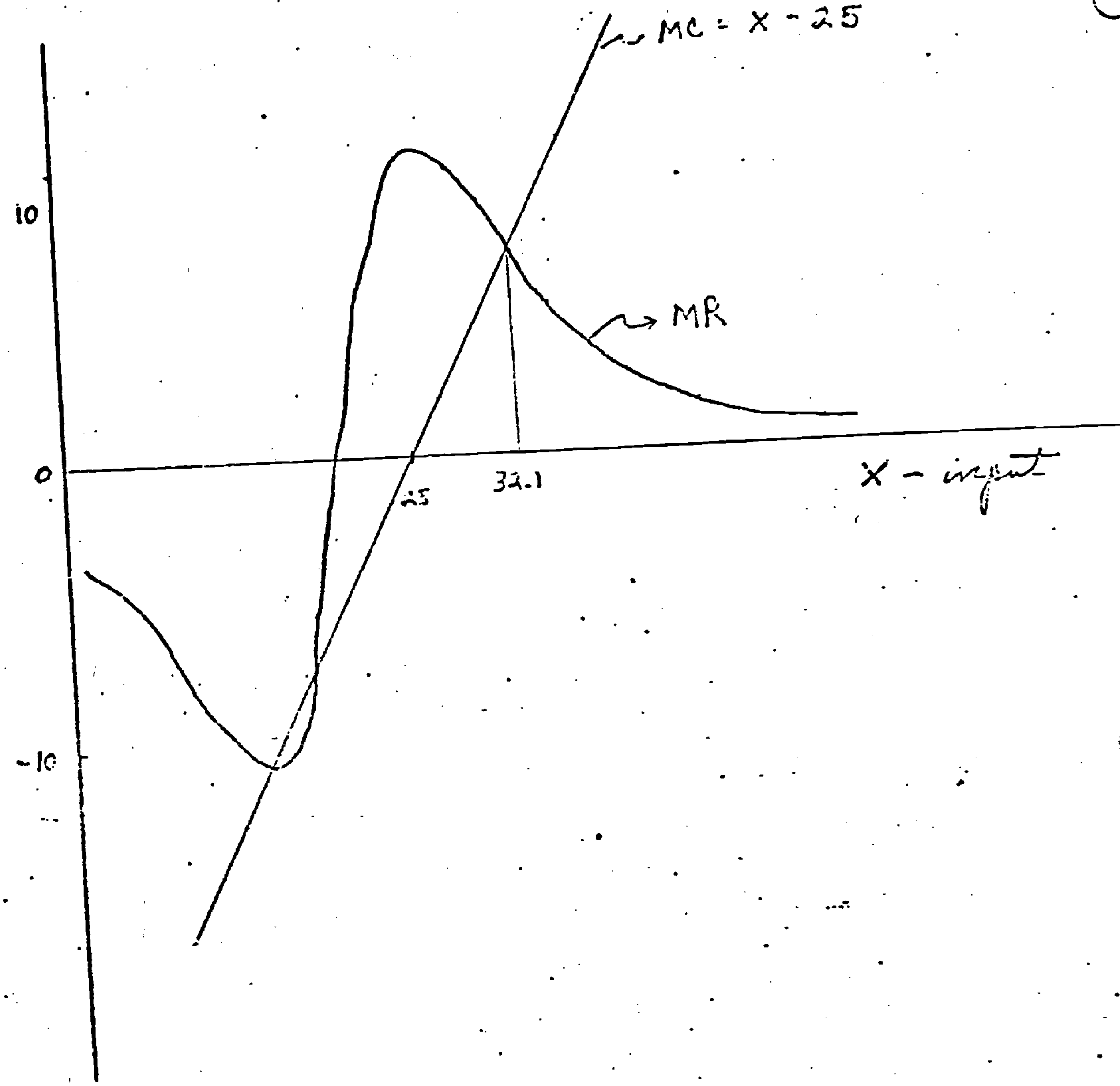

Figure 1. Marginal Informational Revenue and Cost - - An Illustration. (In the hypothesized function, there are three real roots for the net revenue function).

Returns Through Time. To this point, the problem of timing experiments has not been addressed. That information has a future was captured simply by the summary statistic ψ. Unfortunately the problem of time is not simple. Not only does the value of information depend upon all past observations (experiments) but it depends upon future experiments as well. With the static technology depicted by equation (11), tomorrow's value of today's experiment depends only on past experience together with today's experiment, but the value of today's experiment the day after tomorrow depends also upon tomorrow's experiment. And tomorrow's experiment will diminish the value of today's information the day after tomorrow and in all future periods.

The optimization criterion is that the present worth of today's experiment net of experimental cost should be maximized. But since future values depend upon future experiments, the entire future plan must be determined jointly with today's plan. If the firm faces a finite life, it will not experiment in its terminal period, and a basis exists for determining the experimental path as a solution to a dynamic programming problem. Rather than pursue this approach, consider an alternative.

Define the factor ψ as the reciprocal of the effective rate of discount and consider the orders of magnitude of this discount rate.

Notice first that implicit in the model is an incentive for

$$\text{sign}\,(z-z_*) = \text{sign}\,(z_*-\bar{z}).$$

The information value function is minimized in the vicinity of $\bar{z}$.

Clearly by introducing an experiment, $z_1 = \bar{z}$, regression estimates of slope coefficients and their variances are unchanged. The only element of ambiguity lies in the estimate of the intercept, for, in general, the greater is z_1, the greater is its effect in reducing the error variance of the intercept estimate. Thus, in the numeric example information value is not minimized at the historic mean of 20, but at a point between 19 and 20. In any case, the historic mean serves as an approximation of the point of minimum information. For movement of z toward z_* when z lies between $\bar{z}$ and z_*, informational costs are falling and information revenue is rising. An optimum experiment does not then lie between $\bar{z}$ and z_*, but would occur such that z_* lies between z_1, the optimum experiment, and $\bar{z}$, the historic mean.

Second, through time, discrepancies between z and z_* and, possibly, between z_* and $\bar{z}$ will decline. The discrepancy between z and z_* is an index of the "size" of the experiment. As a firm's history extends, the ability of another observation to revise it estimates, i.e., to reduce estimator error variances, declines. And as the value of experimentation falls, the size of experiments falls as well. Recall, however, that z_* is a function only of state parameters that are known prior to activity selection, estimates of parameters not known at the point of activity selection are given. (In the above production case, prices are presumably known prior to selection of inputs.) If these prior decision parameters are drawn from a distribution that is stable through time, they will mirror a distribution of z_* that is itself stable through time and $\bar{z}$ will move toward the mean of that distribution. Define $(z_*'z_*)$ as the steady state moment matrix such that at time, t, $(z_*'z_*) = TC$ where C is the stationary covariance matrix. C, then, serves as an index of the amount of experimentation nature subsidizes. The more definite is C, the larger its determinant, the larger is the variance in Z_*, i.e., the larger is the activity variance associated with intraperiod optimization, and the smaller is the value of information. When prior decision parameters are constant C contains non-zero elements only in the row and column referring to the regression intercept.

To determine the effective rate of discount, it is necessary to know the time-path over which Z'Z, the moment matrix indicating actual experience, moves toward Z'_*Z_*, but this requires full knowledge of the experimental path. We do know that incentives for experimentation are such that at any point in time $|Z'Z| > |Z_*'Z_*|$, i.e., experimentation augments the determinant of the design matrix. With this, we can determine upper and lower bounds for the discount factor. Consider two boundary paths: one in which all future experiments augment the design matrix in a manner similar to the average of all past experiments and another in which no experimentation has occurred or will occur. For the first, if

$$Z_o'Z_o = T_o\left(\frac{Z_o'Z_o}{T_o}\right)$$

then at future point, t,

$$(Z'Z)_t = t\left(\frac{Z_o'Z_o}{T_o}\right).$$

Since incentives for experimentation diminish as experience deepens,

$$|Z_o'Z_o| > \left|t_o\left(\frac{Z_t'Z_t}{t}\right)\right| > \left|t_o\left(\frac{Z_*'Z_*}{t}\right)\right|$$

where $Z_t'Z_t$ indicates actual experience.

Consider the marginal informational revenue in future period, t, accruing to the current period experiment. Define $Z_t'Z_t$ as the design matrix that includes $Z_o'Z_o$ plus moments on all future observations to t exclusive of the current observation.

(18.a) $$MVE_t = \frac{A_t z_1}{(z_1'B_t z_1+1)^2} \frac{\partial z_1'}{\partial x_1}$$

with

$$B_t = (Z_t'Z_t)^{-1}$$

and

$$A_t = B_t \frac{\partial x'}{\partial \beta} \{-\Pi_{ij}\} \frac{\partial x}{\partial \beta} B_t.$$

Define

$$\tilde{Z}'\tilde{Z} = t_o \left(\frac{Z_t'Z_t}{t}\right)$$

so that

$$\tilde{B}_t = (\tilde{Z}'\tilde{Z})^{-1}, \quad B_t = \left(\frac{t_o}{t}\right)\tilde{B}_t, \text{ and } A_t = \left(\frac{t_o}{t}\right)^2 \tilde{A}_t.$$

Then,

$$MVE_t = \frac{\tilde{A}_t z_1}{(z_1'\tilde{B}_t z_1 + \frac{t}{t_o})^2} \frac{\partial z_1'}{\partial x_1}.$$

For the boundary cases $\tilde{A}$ and $\tilde{B}$ are constant so that

$$MVE_t = MVE_{t_o} \left(\frac{1}{1 + \frac{t-t_o}{t_o b}}\right)^2; \; t = t_o, t_o+1, t_o+2, \ldots$$

with

$$b = z_1' \tilde{B} z_1 + 1$$

The present worth of the information flow is

$$\psi \, MVE_{t_o} = MVE_{t_o} \sum_{i=1}^{\infty} \left(\frac{1}{1+r}\right)^i \left(\frac{1}{1+\frac{i-1}{t_o b}}\right)^2.$$

If $t_o b$ is "large",

$$\left(\frac{1}{1+\frac{i-1}{t_o b}}\right)^2 \simeq \left(\frac{1}{1+\frac{2}{t_o b}}\right)^{i-1}$$

and

$$\psi \simeq \frac{1}{\left(r + \frac{2(1+r)}{t_o b}\right)}$$

with r being the market rate of interest. The effective rate of discount, ψ^{-1}, as then bounded by

$$r + \frac{2(1+r)}{t_o b_1} < \psi^{-1} < r + \frac{2(1+r)}{t_o b_2}$$

where

$$b_1 = z_1'(Z_o'Z_o)^{-1}z_1 + 1$$

and

$$b_2 = z_1'(Z_*'Z_*)^{-1}z_1 + 1$$

with Z_* being normalized to t_o observations.

The point is that

$$b_2 - b_1 = z_1'[(Z_*'Z_*)^{-1} - (Z_o'Z_o)^{-1}]z_1$$

may be "small" so that the bounds on the effective rate of discount are "tight".

For the moment matrix, X'X, based on t observations, we know that

$$x_*'(X'X)^{-1}x_* = \frac{1}{T}(1 + (x_* - \bar{x})'S^{-1}(x_* - \bar{x}))$$

where x_* is an arbitrary vector, $\bar{x}$ is the mean of X and S is the covariance matrix,

$$S_{ij} = \frac{1}{T}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j).$$

Consider the diagonal matrix λ whose i-th diagonal element is the standard deviation of x_i in x, i.e., $\lambda_{ii} = \sigma_i = (\frac{1}{T}\Sigma(x_{it} - \bar{x}_i)^2)^{\frac{1}{2}}$.

The matrix of zero-order correlation coefficients,

$$R = (r_{ij}) = \lambda^{-1}S\,\lambda^{-1}$$

and

$$S^{-1} = \lambda^{-1}R^{-1}\lambda^{-1}$$

such that

$$(x-\bar{x})'S^{-1}(x-\bar{x}) = \tilde{x}'\ R^{-1}\tilde{x}$$

where $\tilde{x}$ is the vector x expressed in standard deviation units about the mean. But,

$$R^{-1} = \frac{1}{|R|}\ \text{adj}\ (R) = \frac{1}{|R|}\ (r^{ij})$$

when r^{ij} is the co-factor of r_{ij} in R. With $x_t{}'$ being the t-th row in x, the mean

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_t{}'\ R^{-1}\tilde{x}_t = (k-1)$$

where (k-1) is the dimension of R. This follows from the observation that

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{jt} = r_{ij}$$

and that

$$\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} r_{ij}r^{ij} = |R|(k-1).$$

Recall that as z_* varies about $\bar{z}$, variation in z about $\bar{z}$ is exaggerated but that incentives to experiment fall as experience deepens so that variation in z about z_* dampens through time. It therefore seems reasonable to expect that correlation patterns within Z_o to reflect correlations in Z_* and that variance in z_t declines through time to variance in Z_*. On this basis, I assume that $z_1{}'(Z_o{}'Z_o)^{-1}z_1 < \frac{k}{t}$ where k is the number of parameters in the regression function given in equation (11). It follows that the effective discount rate, ψ^{-1}, is understated by

$$(15)\qquad r + \frac{2(1+r)}{T_o + k}.$$

Assume that steady state variance in Z_* is some fraction, s, of historic variance in Z_o and that variance in the current design, $z_1 z_1'$, is intermediate to these extremes. Let,

$$\left|\frac{Z_*'Z_*}{Z_o'Z_o}\right| = s$$

and

$$\left|\frac{t_o(z_1 z_1')}{Z_o'Z_o}\right| = \delta$$

with $1>\delta> s$. So long as the correlation patterns in $Z_o'Z_o$, $z_1 z_1'$, and $Z_*'Z_*$ are identical,

$$b_2 - b_1 = (z_1 - \bar{z})'((Z_*'Z_*)^{-1} - (Z_o'Z_o)^{-1})(z_1 - \bar{z}) = \frac{\delta}{T_o}\left(\frac{1-S}{S}\right).$$

As a numeric example, assume that three parameters are being estimated, that current estimates are based on twenty historic observations, that the market rate of interest in ten percent and that

$$s = \frac{1}{2}$$

with

$$\delta = \frac{3}{4}.$$

On the assumption of common covariances the "effective" discount rate is bounded by

(16) $$r + \frac{2(1+r)}{(k-1)\delta+t_o+1} > \psi^{-1} > r + \frac{2(1+r)}{(k-1)\frac{\delta}{s} + t_o+1}$$

and in this case lies between 19.3 and 18.5 percent. If s is reduced to 0.1, the lower bound falls to 15.3 percent. As the number of historic observations increases the bounds converge to the market rate, r, which also serves as the lower bound for the stationary case in which there is zero variance in prior decision parameters. Since δ cannot be determined independently of z_1 and since, for most purposes s is unknown, equation (15) offers an alternative upper bound. When these bounds are "close", little is gained by pursuing the dynamic solution.

The central idea in this analogy to learning - from - regression is that learning phenomena can, quite naturally, be couched in the ordinary optimizing frame. The major problem is in forming an optimal learning algorithm, and the argument here is that this rule-for-learning falls directly from the yield function. Based on any given history, the optimum way of digesting experience is that which minimizes the cost of ignorance.

For the standard case of the profit oriented firm, the problem is manageable because yield functions are profit functions. For the consumption case, the problem is more difficult because utility is harder to observe. Even so, we are not completely empty handed. In a later section, approximate loss functions for firms and consumers are presented in which errors are expressed as percentage discrepancies in first-order optimizing conditions. There, the costs of mistakes vary directly with expenditure shares and, with demand elasticities. This offers a basis for determining the allocation of resources to, say, search activity to reduce uncertainty concerning quality of goods. It also provides a rationale for the development of "habit" when habit is described as a propensity toward goods whose characteristics are more certain.

These observations are cold comfort for those interested in classroom learning behavior. Their only relevance is to note that so long as something is being optimized, the learner should not be viewed as a purely passive participant in the educational process.

Turn now to three rather straightforward extensions of the learning model: experimentation through space rather than time; the question of technical change and the obsolesence of knowledge; and scale economies in the use of information.

A. Replication of Experiments

In the analysis of learning by regression, activity design is an input vector, Z_1. Learning is possible only by experimentation and experiments are conducted in sequence, one-at-a-time. In application, there may be several modes of gleaning information. One is simply through observation of others. This adds a "public goods" aspect to the picture. If information is obtained costlessly by observing others, this appears from the individual firm's perspective as an autonomous component in state covariances, $V(\beta)$, and in terms of the earlier discussion, reduces the private incentive to experiment by increasing the effective rate of discount.

Another mode of garnering information is by conducting more than one experiment at a time. The problem is, of course, analytically similar to the design problem discussed above. The distinction is that if n experiments are simultaneously conducted the design, Z_1 is n by k rather than ℓ by k in equation (13). Here the experimental cost function can be conveniently partitioned into two parts, a mean and variance.

Consider a multi-plant firm that has the same micro-production function for each plant. Total output across plants is maximized for given input aggregates when marginal physical productivities for each input is equated across the plants. And in this case where plants are assumed to have a common (concave) production function, output is maximized when each plant uses the same input mix as all other plants.

But for given aggregate inputs, an equal allocation among plants would minimize the informational content of the experiment so that incentives are to vary the activity mix. Define an experimental cost function for input aggregates based on the Hessian, $\{-\Pi_{ij}\}$ as before and add a cost-of-interplant variance function which is proportionate to difference between maximum expected output (which requires interplant equalization of activities) and expected output corresponding to any pattern of interplant activity variance. The optimum experiment would equate the marginal informational value of the aggregate experiment relative to its marginal cost to the marginal revenue of variance relative to its cost.

As an example assume that the firm operates with a one-input Cobb-Douglas production process of the form,

$y_i = Ax_i^{\alpha}e^{u_i}$ over n-plants, were y_i is output of the i-th plant and x_i is the input level. Presumably $\alpha<1$ and this concavity gives the basis for the inter-plant variance cost function. So long as $E(e^{u_i}) = E(e^{u_j})$ (for all plants i and j) with given aggregate input $X_o = \sum_{i=1}^{n} x_i$ the firm maximizes expected output with $x_i = X_o/n$ and the expected product function is

$$y_o = \Sigma y_i = \Sigma A x_i^{\alpha} E(e^{u_i}) = AN^{1-\alpha} X_o^{\alpha}$$

(with $E(e^{u_i}) = 1$). Let $y_1 = E(\Sigma y_i)$ when not all values of $x_i = X_o/n$ and consider a second-order Taylor's expansion of y_1 about y_o.

$$y_1 = y_o + \Sigma \frac{\partial y_i}{\partial x_i}\Bigg|_{x_i=X_o/n} (x_i - X_o/n) + \frac{1}{2}\Sigma \frac{\partial^2 y_i}{\partial x_i^2}\Bigg|_{x_i=X_o/n} (x_i - X_o/n)^2 + \ldots$$

Since

$$\frac{\partial y_i}{\partial x_i}\Bigg|_{x_i=X_o/n} = \frac{\partial y_j}{\partial x_j}\Bigg|_{x_j=X_o/n},$$

the linear term becomes

$$\alpha \frac{y_o}{x_o} \Sigma(x_i - X_o/n) = 0$$

Since

$$\Sigma x_i = X_o,$$

the second term is

$$\frac{\alpha(\alpha-1)}{2} \frac{y_o n}{X_o^2} \Sigma(x_i - X_o/n)^2.$$

Let $p_i = \frac{x_i}{x_o}$ with $\bar{p} = \frac{1}{n}$ and $\Sigma p_i = 1$. Which is $y_o \frac{\alpha(\alpha-1)}{2} n^2 \sigma_p^2$ with $\sigma_p^2 = \frac{1}{n} \Sigma(p_i - \bar{p})^2$. The approximation to the production function is

$$y_1 = y_o \left(1 - \frac{\alpha(1-\alpha)}{2} n^2 \sigma_p^2\right)$$

This gives

$$y_o \frac{\alpha(1-\alpha)}{2} n^2 \sigma_p^2$$

as the cost function in physical product units for inter-plant variance which must then be adjusted for product price. To this the cost function for aggregate experiments that deviate from short term optimal ($X_o \neq X_*$) must be added to give the total experimental cost function.

Presumably the main ingredient in determining the optimum experimental mix between deviations of x_o from x_* (variance through time) and variance (through space) in x_i for given aggregates, x_o lies in the plant residuals. Concavity of the production function is of course always relevant in as much as it forms the basis for experimental cost, but concavity enters both space and time functions, and is not central to determining optimal mix. The distinction is on the revenue side, in the content of the alternative forms of information. Suppose the plant residuals can be partitioned into two stochastically independent elements,

$$u_{it} = w_t + v_{it}$$

where w_t is the firm specific error that is held in common by all plants and v_{it} is the plant specific error. The variance is then partitioned as:

$$\sigma_v^2 = \sigma_w^2 + \sigma_v^2.$$

The greater is the ratio σ_w^2/σ_u^2 the greater is the incentive to gather information via spatial variance. At a point in time, w is fixed so that the greater is σ_w^2/σ_u^2 the greater is the informational content of spatial variation in estimating the production elasticity, α.

B. Change and Obsolesence of Knowledge

To this point, discussion is restricted to static processes. Both production technology and the distributions of prior decision parameters are fixed so that, through time, stable activity distributions evolve. Instead of these static assumptions, suppose that technology and the distributions of prior parameters change over time[1]. For purposes of experimental design, these time based trends have two effects. First, they "subsidize" experimentation at functional extremes. That is, they introduce trends in x_* (informationally myopic observations) which adds variance to the design matrix and reduce incentives for current experimentation. Operationally this is equilivent to increasing the effective rate of discount. The second effect is that prediction problems are exacerbated. Because of trends in x_*, yield response tends to be predicted at observational extremes where predictions are least reliable. Recall that the error variance of the predictions,

$$y_p = z_o'\hat{\beta} \text{ is } z_o'V(\hat{\beta})z_o = \frac{1}{T}(1+(z_o-\bar{z})'S^{-1}(z_o-\bar{z}))$$

where S refers to the covariance matrix of historic observations and $\bar{z}$ is the historic mean of these observations. The larger the discrepancy between z_o and $\bar{z}$, the larger is the prediction error.

To get an idea of the order of importance of these contradictory forces, consider a few simple models. First for a standard of comparison, consider a fully stationary process:

---

[1]I consider only trends in means of distributions of prior decision parameters here. If, for example, variances of these distributions expand with time so that the variance of x_* is increasing, then nature is subsidizing larger future experimentation and less experimentation is required in current periods. Alternatively, if variance of x_* is declining with time, larger current experiments are optimal. In each case, these trends in dispersion can be viewed as adjustments in effective discount rates.

$$y = a + b\,x + u$$

where there is no trend in technique (a and b) or in factors determining optimal activity, x_*. Suppose that on average we are interested in predicting y for values of x that differ from $\bar{x}$, the historic mean by one standard deviation. Prediction error variance, from least squares estimates,

$$y_p = \hat{a} + \hat{b}(\bar{x} \pm \sigma_x)$$

is

$$\Theta_1 = 2\,\frac{\sigma_u^2}{n}$$

where n is the number of observations on which the estimates, $\hat{a}$ and $\hat{b}$, are computed.

Now suppose that technique is static but that x_* is trended so that

$$x_t = t + V_t$$

where V_t is stochastically independent of time, t. The trend in x adds variance (since the mean grows only half as fast as the activity level), but since for prediction problems we are concerned only with extreme activity levels, there is a trade-off between the reduced coefficient estimator variances and the necessity of evaluating the estimated function at observational extremes. Define Θ_2 as prediction error variance at activity level, $x = t \pm \sigma_v$. In this case

$$\lim_{n\to\infty} \frac{\Theta_2}{\Theta_1} = \frac{7}{8}.$$

As sample size increases, predictions tend to be more accurate than in the fully static case. That is, the added variance in the historic observations reduces the variance of the coefficient estimates and this dominates the effect of the reduced reliability associated with prediction at observational extremes.

As a third alternative, suppose that x_x is trended, but that the function to be estimated is also trended, such that

$$x = t + V_t$$

(V is stochastically independent of t) and

$$a_t = a_o + a_1 t,$$

the intercept shifts with time. For prediction problems, this is the worst of possible worlds. Not only is prediction required at observational extremes, but the time dependent movement in x is not a source of independent activity variance for improving the reliability of the estimate, $\hat{b}$. Let Θ_3 indicate prediction error variance at

$$\hat{a}_t = \hat{a}_o + \hat{b}_o t \quad \text{and} \quad x_t = t \pm \sigma_v.$$

Here,

$$\lim_{n\to\infty} \Theta_3/\Theta_1 = 2.5$$

so that as the number of observations increases, this fully dynamic process results in larger prediction error than is associated with the static process.

In each of these cases, prediction errors decline as the number of observations increases and incentives for experimentation diminish through time. With the stationary case as a benchmark, it appears that the quasi-dynamic case in which activities are trended and technique is not results in reduced incentives to experiment. On the other hand, when both techniques and activities are trended incentives to experiment are enhanced[1].

---

[1] The omitted case corresponds to a trend in technique, $a_t=a_o+a_1t$ with no trend in activity. Evaluated at $\hat{a}_t = \hat{a}_o + \hat{a}_1 t$ with $x_t = x \pm \sigma_x$, prediction error is less than for the static case, but converges to that of the static case as the number of observations increases.

It is obvious that this view really begs the fundamental informational question of systems change. The point is that when a system progresses smoothly with time, there is a tradeoff between the informationally enhancing effects of increased activity variance and the errors associated with prediction at observational extremes. The most interesting question from an informational perspective arises when the system not only progresses with time, but is actually subject to change. For example in the earlier cases suppose that the intercept not only depends on time, but that the dependence is stochastic, i.e., $a_t = a_{t-1} + \varepsilon_t$.

More generally, consider the production process described by

(17.) $$y_{*t} = z_t' \beta_t + u_t$$

(17.a) $$\beta_t = A\beta_{t-1} + v_t .$$

where y_* and z refer to transformations of product and input that linearize (<u>vis a vis</u> state parameters, β_t) the process, u_t is an unobserved residual and A is the k by k transitional matrix (which for simplicity is presumed known) that along with the unobserved residuals, v_t k by k transform state t-1 parameters into current state parameters β_t. Viewing past observations in terms of current system technology, gives

$$y_{*t} = z_t' \beta_t + u_t$$

$$y_{*t-1} = z_{t-1}' A^{-1} \beta_t - z_{t-1}' A^{-1} V_t + u_{t-1}$$

and for the n-th previous observation,

$$y_{*t-n} = z_{t-n}' A^{-n} \beta_t - z_{t-n}' \sum_{i=1}^{n} A^{-i} v_{t-n+i} + u_{t-n}.$$

Estimation of β_t is obviously complex, even when A is known[1]. Not

[1] For a discussion of this problem with proposed solutions, see Deutsch ( ), chapter 8, 11, and 12.

only do more distant observations reflect todays technique with larger error, but the errors are serially correlated. This, however captures the essential ingredient of change, i.e., the obsolesence of knowledge. With respect to today's technique, yesterday's observation is blurred by technical change between yesterday and today. Similarly, the observation from the day before yesterday is twice penalized in counting for its value in estimating today's structure.

Presumably the generalized least squares estimates for this system are informationally efficient since they minimize $V(\beta_t)$ and therefore minimize the cost of ignorance.

The most significant distinction between this kind of approach to systems change and the cases discussed earlier, in which systems are either static or progress smoothly with time, is that in this view of dynamics incentives to gather information do not vanish through time. In terms of current systems states, the informational content of a given observation declines through time: Information is subject to obsolesence which creates an incentive for updating information.

Obsolesence of information implies that at any point in time there is less of a basis for forming estimates. This creates an incentive for current experimentation. But, since obsolesence renders past observations less relevant to today's system, future systems change makes todays observations less relevant in the future. This side of the obsciesence issue reduces the value of present experiments. The net effect of these contradictory forces is uncertain.

While it is clear that the existence of systems change, of the obsolesence of knowledge, provides incentives for steady state experimentation, and there is none in the static case, I am still unable to derive definitive results showing that even for simple models of obsolesence, experime

V. Loss Functions for Firms and Consumers

To this point the discussion has been couched in terms of a generalized yield function. In this section I present formulae for two cases of special interest to economists, the profit maximizing firm and the utility maximizing consumer. The idea is that as a result of imperfect information at the time activities are selected, mistakes occur. This section offers second-order approximations to the cost of a mistake, a sub-optimal activity, for a given state of the world.

a) The firm. The firm is assumed to err in two ways: first, at the observed or realized output level costs are not minimized, i.e., marginal rates of input substitution are not equal to marginal rates of input exchange; the second mistake is simply that output is not optimal, marginal cost is not equal to marginal revenue. Define π as the profit the firm earns from a given activity, a given choice of inputs, and define π^* as the maximum that could have been earned had the activity been optimal. We have then,

$$(18) \quad \pi = \pi^* + d\pi$$

where $d\pi$ represents the cost of incorrect choice. Let x_i indicate the quantity of the i^{th} input used when x_i^* is the optimum. And for the competitive firm let p indicate marginal revenue or product price and p_i, the price of the i^{th} input is also marginal factor cost. It follows that

$$(19) \qquad d\pi = \sum_i (p\bar{f}_i - p_i)(x_i - x_i^*)$$

where $\bar{f}_i$ is the average of the marginal physical product of x_i over the interval, $(x - x^*)$. Note that when x and x^* are not subscripted, reference is to the full input vector. The product, $(p\bar{f}_i - f_i)(x_i - x_i^*)$, is the average difference between value of the marginal product and input price times the size of the input error, $(x_i - x_i^*)$, i.e., it is the unit cost of the mistake multiplied by the size of the mistake. Since π is assumed concave in x, this product most be negative: when an input is underutilized, $x_i - x_i^* < 0$, the value of its marginal product exceeds its price, $p\bar{f}_i - p_i > 0$. For the second-order approximation, let

$$p\bar{f}_i - p_i = \frac{p}{2} d\, f_i \quad \text{with}$$

$$d\, f_i = f_i(x) - f_i(x^*) = \sum_j f_{ij}(x_j - x_j^*)$$

$$\text{where} \quad f_i(x^*) = p_i/P \quad \text{and} \quad f_{ij} = \frac{\partial^2 f}{\partial x_i \partial X_j} .$$

Then,

$$(20) \qquad d\pi = \frac{p}{2} \sum_i \sum_j f_{ij}(x_i - x_i^*)(x_j - x_j^*)$$

is quadratic in activity errors is the loss from $x \neq x^*$.

This, of course says nothing about the source of the errors, $x - x^*$. Presumably, errors occur because of incorrect price forecasts or because of uncertain technology so that it is convenient to express the cost of an activity mistake in these terms. In equation (19) let $\lambda^* = p_i/f_i(x^*) = p$ and define x' as the least cost input combination for producing the output consistent with x, i.e., $f(x)=f(x')$ with $p_i/f_i(x') = p_j/f_j(x')$ for all inputs, x_i and x_j. Substitute

$$x-x^* = (x-x')+(x'-x^*)= dx' + dx^* = dx$$

so that the movement, dx' corresponds to movement on the isoquant, $f(x)$, and dx^* refers to cost effecient changes in output between $y = f(x')$ and $y^* = f(x^*)$. Indicate $y-y^*$ as dy and $p= p-\bar{\lambda} + \bar{\lambda}$ with $\bar{\lambda}$ being the average level of marginal cost between y and y^*. This substitution into equation (19) gives

$$\text{(19.a)} \qquad d\pi = (p-\bar{\lambda}) \sum_i \bar{f}_i \, dx_i + \sum_i (\bar{\lambda}\bar{f}_i - p_i) dx_i .$$

But according to the definitions of dx_i', dx_i^*, and $\bar{\lambda}$,

$$\sum_i \bar{f}_i \, dx'_i = 0, \quad \sum_i \bar{f}_i \, dx_i^* = dy,$$

and

$$\sum_i (\bar{\lambda}\bar{f}_i - p_i) dx_i^* = 0.$$

So that

$$\text{(19.b)} \qquad d\pi = (p-\bar{\lambda}) dy + \sum_i (\bar{\lambda} f_i - p_i) dx_i' .$$

The first term on the right hand side of equation (19.b) is the production effect, the cost of producing the incorrect output and the second term is the cost effect, the loss resulting from the failure to minimize cost.

Since $p = \lambda^*$, $\lambda^*-\bar{\lambda}$ is approximated as $\frac{1}{2}\, d\lambda$ with $d\lambda = -\frac{d\lambda}{dy} dy$[1] so that the production effect is approximated as $-\frac{1}{2} \frac{d\lambda}{dy} \overline{dy}^{\,2}$. Introducing the elasticity notation $E_a = d \ln a$,

[1] The substitution of $d\lambda = -\frac{1}{2} \frac{d\lambda}{dy} dy$ is in recognition of the opposing nature of prediction errors on marginal cost and output. An overestimate of marginal cost reduces output.

$$dy = \frac{dy}{d\lambda}\left(p\frac{dp}{p} - \lambda \sum_i \frac{E\lambda}{Ep_i}\left(\frac{dp_i}{p_i}\right)\right)$$

where $\frac{dp}{p}$ is the percentage prediction error in product price and $\frac{dp_i}{p_i}$ is the percentage error between input price and marginal product. Also,

$$\frac{E\lambda}{Ep_i} = \frac{p_i}{\lambda}\frac{F_i}{F}$$ is the elasticity of marginal cost with

respect to the price of the i-th input with F being the determinant of the bordered Hessian, $\begin{bmatrix} o & f_j \\ f_i & f_{ij} \end{bmatrix}$, and F_i is the co-factor of f_i. Since $\sum_i f_i F_i = F$, $\sum_i \frac{E\lambda}{Ep_i} = 1$ and the term $\sum_i \frac{E\lambda}{Ep_i}\left(\frac{dp_i}{p_i}\right)$ is a weighted average of the percentage errors in input prices or marginal productivities with weights referring to the effects of these errors in predictions of marginal cost. Let

$$dy = \frac{dy}{d\lambda}\, p \left(\frac{dp}{p} - \frac{d\lambda}{\lambda}\right)$$

so that the production effect becomes

$$(19.b.1) \quad -\frac{1}{2}\frac{d\lambda}{dy}(dy)^2 = -\frac{py}{2}\frac{Ey}{E\lambda}\left(\frac{dp}{p} - \frac{d\lambda}{\lambda}\right)^2$$

where the percentage error in marginal cost

$$\frac{d\lambda}{\lambda} = \sum_i \frac{E\lambda}{\lambda}\left(\frac{dp_i}{p_i}\right)$$

and

$$\frac{Ey}{E\lambda} = \frac{F}{yF_o}$$

with F_o being the determinant of the Hessian, $\{f_{ij}\}$. Thus the production effect is proportionate to total revenue, py, and the squared percentage error between marginal revenue, p, and marginal cost, λ.

Now turn to the cost effect. With the approximation that f_i is linear over dx',

$$\sum_i (\bar{\lambda}\,\bar{f}_i - p_i) \doteq \frac{\lambda}{2}\sum_i df_i dx_i \ .$$

From Allen ( ), we have $\frac{E\,X_i}{E\,P_j} = k_j\sigma_{ij}$ so that

$$dx_i = \sum_j x_i\, k_j\, \sigma_{ij}\, (\frac{dp_j}{p_j})$$

with k_j, the expenditure share of x_j and σ_{ij} as the Allen-Uzawa partial elasticity of substitution between x_i and x_j. Since $p_i x_i = k_i C$ where C is total cost, the cost effect is rewritten as

(19.b.2) $$\sum_i (\bar{\lambda}\,\bar{f}_i - p_i)dx_i' = \frac{C}{2}\sum_i\sum_j k_i\, k_j\, \sigma_{ij}(\frac{dp_i}{p_i})\,(\frac{dp_j}{p_j})$$

where $\frac{dp_i}{p_i} = \frac{\lambda df_i}{\lambda f_i}$ is viewed as the percentage discrepancy between marginal factor cost and marginal factor productivity as in the expression for the production effect. Since $\sum_i k_i\,\sigma_{ij} = \sum_j k_j\,\sigma_{ij} = 0$ the cost effect can be written in an equivalent form by substituting $-\frac{1}{2}(\frac{dp_i}{p_i} - \frac{dp_j}{p_j})^2$ for $(\frac{dp_i}{p_i})\,(\frac{dp_j}{p_j})$. In this form, the cost effect is

(19.b.2)' $$-\frac{C}{2}\sum_i\sum_{j>i} k_i\, k_j\, \sigma_{ij}\, (\frac{dMRS_{ij}}{MRS_{ij}})^2$$

where $\sigma_{ij} = \sigma_{ji}$ and $\frac{dMRS_{ij}}{MRS_{ij}} = (\frac{dp_i}{p_i} - \frac{dp_j}{p_j})$ is the percentage discrepancy between marginal rates of substitution of x_j for x_i and their marginal rate of exchange.

b) The Consumer. Having derived an expression for the cost a firm incurs in not maximizing profit, the cost a consumer incurs in not maximizing utility is easily formulated. It is in fact only a restatement of the cost effect for the firm. The symmetry between the cost minimization problem given an output constraint and the utility maximization problem given an income constraint is obvious. In fact if U is defined as utility realized under goods selection x and U* is the maximum that could have been achieved with optimal selection x* given the budget restraint then the psychic cost of x-x* is

$$(21) \qquad dU = U^{*}-U \quad .$$

To translate this utility loss into a monetary term, divide it by, λ, the marginal utility of income so that dU/λ is a meauure of the income wasted by failing to optimize. Expansion gives

$$(22) \qquad \frac{dU}{\lambda} = \sum_i \left(\frac{\overline{U}_i}{\lambda} - x_i\right)(x_i - x_i^*)$$

where U_i is the marginal utility of X_i. Since $\sum_i p_i x_i = \sum_i p_i x_i^*$ the analogy to the cost effect is obvious. Therefore

$$(22.a) \qquad \frac{dU}{\lambda} \doteq \frac{I}{2} \sum_i \sum_j K_i \, \eta_{ij} \left(\frac{dP_i}{P_i}\right)\left(\frac{dP_j}{P_j}\right) .$$

where income, I, is substituted for total cost in equation (19.a.2) and $\eta_{ij} = k_j \, \sigma_{ij}$ is the income constant elasticity of demand for x_i with respect to P_j. Similarly, substitution in terms of discrepancies between marginal rates of substitution and marginal rates of exchange gives

$$(22.a)' \qquad \frac{dU}{\lambda} \doteq \frac{-I}{2} \sum_i \sum_{j>i} k_i \, \eta_{ij} \left(\frac{dMRS_{ij}}{MRS_{ij}}\right)^2 .$$

To go from these approximations of the cost of suboptimal activity to measures of the cost of ignorance requires that expected values be taken over the loss-weighted activities and the activity, X, that is selected to minimize expected cost, is the mean of X* in the loss-weighted densities. Clearly, this approximation to the cost of ignorance is a linear function of percentage error variances in prices and factor productivities, and the value of learning is a linear function of foreseen reduction in these error variances.

In the earlier sections of this chapter, the only behavioral issue is concerned with learning from experience. It is of course obvious that a lack of full knowledge has other behavioral implications. One important application of concepts of imperfect information is the search models forwarded by Stigler: actors perceive price dispersion and search for the best deal. In there simplest form, the search models are prcb- lems in order-statistics - the best deal is the minimum price obtained in n trials. More recently, Evenson and Kislev have extended the search framework to applied research in the biological sciences. Within a fixed genetic distribution researchers search for the "best" attributes in plants and animals.

The loss functions for firms and consumers presented here serve as a basis for extending behavioral implications of decisions under uncer- tainty. Here, I consider two special cases; one involving goods of uncertain quality and in the other case I consider the question of scale economy in using information.

A. Uncertain Quality

Unlike the general models for which the question of attitudes toward risk is foremost, the case of uncertain quality is fairly simple. This is because actors are definitionally risk adverse with respect to quality. The concavity of the production function guarantees that input quality dispersion reduces expected output and concavity of the income-constrained utility function quarantees that dispersion regarding goods quality lowers expected utility.

Thus for the consumer increased quality uncertainty lowers real income and the consumption of the consumption bundle shifts in favor of income inelastic goods. For firms, this effect is ambiguous at each level of input scale - corresponding to given cost as in the consumption case with fixed income - input composition shifts in favor of expenditure inelastic factors, but input scale can also be affected. For a competitive industry, quality uncertainty increases costs and therefore product price. If demand is elastic, industry revenue will decline and the associated reduction in input scale will reinforce the shift toward expenditure inelastic factors. But if demand is inelastic the increase in input scale associated with rising revenue will twart the tendency away from the expenditure elastic inputs.

For the linear homogeneous production function associated with classical perfect competition, all factors have unit expenditure elasticities and the question of input scale is one of industry revenue only.

There is also a question of composition effects for given scale since quality uncertainty may alter rates of substitution. These "substitution" effects depend upon (1) the relation between the quantity of a commodity used and the uncertainty associated with the content of the commodity and (2) the concavity or convexity of marginal functions. This dependence upon concavity or convexity of marginal functions is regretable because our economic intuition is poorly formed insofar as third-order derivatives are concerned. The first effect concerns the relation between quantity and uncertainty.

If a unit of commodity X has quality variance σ^2, then the variance of N units of this commodity is $N\sigma^2$ if the units are drawn independently and is $N^2\sigma^2$ if each unit is of the same quality as all others in the sample.

Consider a production process that is a known function of intrinsic input attributes, X_*. But assume that instead of purchasing these "goodies" the firm purchases "goods", X, of quality, q, where quality is goodies per good. According to the second-order approximation,

$$f(x_*|x) = f(x) + f'(x)(x_*-x) + \frac{1}{2}(x_*-x)'\{f_{ij}\}(x_*-x).$$

Add the assumptions, $E(q_i) = 1$ (goods and goodies are measured in the same units), $E(q_i^2) = \sigma_i^2$ /1 and $E(q_iq_j) = 1$ (qualities are independently distributed). Expected values are

$$E\, f(x_*|x) = f(x) + \frac{1}{2}\sum_{i=1}^{n} x_i^2 f_{ii}\sigma_i^2$$

for n-inputs. Expected marginal products,

$$MP_j = \frac{\partial Ef(x_*|x)}{\partial x_j} = f_j + x_j f_{jj}\sigma_j^2 + \frac{1}{2}\sum_i x_i^2 f_{iij}\sigma_i^2.$$

In this expression, the first term is the certainty equivalent marginal product, the second is the direct effect linking quantities and uncertainty and the third term involves the concavity or convexity of the marginal functions as noted above. For quadratic functions, third-derivatives vanish and the effect of quality uncertainty is simply a price effect; a good subject to quality uncertainty is seen simply as having a higher price than in the certain case. More generally the effect on the rate of substitution,

$$\frac{d\, MP_i/MP_j}{d\,\sigma_k^2} \sim f_jf_{kki} - f_if_{kkj};\ k\neq i\neq j$$

and

$$\sim \frac{1}{2}x_k^2(f_kf_{kki} - f_if_{kkk})\quad x_if_if_{ii};\ k \neq j = i.$$

/1 Notice that by assumption, quality is homogeneous within x. If quality were independently distributed, in the sample of x units variance would be $x_i\sigma_i^2$.

Presumably the third-derivative is an index of similarity between inputs i and k. If $f_{kki} < o$ an increase in x_i increases the rate of productivity decline associated with x_k so in a sense x_i substitutes for x_k. The impact of these secondary effects is to shift input composition away from inputs of uncertain quality and away from close substitutes in this restricted third-order sense of substitutes.

Interestingly enough for the two factor linear homogeneous process,

$$y = f(a,b)$$

the Euler equations give

$$a^2 f_{aa} = b^2 f_{bb}$$

so that the expected product function is

$$E(y|a,b) = f(a,b) + \frac{a^2}{2} f_{aa}\sigma^2; \; \sigma^2 = \sigma_a^2 + \sigma_b^2$$

Here, since marginal rates of substitution depend only upon input ratios, the origin of uncertainty is irrelevant: effects on expected output are symmetric between σ_a^2 and σ_b^2 and the effect of quality uncertainty on input composition is independent of origin. There is however a composition effect which for the constant elasticity of substitution function is

$$\frac{dMP_a|MP_b}{d\sigma^2} = \frac{(k_b - k_a)(1-\varepsilon)}{\varepsilon^2} .$$

Here, k_a and k_b refer to input shares and ε is the elasticity of substitution. An increase in uncertainty shifts composition away from inputs accounting for greater shares of cost when substitution is inelastic and toward these inputs in the substitution elastic case.

VI. Scale Economy

Many forms of information are environmental: they are like public goods in the sense that the use of information does not deplete the stock and the value of knowing that A is superior to B may be in direct proportion to the use of A or the non-use of B. In a situation in which the cost of information is independent of its intended use and value is proportional to use, scale economies are created.[1]

Define r_o as the <u>ceteris paribus</u> ignorance level that would exist at a point in time if no alternatives for purchasing information existed. The cost of this ignorance is presumed to be proportional to scale so that the firm's cost function is given as $C(y) = C^*(y)\ (1 + r_o)$ where $C^*(y)$ corresponds to the minimal possible cost for achieving output, y, and r_o is the fraction by which realized cost exceeds minimum cost. The cost of ignorance is $r_o C^*$. Now until this point ignorance, r_o, is treated as independent of the actions of the firm. But suppose that ignorance can be controlled-- that information can be purchased either explicitly by consulting with "experts," subscribing to trade journals, etc., or implicitly by experimentation and "visiting the friendly neighborhood research station."

Let $g(-r)$ represent a cost-of-information function such that $g'(-r) > o$ (marginal information costs are positive). The equilibrium conditions for purchasing information are $g'(-r) = \psi C^*$ and $g''(-r) > o$. The first states that the marginal cost of information (minus ignorance) is equal to marginal revenue. In this expression, ψ is a present-worth summary statistic

---

[1]This point is not original. Nelson (10) carefully delineates the size bias <u>vis a vis</u> a firm's R & D investments.

Education and Allocative Efficiency

This chapter describes a number of empirical studies of the sources of education's productivity in agriculture. The first examines determinants of relative wages among groups of workers with differing amounts of schooling. The main result is that the return to schooling is directly related to rates of technological changes.

The second study summarized here explores relationships between education and returns to farm sizes in the United States. Agricultural productivity advances appear to be closely related to growth in farm size. The arguement advanced is that technological growth creates scale economies in using information which alongside rising levels of farm operator education results in larger more efficient farms. The remaining sections report work completed by other. Nabil Khaldi uses an estimated aggregate agricultural production function to compare observed input mix to least-cost input combinations. His conclusion is that technological growth results in production uncertainty such that allocative efficiency declines as the rate of technological change rises. Alongside this, he finds that in a dynamic production environment more educated farmers are more efficient in selecting input bundles.

Fane uses a similar approach for a different body of data with similar results. Wallace Huffman explores the link between rates of response to reductions in Nitrogen fertilizer prices and farmer schooling. He concludes that more educated farmers respond more quickly to change.

The central theme of all of these studies is that education enhances allocative efficiency in an environment of rapid technological change.

The Productive Value of Education

Standard competitive theory, in its assumption of perfect information, rules out allocative ability as a source of the return to a factor. With complete information, there is no room for the concept of a superior alternative since in equilibrium all alternatives are equally good at the margin. That is, the perfect information assumption implies that the return to a factor is proportional to its marginal contribution to physical product. But, for education and some other intangibles, it is not clear that the direct contribution to physical production accounts for the total contribution to revenue. There have been attempts to modify the competitive model to allow for "entrepreneural capacity," but the return to this ability is almost always computed as a residual, total revenue less the cost of other things, which does not facilitate marginal analysis. Yet, firms clearly make marginal decisions vis-a-vis allocative abilities. They sometimes hire new "managers" and invest both in market and production information. As an alternative to computing marginal factor revenue as being proportional to marginal physical product in which all other things are held constant, I explore the implications of variations of an input (education) whose function, in part, is to vary the use of other inputs.

It seems plausible that the productive value of education has its roots in two distinct phenomena. Increased education simply may permit a worker to accomplish more with the resources at hand. This "worker effect" is the marginal product of education as marginal product is normally defined, that is, it is the increased output per unit change in education holding other factor quantities constant. On the other hand, increased education may enhance a worker's ability ability to acquire and decode information about costs and productive characteristics of other inputs including, perhaps, the use of some "new" factors that otherwise would not be used. The return to education is therefore considered as consisting of two effects: a "worker effect" and an "allocative effect."

In recent years, we have stressed the importanct of education as a factor of production and have included it, often as an adjustment for quality of labor, as a variable in estimates of production functions. Consider three "production functions" to distinguish the role of education in each: (1) the engineering production function of a single commodity; (2) a production function of gross sales; and (3) a production function of value added by some subset of factors supplied by the firm or industry, the other inputs being "purchased." As I show, when the marginal product of education is treated as a partial derivative, the composition of the bundle of "other things" held constant is crucial.

In each case, production is assumed to be technically efficient in the sense that for given inputs, physical output is maximized.

For the engineering function, we have

$$Q = q(X,E),$$

where Q, physical output, is a function of education, E, and other inputs, X. In this case, the marginal product of education is $\partial q/\partial E$ and refers only to the worker effect. As noted earlier, it refers to the ability to accomplish more (physical output), given the resources at hand. By including education or "knowledge" as an explicit factor of production, the concept of technical efficiency becomes something of a tautology. Production is technically efficient if producers do not knowingly waste resources. If they wast resources but are ignorant of doing so, the loss is attributed to a lack of knowledge. Presumably, the worker effect is related to the complexity of the physical production process. In the engineering function there is no room for allocative ability, since questions of allocation do not arise. In the remaining functions, education is excluded as an explicit factor. To include it would only reiterate the worker effect which is obvious in the engineering function.

Now consider gross sales for firms producing more than one product. With two commodities, we have,

$$Q = p_1 q_1(x_1) + p_2 q_2(x_2),$$

where p_1 and p_2 refer to the prices (assumed exogenous to the producer) of the respective commodities, q_1 and q_2. Both commodities are assumed to be functions of the input vector, X. The quantity of X used in producing q_1 is denoted by x_1 and similarly for x_2. In this case, assume that x is given, but that its allocation among competing uses x_1 and x_2 is not. Here technical efficiency refers to being on the product transformation frontier, that is, of maximizing q_1, given q_2 and X, and does not correspond to maximization of sales, Q, given X. To maximize Q, we have

$$\frac{\partial Q}{\partial x_1} = p_1 \frac{\partial q_1}{\partial x_1} - p_2 \frac{\partial q_2}{\partial x_2} = 0$$

as the first-order condition. Maximization of sales requires technical efficiency and that the marginal value product of x be equated between its competing uses. Suppose that productive capacities of some factors are not equally understood by all who use them so that Q, given X, is not necessarily maximized. Suppose further that the allocation of X among its alternatives is a function of education, that is, $x_1 = x_1(E)$. In this case, the marginal product of education,

$$\frac{\partial Q}{\partial E} = \left(p_1 \frac{\partial q_1}{\partial x_1} - p_2 \frac{\partial q_2}{\partial x_2}\right) \frac{dx_1}{dE}$$

is positive if education enhances allocative ability. Thus when education is treated as a factor in functions producing gross sales, if the allocation of inputs among alternatives is not an explicit part of the function, we have the inference that the marginal product of education includes gains in allocative efficiency as well as the worker effect.

In considering value added, assume that there is only one product. Nothing is gained by multiple products since the question of allocation between competing uses is obvious in the previous example. Value added is expressed as $Q = pq(X,Z) - p_x X$, where p refers to commodity price and q, physical product, is a function of purchased inputs, X, and inputs supplied by the firm or industry, Z. The prixe of X is p_x, and both p_x and p are assumed exogenous to the producer. Here, maximization of Q with respect to X gives $\partial Q/\partial X = p(\partial q/\partial x) - p_x = 0$, which is the marginal productivity theory, that is, in equilibrium the value of the marginal product of X should equal its price. When Q is maximized with respect to X, we have the inference that value added is a function of Z only. But, again assume that producers are not equally adept at assessing productivity and that the quantity of X purchased is a function of education. In this case, the marginal product of education,

$$\frac{\partial Q}{\partial E} = (p\,\frac{\partial q}{\partial x} - p_x)\,\frac{dX}{dE}\,.$$

Here, the question of the ceteris paribus bundle is obvious. If a production function of value added is estimated and X is introduced as an explanatory variable, a positive marginal product denotes underutilization (at the mean) and vice versa for a negative marginal product. Alternatively, X can be excluded as an explicit variable and education included, in which case the marginal product of education reflects conovement between X and E with any resulting allocative gains or losses. Thus, if a value-added function, based on multiple products, is estimated which specifies the quantity of supplied inputs, Z, but does not specify allocation among competing uses, and if purchased inputs are omitted, the marginal product of education will contain three elements. First is the worker effect, then there is the question of selecting the quantity of other inputs, and finally the allocation of these inputs among their alternatives.

These effects can be combined by considering a value-added production function in which there are two products produced, q_1 and q_2, and each is a function of three inputs: education, E; other inputs supplied by the firm, Z; and purchased inputs, X. The respective commodity prices are p_1 and p_2, and p_x is the price of X. We have value added by education and other supplied inputs,

$$Q = p_1 q_1(x_1, z_1, E_1) + p_2 q_2(x_2, Z_2, E_2) - p_x X.$$

Where $E = E_1 + E_2$, $Z^o = z_1 + z_2$, $X = x_1 + x_2$; and

$$1 = \frac{dE_1}{dE} + \frac{dE_2}{dE}, \quad 0 = \frac{dz_1}{dE} + \frac{dz_2}{dE}, \text{ and } \frac{dX}{dE} = \frac{dx_1}{dE} + \frac{dx_2}{dE}.$$

If value added is taken as a function of the total quantities of education and supplied inputs, $Q = f(E, Z^o)$, the marginal product of education,

$$\frac{\partial f}{\partial E} = p_2 \frac{\partial q_1}{\partial E} + (p_1 \frac{\partial q_1}{\partial E} - p_2 \frac{\partial q_2}{\partial E}) \frac{dE_1}{dE} + (p_1 \frac{\partial q_1}{\partial z} - p_2 \frac{\partial q_2}{\partial z}) \frac{dz_1}{dE}$$
$$+ (p_1 \frac{\partial q_1}{\partial x} - p_2 \frac{\partial q_2}{\partial x}) \frac{dx_1}{dE} + (p_2 \frac{\partial q_2}{\partial x} - p_x) \frac{dE}{dE}.$$

Where the first term is the "own" value of the marginal product of education, the worker effect, the next three terms refer to the gains from allocating the respective factors, education, supplied inputs, and purchased inputs efficiently between competing uses, and the last term refers to the allocative gain from selecting the "right" quantity of purchased inputs, X. If this were a production function of sales, dX/dE would equal zero and the effects of selecting the input bundle would be lost.

Since returns to education include allocative ability, estimates of the productive value of education should include a provision for these returns. This can be done explicitly by the use of total derivatives or implicitly through profit or value-added functions.[1] Production functions of gross revenue include the worker effect and the effect of allocating factors between competing uses but exclude the effect of selecting the "right" quantities of other inputs. Engineering production functions reveal only the worker effect.

Perhaps this helps reconcile the inconsistency between the estimates of Griliches (1964) and Kislev (1965) of the productive value of education in agriculture. Using similar data, both estimated an agricultural production function of gross revenue for 1959. At the state level of aggregation, Griliches found schooling to be an important source of productivity, whereas Kislev, working with county data, found little or no return to schooling. The level of aggregation may be the key to understanding the difference in their estimates. While both used gross revenue as their measure of output and therefore permit education to capture the gains from allocating resources between competing uses, it is clear that agriculture at the state level is much more diversified, vis-a-vis product, than at the county level. Thus the state aggregate permits more "room" for allocative ability than does the county. It is

[1] Value-added functions become profit functions in the extreme as all inputs are treated as purchased (variable).

also clear that to the extent that education affects the choice of which inputs to use, both Griliches and Kislev understated the productive value of education because both held other input quantities (including purchased inputs) constant in estimating marginal productivity.

This also helps interpret a peculiar result of an earlier attempt of my own (Welch 1966) to analyze the determinants of the value of schooling in U. S. agriculture. The dependent variable in my analysis, the return to eight years of schooling, was assumed to be the value of the marginal product (as a partial derivative) of schooling in agriculture. Clearly, when the return to schooling is estimated from wages, it includes gains to allocative ability. The coefficient estimates indicated that the share of labor in agriculture is about three-fourths of total output and the share of nonlabor inputs is one-fourth; factor shares more relevant to value-added than to gross sales.[2]

The first clear-cut distinction between worker and allocative effects is provided by Chaudhri (1968). In trying to assess the impact of education on Indian agriculture, he estimates an aggregate production function (at the state level) of gross revenue. Although statistical problems of few observations and large error-variance in coefficients preclude strong statements, Chaudhri fails to demonstrate that education is an important source of productivity. He argues (partly in error) that, in his estimated function, marginal product of education refers to the worker effect alone, and to capture the allocative effect he provides evidence showing

---

[2] I failed to recognize the significance of this result, and although I referred to the underlying production process as one of value added, the measure of nonlabor inputs included purchased inputs.

in case after case that the composition of the "other" input bundle varies with the incidence of secondary schooling in the farm labor force. His conclusions reinforce the Nelson and Phelps (1966) contention that education enhances innovative ability as he demonstrates that the use of modern, as opposed to conventional, inputs is positively related to education. In this context, innovative ability is one dimension of allocative ability.

The Case of Agriculture

The empirical analysis of factors determining the productivity of schooling is restricted to agriculture in the United States. There are two good reasons for doing so. First, the data are fairly accessible. The effort of others, particularly Griliches in his work on the sources of measured productivity growth (1963a, and 1963b, 1964) shows many relevant considerations; also the work of Evenson (1967, 1968) refines some aspects of the original Griliches measures. The ability to build on this kind of empirical foundation does not exist outside of agriculture. The second reason is that U.S. agriculture is highly dynamic technically. The well-known concept of the farmer on the treadmill places peculiar emphasis upon innovative effort. A rapid rate of "technical change" together with an inelastic aggregate product demand implies that there is continual pressure on some factors, particularly labor, to leave agriculture, and the ability to stay current with respect to productive techniques determines whether a firm will exist in the long run. While these factors favor the selection of agriculture, there is one shortcoming.

Agriculture is probably atypical inasmuch as a larger share of the productive value of education may refer to allocative ability than in most industries. Farming usually includes a diversified set of activities for which allocative decisions are made continuously as part of the normal routine. In other industries, the jobs of a large portion of the work force do not involve decisions for which prices are relevant. Too, in most industries jobs performed by persons with different education are more sharply differentiated than in agriculture, and, in these cases, the physical productivity of education is more easily understood. In agriculture, differences in job complexity associated with differences in education are less noticeable, and the product of education is more likely to be associated with allocative efficiency. Does education enable one to pick more grapes or do a better job of driving a tractor? Even if it does, these "worker" effects are probably small when compared with the considerable differences revealed in income. Allocative ability plays a key role in determining education's productivity in agriculture and is most relevant in a dynamic setting.

The relevance of dynamical factors is stressed by Schultz (1964) when he suggests that, in economies in which agricultural production is accomplished almost soley by the use of "traditional" factors, there is reason to believe that factors are more efficiently allocated than in "modern" agricultural economies. Schultz's interpretation is that traditional agriculture is close to an economic equilibrium in adjusting to relatively stationary techniques. Because of this, judgments about factors are based upon extensive observation; the stationary technology guarantees ample time to explore the potential of factors being used.

In contrast, in a technically dynamic agriculture, a factor may be obsolete before its productivity can be fully explored. Herein, I think, lies the explanation of education's productivity. If educated persons are more adept at critically evaluating new and reportedly improved input varieties, if they can distinguish more quickly between the systematic and random elements of productivity responses, then in a dynamical context educated persons will be more productive. Furthermore, the extent of the productivity differentials between skill levels will be directly related to the rate of flow of new inputs into agriculture.

In the empirical analysis that follows, I concentrate on determinants of relative wages among three skill classes: college graduates, high school graduates, and persons with one to four years of schooling (functional illiterates). As is described in Appendix A, laborers in intermediate classes are treated as linear combinations of persons in these three classes. For example, a person with eight years of schooling is "counted" as 0.46 of a person with one to four years of schooling, 0.53 of a high school graduate, and so on for the schooling classes: 0, 5-7, 9-11, and 13-15 years. Wages refer to males 45-54 years old, and the associated number of persons in each skill class refers to a white male, 45-54 years old, equivalent wage earner. It is assumed that a laborer's wage is his marginal product from an underlying production process of value added by labor and inputs supplied by farms. Nonlabor inputs include a measure of the flow of services from land, machinery, and livestock inventories. Purchased inputs are excluded.

The measure that I use for the rate of flow of new inputs is a weighted average of expenditures per farm for research over the past nine years. As a measure of the availability of information about new inputs, I include an average over the past four years of the number of days spent (per farm) on farms by state and federal extension staff.

The unit of observation is the "state" of which there are forty-nine. The thirty-nine "states" used by Griliches (1964) are included together with a breakdown of ten Southern states[3] into white and nonwhites. The white-nonwhite specification is in recognition of the segregation of the Federal Extension Service prior to 1962. While factor ratios and research are treated as the same level in the white and nonwhite sector of each of the ten states, the extension variable refers to days spent on white and nonwhite farms, respectively.[4]

Table 1 provides estimates of factors affecting relative wages in agriculture. Although coefficients on inputs are reported separately, the equations are estimated using factor ratios so that the sum of the coefficients on the inputs is constrained to equal zero. All variables except research and extension are in logarithms.

In Table 1, regression equation (1) and (2) raise as many questions as they answer. Notice that in each equation the coefficient estimates indicate that the relative productivity of a college graduate is increased by increasing the number of college graduates, although not "significantly" so. There is another peculiarity of these two equations. In equation (1) an increase in the number of functional illiterates reduces the wage of high school relative to college graduates. In equation (2) an increase in the number of high school graduates increases the wage of functional illiterates relative to college graduates. That is, high school graduates . . .

[3]The states are Alabama, Arkansas, Georgia, Louisiana, Mississippi, North Carolina, South Carolin, Tennessee, Texas, and Virginia.

[4]For an enlightening discussion of the segregation of the Federal Extension Service, see U.S. Commission on Civil Rights (1965).

Table 1

Estimates of Factors Affecting the Productivity of More Relative to Less Schooled Persons in U.S. Agriculture, 1959

Independent Variables	Dependent Variables			
	W_{16+}/W_{12}	W_{16+}/W_{1-4}	W_{12}/W_{1-4}	
	(1)	(2)	(3)	(4)
1. Functional illiterates	.054 (.037)	.442 (.048)	.388 (.039)	.359 (.038)
2. High School Graduates	.034 (.084)	-.426 (.109)	-.460 (.089)	-.359 (.038)
3. College Graduates	.048 (.064)	.062 (.083)	.014 (.068)	....
4. Nonlabor inputs	-.136 (.034)	-.078 (.045)	.058 (.036)	...
5. Research expenditures ($00) per farm	.056 (.094)	.179 (.122)	.123 (.100)	....
6. Day per farm by extension Personnel	-.130 (.078)	-.136 (.101)	-.006 (.082)	....
7. Nonwhite	.138 (.043)	-.122 (.056)	-.260 (.045)	-.281 (.045)
8. Intercept	1.742	2.008	.266	.744
R^2	.578	.737	.722	.671
Residual sum of squares (degrees of freedom)	.436 (42)	.737 (42)	.486 (42)	.574 (46)

Note. Subscripts indicate years of school completed. Standard errors of the coefficient estimates are in parentheses. Variables other than research, extension, and nonwhite (a dummy variable) are in logarithms.

Definitions of Variables:

1. Wages. - Wages refer to total income in 1959, for males with income, age 45-54 years. The estimation procedure is discussed in the Appendix.

2. Number of persons. - The number of persons in each schooling class is estimated in terms of age constant, white, male equivalent earners. The estimation procedure and the procedure for reducing the schooling distribution to three classes is described in the Appendix. The procedure described there results in estimates of number of persons in the rural farm population, not in the farm labor force per se. To adjust for this overstatement of numbers of persons, the number in each schooling class was multiplied by the ratio, R, for each state. Where R = number of employed male farmers, farm managers, farm laborers, and foremen in 1960 divided by the number of rural farm males employed in 1960 (1960 Census of Population, tables (s) 121). The national average for this ratio is .965.

3. Nonlabor inputs supplied by farms. - A linear aggregate of the estimated flow of services from land and buildings, machinery, and livestock inventories. The measure includes:

a) 3 percent of the value of farm land and buildings. The land and buildings variable is taken from Griliches (1964). His series adjusts for differences in quality of crop, irrigation, pasture land, and so forth, using relative prices for 1940. A uniform price index adjustment was used by Griliches to express values in 1949 dollars, and I mjltiplied his values by 1.69 (the ratio of the 1959-1949 price indexes for farm real estate) so that value is in 1959 dollars. The 3 percent refers to an assumed 8 percent competitive rate of return (a la Griliches) and an assumed 5 percent rate of appreciation in farm real estate values.

b) 15 percent of the value of machinery on farms. A measure of the value of machinery on farms is constructed using price indexes supplied by Kislev (1965) and Census of Agriculture estimates of the stock of machines on farms. The 0.15 refers to an assumed average machine life of ten years and an 8 percent rate of interest.

c)8 percent of the value of the livestock inventory: U.S. Census of Agriculture, 1959 estimates.

4. Research per farm. - A weighted average of total research expenditures refer to all federal and state expenditures by Census of Agriculture number of farms. Research expenditures refer to all federal and state expenditures (including farm management research). These data are provided by Evenson (1968). The annual weights are: .04, .08, .12, .16, .20, .16, .12, .08, and .04 for the years 1959 to 1951, resectively.

5. Days per farm by estension personnel: A weighted average over four years of the total days spent by extension personnel in eight selected activities divided by Census of Agriculture number of farms. The included activities are: crops, livestock, marketing, soils, planning and management, land, buildings and machinery, and forestry. These data are obtained from unpublished reports of the Federal Extension Service. The annual weights are: 1/3, 1/3, 1/6, 1/6 for the years 1959 to 1956, respectively. For the Southern states, days per farm is computed from separate statistics for the Negro and white extension services, and number of farms similarly refers to Negro and white farm operators.

In the Southern states, the number of persons in a schooling class refers to the number of whites plus the product of the number of nonwhites and the relative wage of nonwhites. This number is interpreted as "white" equivalent laborers.

enhance the relative productivity of functional illiterates, but functional illiterates detract from the relative productivity of high school graduates. Strictly speaking these results are not necessarily contradictory because the coefficients refer to the elasticity of the relative wage of two factors with respect to a third and are not transitive. Nevertheless, this result is contrary to the usual interpretation of factor substitutability and questions the form in which equations (1) and (2) are specified.

Before discarding these two equations, consider the remaining results. In each case, nonlabor inputs supplied by farms detract from the relative productivity of college graduates. Too, the evidence is that research activity, the rate of flow of new inputs, enhances the realtive productivity of college graduates and that extension activity, the flow of information about new inputs, detracts from the relative productivity of college graduates. This is as it should be. If education enhances the ability of a producer to decode information about the productive characteristics of new inputs, then the more rapid the rate of flow of new inputs, the greater will be the productivity differential associated with additional education. Further, if the advantages associated with added education refer to a differential ability to acquire and decode information, then an activity of disseminating information (extension) can short-circuit the gains to education. In a sense, the Extension Service may serve the purpose of overcoming the disadvantages associated with insufficient schooling. Unfortunately, the effect of extension seems more apparent than real, inasmuch as in considering a (hopefully) superior specification of equations (1) and (2), the effect of extension disappears.

Regression equations (3) and (4) provide the most valuable information of Table 1, In fact, equation (3) is implied by equations (1) and (2) and can be calculated simply as equation (2) less (1). Nevertheless, when (3) is compared with (4), the evidence is that the wage of high school graduates relative to functional illiterates depends neither upon the number of college graduates, the quantity of nonlabor inputs, research, nor upon extension activities. Equation (4) is estimated with the constraint that the coefficient on each of these four variables is zero. Deleting these variables reduces R^2 by only .051 (from .722 to .671), indicating that the partial R^2 of these four variables with the relative wage is .16. In testing for the joint significance of these variables, the the computed $F_{(4,42)}$ statistic is 1.90, whereas the associated critical value of F at a confidence level of .05 is 2.59.[5] I therfore accept the hypothesis that the marginal rate of substitution (assumed equal to the relative wage) of functional illiterates for high school graduates is a function of the ratio of high school graduates to functional illiterates only. Under this hypothesis the coefficient on the factor ratio, .359, can be interpreted as an estimate of a special kind of elasticity of substitution. It is not the partial elasticity in its most general form, but is is the elasticity of the factor ratio with respect to the marginal rate of sbustitution. The point estimate of the elasticity of substitution is 2.8 (the reciprocal of .359). The evidence here is that the elasticity of substitution between these classes is significantly different from unity so that the commonly used Cobb-Douglas and linear forms for combining inputs seem inappropirate.

---

[5] The test that a subset of coefficient in a regression equation is equal to zero is given in Graybill (1961, pp. 133-40).

Solow (1956), a la the Leontief separability theorem, has pointed out that, if the marginal rate of substitution between two inputs is independent of other inputs, production can be considered as a multistage process in which the two are first combined into an intermediate good which is then combined with other inputs to form the final product. The evidence here is that functional illiterates and high school graduates can be aggregated into an intermediate good. Call it "conventional labor" in contrast to the "modern" skills acquired in college. If the production process is viewed as a nested C-E-S function of the form suggested by Mundlak and Razin (1967) and if high school graduates and functional illiterstes belong in the same subaggregate, then equation (4) is correctly specified, and equations (1) and (2) are not.

Table 2 provides estimates of regressions when high school graduates and persons with one to four years of schooling are aggregated using the C-E-S form into conventional labor, CL, and the wage is estimated as the average cost of CL. The aggregate is:

$$CL = (\delta N_{12}^{-\beta} + (1 - \delta)N_{1-4}^{-\beta})^{-1/\beta}$$

$$1 + \beta = .359$$

$$\log_e(\frac{\delta}{1-\delta}) = .744,$$

where N_{12} N_{1-4} indicate the numbers, respectively, of high school graduates and functional illiterates. The wage of the aggregate is computed as

$$W_{CL} = \frac{W_{12}N_{12} + W_{1-4}N_{1-4}}{CL} .$$

These results appear superior to those provided in equations (1) and (2) of table 4. The major change in specification is the form of the labor variables, and coefficients on labor inputs have much smaller

Table 2

Estimates of Factors Determining the Productivity of College Graduates Relative to Laborers with Conventional Skill, an Aggregate of Functional Illiterates and High School Graduates, in U. S. Agriculture, 1959

Independent Variables	Dependent Variables, the Relative Wage (1)	(2)	(3)
1. The aggregate of functional illiterates and high school graduates	.699 (.062)	.699 (.061)	.711 (.081)
2. College graduates	-.377 (.084)	-.377 (.084)	-.711 (.081)
3. Nonlabor inputs	-.322 (.056)	-.322 (.055)	... ...
4. Research expenditures ($00/farm	.485 (.163)	.482 (.150)	.663 (.194)
5. Days per farm by extension personnel	-.009 (.148)	...	...
6. Nonwhite	-.637 (.084)	-.637 (.079)	-.540
7. Intercept	1.167	1.157	-2.314
R^2	.803	.803	.648

Note. - Standard errors are in parentheses. Inputs and wages are in logarithms. The equations are estimated subject to the constraint that the coefficients on inputs (excluding research and extension) sum to zero. For definitions of variables, see the notes to table 4.

standard errors relative to the estimated coefficients than in the earlier equations.

If the average cost of the combined high school graduates and functional illiterates is considered as the marginal product from an underlying production process, then the relations between the wages (marginal products) of high school graduates and functional illiterates are given as:

$$W_{12} = W_{CL} \frac{\partial CL}{\partial N_{12}}$$

and

$$W_{1-4} = W_{CL} \frac{\partial CL}{\partial N_{1-4}}$$

From this it follows that equations (1) and (2) of table 4 should be:

$$\frac{W_{16+}}{W_{12}} \quad \frac{W_{16+}}{W_{CL}} \cdot \delta \left(\frac{N_{12}}{CL}\right)^{1+\beta}$$

and

$$\frac{W_{16}}{W_{1-4}} = \frac{W_{16+}}{W_{CL}} (1-\delta) \left(\frac{N_{1-4}}{CL}\right)^{1+\beta}$$

Since the function W_{16+}/W_{CL} is estimated in table 2 and $(1+\beta)$ and δ are estimated in equation (4) of table 1, estimates of these relative wage equations are easily derived. Too, when viewed this way, the misspecification of equations (1) and (2) of table 1, is obvious. The variable, conventional labor, is simply left out.

Thus, table 2 together with equation (4), of tablel, provides an internally consistent set of estimates of factors determing relative wages in agriculture.

In table 2, equation (1) indicates that extension activities are not important in determining relative wages, and equation (2) presents the regression estimates when this factor is deleted. It represents my "best" estimate of the equation. The third column also excludes nonlabor inputs to provide an estimate of the long-run (since other factors are left free to vary) elasticity of substitution between college graduates and conventional labor. That estimate is 1.41 (the reciprocal of .711).

One important by-product of these estimates is that the marginal rete of substitution between college graduates and other labor appears to be significantly related to the quantity of nonlabor inputs. As such, it does not appear that all forms of labor can be aggregated into a single input.

From regression equation (2) of table 2 the coefficient on research can best be interpreted by asking the question: What would happen to the relative wage of college graduates if the research variable were to become zero? At the samples geometric mean, the wage of college graduates relative to high school graduates is 1.62, and relative to functional illiterates it is 1.75. The sample mean value of research expenditures per farm is $26.74. If this value were to fall to zero, the estimate here is that the relative wage in each case would fall by about 14 percent or to 1.39 and 1.50, respectively. Thus, about one-third of the productivity differential between college graduates and either high school graduates or functional illiterates is directly attributable to research. In fact, this is probably an understatement of the impact of research. If production were to become technically static, eventually the productive characteristics of all inputs would become fairly well understood. This common information would be passed by word of mouth from one generation of farmers to the next, and under such conditions it is difficult to understand how education could enhance allocative efficiency. In a dynamic setting discretionary abilities may be the key to allocative efficiency. In a static setting these abilities seem unimportant.

That the partial effect of nonlabor inputs supplied by farms is negative (with respect to the relative wage of college graduates) is consistent with many explanations that cannot be distinguished with these data.[6] It is likely that the productive characteristics of land, buildings, machinery, and livestock (the supplied inputs) are more commonly understood than are the characteristics of the purchased inputs, seeds, commercial fertilizers, pesticides, and so forth. If true, we would expect farmers who are adept at assessing the productivity of modern inputs to rely more heavily on them, that is, to allocate a larger share of their input bundles to modern inputs than would farmers less certain of the capacities of modern inputs; and if this is true, a plausible explanation of the negative effect of supplied inputs is that, for college graduates, the productivity gains associated with increments in supplied inputs are less than proportionate to relative wages. Recall that, at the sample mean, the relative wage of college to high school graduates is 1.62. For supplied inputs to be neutral between high school and college graduates, the rate of increase in pro-

[6] One possibility is that, given existing factor and product prices and the quantities of labor inputs, there will exist a corresponding combination of nonlabor inputs that is optimal. Call that combination K^*. If we assume that the farm enters the production period with K of these inputs (K denotes the inputs supplied by the farm), then $K^* - K$ remain to be purchased. If more educated persons possess superior allocative ability, the gains to this ability will be positively related to the "room" for selecting inputs, that is, to $K^* - K$, and will therefore be negatively related to K. This, of course, is true only if K^* is independent of K, which seems unlikely, and if the "superior" allocative ability is superior only in the short run. Otherwise, this ability would be reflected in the previous selection of K.

ductivity of college graduates from increments in supplied inputs would have to be 1.62 times the corresponding rate for high school graduates. The evidence here is that, while it is possible that an increment in supplied inputs increases the productivity of college graduates by more in absolute terms than for less schooled persons, the ratio of the absolute increments is less than the prevailing relative wages. College graduates presumably get the added leverage (that results in relative wages being what they are) through the use of other inputs.

Since the effects of extension activities appear neutral, that is, they do not seem to alter relative wages, a question arises as to whether the nonwhite"states" should be included in the analysis. This is because extension is the only variable other than the dummy for the "level effect" distinguishing the Southern white and nonwhite observations. To "test" for the sensitivity of the coefficient estimates to the nonwhite observations, the regressions presented in tables 1 and 2 are estimated for the thirty-nine white "states" and are provided in the Appendix in tables 4 and 5. There is marked agreement in the two sets of estimates.

Summary

The role of education in production is stressed here, showing that, while it can be considered as any other factor in the sense that it may directly contribute to physical product, the effects of allocating other factors must also be recognized. If education enhances allocative ability in the sense of selecting the appropriate input bundles and of efficiently distributing inputs between competing uses, the return to this ability is part of the return to education. The empirical analysis refers to determinants of relative wages in agriculture.

The evidence is that, while factor ratios are important, much of the "leverage" associated with added schooling is drawn from the dynamical implications of changing technology. But this appears to hold only for skills that result from college. Relative wages for persons who have not attended college are determined by labor ratios only.

Consider the effects of research on the return to college education. Research expenditures per farm were \$4.30 in 1940 and \$28.40 in 1959.[7] Based on coefficient estimates in table 2, if research were to fall from \$28.40 to \$4.30, holding factor ratios constant, the relative wage of college to high school graduates would fall from 1.62 to 1.43, indicating that about one-third of the wage differential would disappear. Too, purchased inputs have become relatively more important. In 1939, inputs purchased from other industries accounted for 38 percent of agricultural output, and by 1959 the share had increased to 48 percent.[8] This trend should have increased the role of the innovator-allocator.

[7] Constant 1959 dollars (Evenson 1968).

[8] Agricultural output excludes the intermediate goods, feed and livestock (Welch 1969b).

II. Returns to Scale in U.S. Agriculture

One result common to virtually all estimates of aggregate production functions for U.S. agriculture is that larger farms are more efficient than smaller ones.[9] In Griliches' (3) accounting for growth in measured productivity, 1940-1960, more than one-half of the total gain is attributed to increased farm size. There are a number of *ad hoc* arguements that can explain coincident productivity growth and economies of scale. For example: technical change is embodied in large, indivisible inputs -- the productivity of machinery increases more rapidly with size than do machine costs. Also, integration of many activities within a single managerial unit can reduce transaction cost. Finally, there is a systematic relationship between farm size and the activity mix. Possibly productivity advances have most effected products and inputs on which larger firms are more dependent. But, is there any *prima facie* reason to expect these relationships?

An alternative view is that scale economies exist and persist as part of the purely technical relationships of farm production. These economies are not a byproduct of productivity advance; instead they are the vehicle of this advance. But if these economies have always existed why have farmers waited so long to capture the returns to size expansion?

Yet another view is that productivity change, changes in the rules of the game, create production uncertainty and there are scale economies

---

[9] See the Griliches summary (4), Huffer (5), Madden (8) and Tweeten (12). Kislev (7), in the introduction of his paper, offers an interesting summary of the scale evidence for micro (farm-level observations) and macro (state agregates) studies.

in using information. A model of this sort is described in the preceding chapter. Because of technical change, actual costs of production exceed minimum possible costs. Excluding apportunities for acquiring information, the percentage descrepancy between actual and minimum cost is independent of scale. That is, since the relative cost of ignorance is independent of scale, the absolute cost is proportionate to scale. When opportunities for acquiring information are considered, if cost-of-information is independent of scale it is clear that it will pay larger firms to purchase more information. Further, for the combined cost function which adds information cost to explicit production cost, the ratio of average to marginal cost is greater at every level of output than for either the "most efficient" cost function or for the cost function that ignores information options. This situation is depicted in figure 1, where . . .

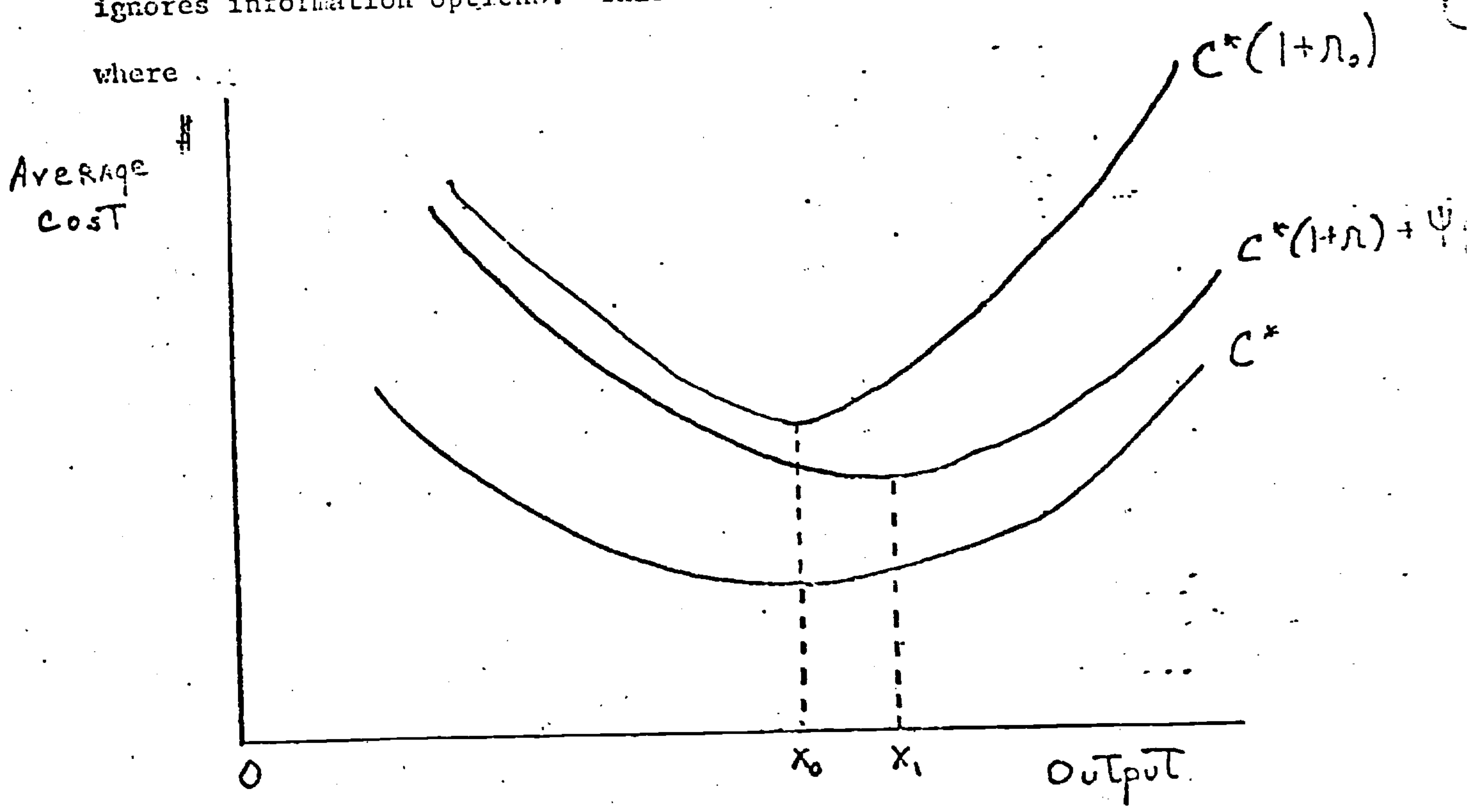

Figure 1

C^* represents average cost when costs are minimized, r_o is the <u>ceteris parabus</u> level of ignorance - the proportionate discrepancy between minimum and actual cost that would exist if there were no opportunities for acquiring information. The full cost function, $C^*(1+r) + \psi\, g(-r)$, includes the current cost of information $\psi\, g(-r)$ and realized production cost $C^*(1+r)$. The C^* and $C^*(1+r_o)$ functions are minimized at x_o and the composite cost function is minimized at output, x_1.

Suppose that a technological innovation shifts the cost function for all firms in a perfectly competitive industry downward from $C^*(1+r_o)$ to C^*, but that firms can realize this cost saving only by adopting the new technique. First, if no new information is available, there is no effect - firms will simply continue producing as before. With information purchasing opportunities firm size will orginally organically expand and will then return through time to its original level as r_o falls as a result of prior investments in information. If, on the other hand, technical change occurs continuously then r_o is an index of the pace of change. In the Nelson-Phelps ( ) terminology, it is the "technological gap". The larger the technological gap the greater is the return to investing in information. If information costs are independent of r_o, the pace of change, then the greater is the change the greater is the scale bias (the discrepancy between x_1 and x_o) associated with investments in information.

In Table 1, post-World War II trends in farm size and agricultural productivity are summarized. The important evidence is that in periods during which productivity was advancing most rapidly, farm size was growing most rapidly. This is at least suggestive that productivity growth offers incentives for size expansion.

Table 1

Trends in Farm Size and AGricultural Productivity, 1945-64

Year	Numbers of Farms (millions)	Gross-Receipts per Farm (1957-59$)	Output per unit Inport (1957-59=100)	Growth (percemt per year)	
				Receipts per Farm	Output per Unit input
1945	5.9	5,240	82	1.7	0.7
1950	5.4	5,680	85	4.9	1.8
1954	4.8	6,880	91	8.2	2.2
1959	3.7	10,220	101	6.2	1.6
1964	3.2	13,780	109		

Source: Various issues of Agricultural Statistics (15). The price deflator is the U.S.D.A. index of prices received by farmers.

My purpose here is to examine determinants of returns to scale in U.S. agriculture with special reference to scale economies in the use of information.

Scale Economies: Although the concept of scale economies is used in various ways, one consistent theme is that in a competitive industry the existence of returns to scale is taken as a signal of disequilibrium such that adjustments result in larger firms with lower unit costs. When the term, scale economies, is used, I mean simply that larger firms have a comparative advantage over smaller ones. In the case of a single product, scale economies exist if average cost is falling, and if firms produce multiple products, scale economies refer to the cost/revenue ratio declining as total cost or revenue rises. I assume that firms are competitive so the existence of scale economies implies that something in the total environment has changed to the relative advantage of larger firms and that adjustments have not exhausted this advantage.

To allow for the production of more than one product, aggregation is necessary. The assumption here is that firms are efficient in the sense that revenue is maximized given cost so that for given market conditions (prices) there is a functional relationship between revenue and cost. If a marginal expenditure on any given input is allocated to produce a particular product the efficiency requirement is that an alternative allocation not exist that would produce more marginal revenue. If the quantity produced of one product does not depend upon the allocation on inputs to another then competitive firms will specialize in producing a single product whenever marginal costs are falling. Diversification occurs when production functions are interdependent or for independent production functions whenever marginal costs (deflated by marginal revenue) are rising.[10] The revenue constant cost minimizing condition is that the ratio of marginal cost to marginal revenue be common among all products and that the ratio of marginal factor cost to marginal factor revenue be common among all inputs in the production of all commodities. If these conditions are satisfied and production is diversi-

---

[10]There is an exception for the independence case. If marginal cost (deflated by marginal revenue) is falling for one product at a slower rate than the aggregated deflated marginal cost functions for all other products is rising, then the firms can reach a diversified equilibrium for given total cost. But in this case as aggregate output expands the expansion is accomplished by increasing production of the product with falling marginal cost and reducing the production of all other products.

fied the firm's aggregated marginal cost function must be rising.[11] Thus for analyzing scale economies in which average costs are falling we are restricted to the range of production in which marginal costs are rising. This, of course, corresponds to levels of production in which short-term equilibria can be established but not maintained in the longer term and points to the linkage between the presence of scale economies are pressures to adjust firm size.

Throughout, scale elasticity is defined as the elasticity of revenue with respect to cost and scale economies exist when this elasticity is greater than unity. For a single product the scale elasticity is the ratio of average to marginal cost and if the production function is homogeneous, this measure is the degree of homogeniety.

Table 2 reports measures of farm size for the six commercial classes of farm specified by the U.S. Department of agriculture[12] when farm size is measured alternatively by total output and by total inputs. The third row gives the normalized cost/revenue ratio and the fourth reports the elasticity of revenue with respect to cost between size classes.

---

[11] The aggregated marginal cost function is defined as the horizonal summation of the individual product marginal cost functions when each is deflated by its marginal revenue and the quantities being summed are either total costs or total revenues.

[12] For some purposes, large scale farms, having sales above $100,000 are distinguished. For economic Classes I-VI, respectively, stratification is based upon the value of farm products sold as: $40,000 or more; $20,000-$39,999, $10,000-$19,999; $5,000-$9,999; and, $50-$2,499 with operator under 65 and working off farm 100 or more days. Farms dubbed "non commercial" include, Part-time: farms with sales of $50-$2,499 whose operators work off the farm more than 100 days. Part-retirement: farms with sales of $50-$2,499 with operator 65 years old or over. And, abnormal: institutional (hospitals, penitentiaries, schools, etc.) and Indian reservations.

As reported, the normalized cost/revenue ratio falls from 2.63 for the smallest class of commerical farms to 1.61 for the next smallest class, giving an interval elasticity (computed from row 1 and row 2 as the percent change in revenue relative to the percent change in cost) of 2.21. As farm size increases, the scale elasticity falls consistently to a low of 1.18 for a comparison of the two largest size classes.

Contributing Factors. I assume that scale economies exist because changes have occurred and perhaps are occurring that give comparative advantage to larger firms and that adjustments in firm size have been insufficient to exhaust these advantages. Thus an explanation of why scale economies exist is in fact an answer to the question: What changes?

Here two sources of pro-scale bias are explored. The first feeds upon the scale-free aspects of information and the second is a discussion of how differences in input and product mix alongside changes in prices and technology alter comparative advantage.

Information: In the context of U.S. agricultrue where relative prices are changing and there is a continual stream of new and reportedly improved input varieties and techniques, it would be surprising if entrepreneurs were consistently able to maximize ex post profits. Instead, mistakes occur and are costly and the anticipation of these mistakes leads to a demand for information. If the value of knowing that A is superior to B is in direct proportion to the use of A or the non-use of B, scale economies in using information exist. ...

Measured scale elasticities are related to interaction between rates of change in educational levels of farm operators (as a function of farm size) and public expenditures on agricultural research for evidence of this scale-information relationship. Farmer education is seen as a shift

Table 2

U.S. Aggregate Estimates of Returns to Scale by Economic Class of Farm, 1959

	Economic Class					
	I (Large)	II	III	IV	V	VI (Small)
Value added per farm by primary inputs relative to U.S. average	5.95	2.13	1.25	0.72	0.38	0.16
Cost per farm of primary inputs relative to U.S. average	3.98	1.70	1.18	0.86	0.61	0.42
Cost/revenue relative to U.S. average	0.67	0.80	0.94	1.19	1.61	2.63
Implied returns to scale (% change in revenue ÷ % change in cost)	1.18	1.44	1.71	1.82	2.21	

Source: Census of Agriculture ( )

Notes to Table 2.

Value added is sales of farm products less purchases of livestock, feed, and seeds and less the imputed cost of livestock inventories @ 10% of estimated value. Cost of primary inputs includes imputed cost of labor, fertilizers, land and machinery. See Table 4 for a discription of these imputations.

operator on the short-term information cost function, and from a long-run perspective the opportunity cost of education is an argument in total informational expense. That operator education is closely related to farm size is evident in Table 3 where the distribution of school completion levels of farm operators by economic class of farm is summarized. Notice that

Table 3

Schooling of Farm Operators by
Size of Farm, 1964

Percent Who	Economic Class of Farm						
	Large Scale (Sales of $100,000+)	Class I less large scale Farms	II	III	IV	V	VI
Did not attend Highschool	20.4	23.4	29.5	39.2	49.6	54.7	66.9
Completed at least one year of college	31.9	22.1	14.6	10.1	8.5	8.5	5.5

Source: Census of Agriculture 1964, volume II, ch. 6.

three of every ten farms with sales exceeding $100,000 in 1964 are managed by persons who have completed at least one year of college and only one in twenty operators on the smallest class of commercial farms had attended college.

Composition: It is trite but true that if the price of a purchased input falls, the firm that otherwise would have purchased the most of that input will have the greatest reduction in cost. It is also true that the firm spending the greatest proportion of total cost on that input will as a result of the reduction in its price realize the greatest proportionate decline in cost. Thus when prices change, comparative advantage shifts among firms in inverse proportion to expenditure or revenue share of the inputs or products whose prices are changing.

Define the scale elasticity as

$$S = \frac{C/y}{\lambda}$$

where C/y refers to average cost and λ is marginal cost: Let p_i refer to the price of the i-th input in the production process $y - f(\underline{x})$; $X = \{x_1, \ldots x_n\}$. Using the notation $Ea = d(\ln a)$ so that $\frac{Ea}{Eb}$ is the elasticity of a with respect to b, we have

$$\frac{ES}{E\ p_i} = \frac{E\ C/y}{E\ p_i} - \frac{E\lambda}{E\ p_i} .$$

It is well known that holding output constant, $\frac{E\ C/y}{E\ p_i} = k_i$, the expenditure share of x_i. Further, from Samuelson (12),

$$\left.\frac{d\lambda}{dp_i}\right|_{y,\ \text{all other prices}} = \left.\frac{dx_i}{dy}\right|_{\text{all prices}}$$

So that

$$\frac{E\lambda}{Ep_i} = k_i \frac{Ek_i}{EC}$$

where C refers to total cost. It is also obvious that

$$\frac{Ex_i}{EC} = \frac{EK_i}{EC} + 1.$$

On combining terms,

$$\frac{ES}{Ep_i} = -k_i \frac{EK_i}{EC} .$$

This result is intuitively obvious: If the expenditure share of x_i increases with firm size the scale elasticity is increased by a reduction in p_i.

A similar result can be deduced for the case of non-neutral technical change. The simplest specification of technical change is in terms of input qualities. Let the production process be specified in terms of effective input quantities which are nominal quantities multiplied by quality, i.e., $y = f(\underline{x}^*)$; $\underline{x}^* = \{x_i^* \ldots x_n^*\}$, where $x_i = q_i x_i$. Now with respect to costs of production, a one percent increase in q_i holding constant all prices is the same as a one percent fall in p_i holding quality constant. So it must be true that

$$\frac{ES}{Eq_i} = -\frac{ES}{Ep_i}$$

In the classic case of perfect competition all firms in an industry are scalar replicas of each other. In this case, $\frac{Ek_i}{EC} = 0$ because cost expansion occurs only with the addition of new firms and comparative advantage is unaffected by external changes in either prices or technique. In U.S. agriculture, the evidence is contrary to the classic case. Large firms differ from smaller firms in many respects other than size: they simply produce different products using different inputs. Because differences in composition of input and product bundles are related to farm size, scale economies will vary as prices vary. In Table 4, the composition of revenue and production cost is summarized for each of the Census economic classes.

The most striking feature is that larger farms are much more dependent upon farm produced inputs such as feed and livestock than are the smaller farms. Notice that Class I farms devote 50 percent of their cost dollar to farm produced inputs as compared to 9 percent for Class VI. This points out one fallacy of measuring size by gross sales. Consider a farm that raises both feed and livestock. By selling its feed and feeder calves to other farms and then repurchasing them, its "size" would be approximately doubled. As measured by gross cost, Class I farms are seventeen times as large as Class VI farms but as measured by cost not of farm produced inputs, the relative size of Class I farms falls to less than ten times that of Class VI farms.

Within the complex of farm produced inputs, larger farms are more dependent on beef cattle and with the exception of the Class I farms, there is an inverse relation between size and expenditure shares for labor. Larger farms appear more land and fertilizer intensive, while small farms are labor and machinery intensive.

Table 4 also provides a summary of revenue shares. The largest absolute discrepancies are for field crops and beef cattle. Class I farms receive 30 percent of their revenue from field crops and 32 percent from beef cattle whereas Class VI, the smallest commercial class, receive 59 and 15 percent respectively from field crops and beef cattle. A change in relative prices such that the price of beef cattle increases by 10 percent relative to field crops will increase the gross revenue of Class I relative to Class VI farms by about 4.5 percent.

The Empirical Specification

The analysis is cross-sectional and is based largely upon Census of Agriculture data for 1959. First, average scale elasticity is estimated for each of the 48 coterminous states. Then summary measures are constructed to capture some of the scale biases that may have emerged

Table 4

Revenue and Cost Shares for Major Commodity and Input Groups by Economic Class of Farm, 1959

Percent of-	Economic Class I (large)	II	III	IV	V	VI (Small)
1. Revenue from						
a) crop sales	47.5	44.8	45.2	47.5	51.6	63.8
b) Livestock and Livestock products	52.5	55.2	54.8	52.5	48.4	36.2
2. Cost attributed to						
a) Labor including operator and family	21.1	21.4	24.9	31.4	38.8	57.5
b) Land, machinery fertilizer and seed	30.0	41.6	46.6	57.5	45.1	33.7
c) Livestock inventories, purchases, and feed purchases	48.9	37.0	28.5	21.1	46.1	8.8

Notes to Table 4

Except when notes data are from the 1959 Census of Agriculture (15).

Definition of Expenditures:

Labor - The cost of labor is taken as the sum of expenditures on hired labor and an estimate of the value of operator family labor. Annual equivalent farmer labor is one less one-half the proportion of farmers in each economic class (except part-time) working off the farm 100 or more days. Days worked per year are assumed to be 250 for commercial farms. For part-time farms, days are 75 and are 150 for part-retirement farms. This estimate of annual farmer days is then multiplied by one plus the ratio of unpaid family to farmer labor. This estimate is reported by Kislev (6) for each state but is not given separately for the economic classes. Farm family days are then multiplied by the state specific average daily wage of hired farm labor (17).

Land - The cost of land is imputed "user" cost. The opportunity cost of capital is assumed to be 10 percent. The cost of land is then .1 less the average annual rate of land value appreciation, in the state, realized between 1954 and 1959 times the reported value of farm lands and buildings. In cases in which average annual appreciation exceeds 8 percent, the cost of land is 2 percent of value. (con'd)

(Notes on Table 4 cont'd)

Machinery - 22 percent of the estimated value of machine inventories plus expenditures for gas and oil plus expenditures for custom hire of machine services. The "22" corresponds to a 10 percent opportunity cost of capital plus 12 percent annual depreciation. The value of machine inventories is computed using physical inventories and prices (regional) reported by Kislev (6).

Fertilizer - Census reported tonnages for fertilizer and lime adjusted by state specific prices (6).

Livestock - 10 percent of the value of livestock inventories (60 percent for poultry) plus purchases of livestock and poultry. Except for poultry, no inventory depreciation is assumed. It is assumed rather that sales are reported net of herd maintenance expense. The proration of livestock and poultry expenses is as follows: For poultry farms, the national average ratio of purchases for livestock and poultry to poultry sales computed (0.59) and for dairy farms the ratio of livestock and poultry purchases to sales is 0.33. Within each economic class of each state the shares of poultry to total livestock sales, S_1, and the dairy share, S_2, are computed. Purchases of poultry and dairy cattle are respectively assumed to be $.59S_1$, and $.33S_2$, is assumed to be purchases of beef cattle.

Feed, seeds and bulbs, plants and trees - Expenditures are reported in the Census.

as a result of changing informational demands. A second class of explanatory variables is constructed to allow for scale economies that are associated with size differences in input and product composition. Finally, using the informational and composition constructs, estimated state scale elasticities are "explained" by a series of ordinary least-squares regressions.

In addition to the statistical question of how significant are the correlations between each of the information and composition variables and the measures of returns to scale there is the related question of attributing the observed scale bias to these forces. For this purpose a final accounting section is added. It concludes that about 20 percent of measured scale biases, at the U.S. average, is due to informational economies, another 55 percent is related to composition, and the 25 percent residual is unexplained.

Notice that to explain scale economies it is not necessary to explain them away. The evidence is overwhelming that forces exist to increase average farm size. It also appears true that as average size increases, aggregate output rises relative to input. One purpose of this exercise is to show that the black box of productivity advances is based largely upon very real and relatively simple changes in underlying conditions.

<u>Scale Elasticities:</u> The underlying <u>Census</u> data from which state specific scale elasticities are estimated stratify farms into nine economic classes. For each class, revenue is defined as sales per farm of all products. Data are not available to permit the inclusion of government payments, the value of home consumption, changes in inventories, or the rental value of home dwellings. The exclusion of these terms may

result in an upward bias in measured returns to scale since they probably account for a larger proportion of revenue for smaller farms.[13] Also, the stratification of farms by revenue permits the regression fallacy and adds to the upward bias in measures of returns to scale.[14] Because of these factors, my estimates are considerably higher than those reported from the aggregate production functions. Nonetheless, the purpose is to analyze state-to-state variations in returns to scale and if biases introduced by stratification and incomplete revenue specification are not systematically related to the explanatory variables then regression estimates (other than the "intercept") will be unbiased.[15] The state specific estimate of scale economies is the slope coefficient from a regression of the logarithm of revenue on costs, where costs are estimated by adding to out-of-pocket expenses, imputed cost for machine

---

[13] Tweenten and Schreiner (13) have shown that government payments are slightly more evenly distributed than are farm receipts. In 1959 farm receipts were $424.9 million, increases in inventories were $4.6 million, and government payments were an additional $3.6 million (16). Further, home consumption and the rental value of farm dwellings constitute a larger fraction of total income for smaller farms.

[14] Because of production uncertainty, revenue is a random variable. Consider two farms operating with the same input scale and assume that one experiences a "good" year with positive deviations in revenue and the other has a "bad" year. Clearly the average cost for the unlucky farm exceeds that of the lucky farm. If size is measured by revenue then we will find that as size increases, average cost falls, so that there are scale economies. In fact, the two farms are the same cost size and we should not be able to say anything about the presence or absence of scale economies from observing only them.

[15] For the regression model $y = a + X\beta + \mu$, suppose the $E(u) = b$, the bias in the scale estimate, and suppose that, $E(X'(u-b)) = 0$; so that the stochastic component is independent of the "explanatory" variables. Then $E(\hat{\beta}) = \beta$ where $\hat{\beta}$ is the ordinary least squares estimator and $E(\hat{a}) = a+b$. The bias is only a "level" effect and is transferred to the intercept.

and livestock inventories, farm family labor and user cost for land.[16] the observations are weighted by number of farms and are for the nine economic classes reported in the Census.

Although the form of the equation estimated holds the scale elasticity constant, the evidence is that, in fact, it declines with increasing farm size. Recall in Table 2 the U.S. summary showing that the cost/revenue ratio falls from a factor of 2.63 to 1.61 between Classes VI and V as cost per farm increases by 37 percent implying an interval elasticity of 2.21 between these two smallest classes whereas the comparable elasticity between Classes II and I is 1.18. Cost per farm is 9.5 times as great on the Class I as compared to Class VI farms. This is as expected; earlier, I observed that if farms are diversified, marginal costs must be rising. But if scale economies exist, if average costs are falling and marginal costs are rising, the scale elasticity must fall as size increases. The fact that the scale elasticity declines as farm size rises presents a problem concerning how states should be compared. Ideally, a measure of scale economies is an indication of the comparative advantage of larger size and the measure used here is an average of this advantage over the various classes. Since observations within each state are weighted by the number of farms, the points at which the scale elasticity-size function is evaluated will vary from state to state. As an example, suppose that all states have exactly the same relationship between classes as that indicated in Table 2, and differ only in their size distribution of farms. In this case, states with farms whose average size is above

---

[16] A detailed description of the operational definition of costs is provided in the footnotes to Table 5.

the national average would have low scale economies relative to those states with smaller farms. But this is as it should be, for in states with farms of larger than average size, the added advantage (in terms of average cost) to a further increase in size would be less than in states with smaller farms. The scale elasticity estimate is a regression estimate within each state of the elasticity of revenue with respect to cost. Observations are averages per farm for the nine economic classes and are weighted in the regressions by the number of farms in each class.

Information: Although data are not available to identify a farm's informational investments, the picture for U.S. agriculture is not hopeless. The 1964 Census reports the distribution of years of school completed for farm operators in each economic class of each state. The evidence is that in all states except Rhode Island, "average" schooling increases with farm size. If education facilitates learning, then in an uncertain environment more educated farm operators whould be more efficient. The function of agricultural research is one of producing new and better inputs and techniques, and since the research product is new, efficient use requires learning. If more research implies a more rapid flow of new things, then we can assume that the greater the level of research activity the greater is the spread between traditional and "best" methods and the greater the potential role of learning. With learning efficiency being related to education and more educated operators residing on larger farms, the comparative advantage of larger farms would be greater the greater the level of research activity. But extension activity presumably serves the purpose of disseminating the research product.

Unfortunately, there is no data concerning the distribution of the extension effort between farm size classes. Unless this effort is severely skewed in favor of larger farms, we can think of extension activity as simplifying the learning required in conjunction with the research product, which is therefore a reducation in the level of technical uncertainty. If research enhances the comparative advantage of more educated farmers, then this advantage may be eroded by extension.

A set of variables is constructed to see if scale economies have been affected by the research-education-extension complex. First an index of scale-education bias is constructed for each state. Within each economic class average farm operator schooling is computed in which reported years of school completed are weighted by 1959 income levels. The elasticity of average schooling with respect to total cost is computed from a double log regression across economic classes, like the regressions used to estimate scale elasticities. This regression measures the rate of increase in operator schooling with farm size and is dubbed the education bias. Interaction variables between education bias, research and extension are constructed using public research expenditures per farm in each state and an estimate of days per farm spend on farms by personnel of the Extension Service.[17]

Composition: Since relative prices do change it is possible that these changes have altered the structure of comparative advantage among the size classes. To allow for this, a variable is constructed to summarize the effects of recent price changes. Holding the composition of cost and revenue constant, i.e., holding revenue and cost shares constant,

---

[17]Operational definitions are provided in the notes to Table 5.

both total cost and revenue are computed using the average of the 1954 and the 1949 prices. Then, scale elasticities are computed in the same manner as for the cost and revenue values given in 1959 for each state. The difference between the scale parameter computed using 1959 prices and the one calculated when prices are lagged (about 7.5 years) is then dubbed the pecuniary bias.[18]

The calculation of pecuniary bias can differ sharply between the gross and value-added formulations. Consider beef cattle. For Class I farms, they account for 33 percent of sales and 25 percent of cost, but for Class VI farms the corresponding numbers are 15 and 3 percent. A 100 percent increase in the price of beef cattle would increase revenue relative to cost by 12(15-3) percent for Class VI farms and only 8 percent for Class I farms thus increasing the "comparative" advantage of the smaller farms. For the value added formulation, beef cattle appear only in "net" revenue and represent 6.6 percent for class I and 2.3 percent for Class VI. Thus an increase in beef cattle prices will increase

---

[18] In general, the technique is simply to multiply percentage price changes between 1949-54 and 1959 to each of the cost and revenue shares to compute cost and revenue at 1949-54 prices, holding physical quantities constant. For all inputs except labor, percentage price changes are taken only at the national average. For labor, the state specific daily wage rate of hired farm labor is used (17). The user cost of a physical unit of land is 10 percent less the average annual appreciation rate over the previous five years in the state times the national average acre value of farm lands and buildings. For products, antional average price changes are used, but state specific indexes are calculated for field cross-based upon revenue shares for the one to five most important crops in each state. The revenue shares are computed using physical produceltion and state specific average prices as reported in the Census. For the eight states most dependent on fruit and vegetable revenue, state specific indexes are computed using "state" prices and state weights. For other states, the average of these eight indexes is used. Price data are taken from various issues of Agricultural Statistics (15) and the Census of Agriculture (14). Machinery prices are "quality adjusted" using the Fettig (2) series for farm tractors.

the comparative advantage of Class I farms.[19]

At an accounting level, there is evidence that price changes have fairly systematically favored larger farms. At the national average, the value added measured scale elasticity is 0.21 higher when prices are at 1949-54 levels then at 1959 levels. Further, for value added, price changes have increased the measured scale elasticity in 47 of 48 states (Alabama is the exception).

But price changes are not the only way that composition differences lead to changin comparative advantage, non-neutral technical changes may alter scale economies. The logic is the same as for pecuniary effects. As technical change augments one input relative to others, the comparative advantage of firms most dependent upon that input will increase.

To allow for factors other than direct pecuniary bias, two indexes of input mix dissimilarity are constructed. They are elasticities of expenditures for specific items with respect to total cost.[20] One refers to expenditures for intermediate inputs and the other to the combined cost of the traditional land and labor inputs.

---

[19] This only points out that the term comparative advantage should be in quotes throughout the analysis. Obviously the important question should be whether a change in a given price will change profits by more (absolutely) in one class than another. In that respect, the value added form is superior. For example, a 10 percent increase in beef cattle prices would increase average profits of Class I farms by $260 and only by $2 for the average Class VI farm.

[20] These elasticities are computed in the same manner as the scale parameters. Regressions weighted by numbers of farms are calculated within each state across the nine economic classes. The regression is of the form: log of specific expenditure regressed on the log of total cost (per farm).

Regression Results

Regression results are presented in this section for the value-added formulation with the corresponding estimates for the gross sales function being provided in Appendix B.

Table 5 summarizes the value added results. Independent variables are grouped according to whether they refer to compositional effects or to information. Among the compositional variables, the first, direct pecuniary effects, is an interaction constrct between price changes and compositional differences. In addition to the adjustment for price changes, compositional effects are important in the case of non-neutral technical change. Variables I(b) and I(c) measure the rate at which input mix changes as farm size increases and presumably capture pervasive effects due to changing input comparative advantage, i.e., changes in marginal productivity relative to price. Variable I(b) is the expenditure elasticity for the traditional land and labor inputs. Much of the product of the private research effort and the contribution of technical change that occurs outside of agriculture may be embodied in non-traditional inputs such as fertilizers and machinery. Yet embodied technique need not enhance the comparative advantage of the factors being augmented because the existence of complementary inputs introduces the possibility of transferal. This variable is included to see if there is evidence of changing comparative advantage between traditional and non-traditional inputs. Variable I(c), the expenditure elasticity for intermediate inputs, serves a dual role. Like I(b) it can capture the effect of non-neutral technical change, for much of the public research effort is devoted toward improving livestock, feed, and seeds, but it can also be considered as an index of specialization in livestock production as size increases.

Table 5

Regression Estimates of Factors Affecting Measured Returns to Scale in U.S. Agriculture, 1959

Value-added Formulation; Coefficients (Standard Errors)

Variable	Mean	Regression Number 1	2	3	4
I. Composition Effects					
a) Direct pecuniary bias	.214	1.296 (1.123)	1.435 (.096)	1.481 (.092)	1.444 (.084)
b) Expenditure elasticity for land and labor	.960	1.583 (.537)	1.916 (.322)	1.944 (.317)	1.514 (.254)
c) Expenditure elasticity for intermediate inputs	1.682	.515 (.087)	.569 (.045)	.561 (.045)	.543 (.041)
d) b x research expenditures		.034 (.037)			
e) c x research expenditures		.000 (.003)			
II. Information Complex					
a) Education bias	.112	2.239 (1.012)	1.625 (.919)	.222 (.382)	
b) Research expenditure	13.270	-.042 (.038)	-.006 (.003)	-.001 (.001)	
c) Days on farms by extension personnel	.307	1.089 (.448)	.935 (.429)	-.092 (.088)	
d) a x b		.120 (.063)	.104 (.047)		
e) a x c		-7.367 (3.073)	-6.558 (3,045)		
f) a x b x c		-.172 (.075)	-.145 (.067)		
III. Regional Dummies		+	-	-	-
Intercept		-.806	-1.199	-.967	-.540
Dependent variable	2.136 (.235)				
R^2		.942	.933	.922	.912
RSSQ		.1539	.1779	.2081	.2335
d.f.		34	38	41	44

Notes to Table 5

All regressions are for the 48 coterminous states and observations are weighted by the number of farms in each state.

In the value-added formulation, both revenue and cost exclude estimated expenditures for livestock, feed and seeds, inputs considered as farm produced. State specific estimates fo returns to scale, direct pecuniary effects, expenditure elasticities for land and labor and farm produced or intermediate inputs and the farm operator education bias are each estimated as slope coefficients from within-state double logarithmic regressions in which the independent variable is estimated total cost per farm and are described in the text.

The research variable is provided by Evenson and is similar to the variable reported in (1). The main difference is that some expenditure items such as "economic research" which are not seen as directly contributing to agricultural production have been omitted. Evenson estimates total research expenditures for each of nine years ending in 1959 and then averages them by using inverted "V" weights over the period. Finally, average annual research is divided by the number of farms in the state.

The extension variable is an estimate of average annual days per farm spent on farms by personnel of the Extension Service. The data are unpublished but are provided by the Federal Extension Service of the U.S. Department of AGriculture. Each year, each country staff reports a distributio- of days spent on a list of about twenty activities. Of these, six are considered as being directly related to agricultural production and may well have been days spent on farms. The annual average is computed using the weights .16, .34, .34, .16 for the years 1959 to 1956, respectively. The two regional dummy variables refer to the south only. One is for the East South Central Division and the other is for both the South Atlantic and the West South Central. In the gross value formulation, the south, especially the East South Central Division, had lower estimated returns to scale than the rest of the country.

Because data for public research expenditures are available, variables I(d) and I(e) are designed to capture state specific compositional effects induced by research. They are simply the product, research expenditures times variables I(b) and I(c), respectively. These last two variables should be important if the impact of composition effects is conditioned by the local research effort.

Variables referring to scale economies originating with the use of information are intrinsically interaction variables. For example, the comparative advantage of more education is conditioned by learning possibilities. Here it is presumed that the more rapid the research flow and the less active the extension service, the greater is the comparative advantage of education. Thus the interaction terms II(d), II(e), and II(f) are introduced. The first is the information hypothesis pure and simple. If the comparative advantage of education is enhanced by research activity then it seems reasonable to interpret this advantage as having its roots in learning capacities associated with education. The remaining interaction variables are relationships between education and extension. The earlier discussion assumes these effects to be negative, which is to assume that more educated persons do not themselves have a comparative advantage at interpreting the information supplied by extension personnel.[21] And, according to the assumption that education

---

[21]Actually, evidence of interaction between extension and education cannot contradict the information hypothesis. Negative interaction says that they are substitutes and, therefore, supports the hypothesis, but positive interaction would also, while "no effect" would not contradict it, but would at least be consistent with a contradiction. Extension clearly refers to information dissemination and if education is seen as augmenting the spread of information it is not clear how they interact. The key to breaking into this phenomenon would be an understanding of

(Cont'd on the next page)

and extension are competitive, if the magnitude of the education effect is conditioned by research, then the effect of research on education is conditioned by extension--variable II(f). Variable II(e), extension-education-interaction, is designed to capture any effects between education and extension not in proportion to research expenditures. Variables II(a) and II(c) are the "own" effects of education, research, and extension. While it is clear that each can have an effect upon the comparative advantage of size via information, other effects can exist as well. Education is the most obvious example: In an opportunity sense, more educated farm operators cost more. Yet, the cost estimates used are independent of operator education. Since operators of larger farms are more educated, the absolute effect of this omission increases with size and may have introduced an upward bias to the return to scale measure.[22]

In Table 5 the most striking evidence is that direct pecuniary effects literally swamp the regressions, both in the sense of partial correlations and, as is later demonstrated, in accounting for the level of measured returns to scale. Clearly, price movements during the decade of the 50's enhanced the comparative advantage of larger farms. In the

---

(Footnote 21 continued from previous page) (1) the level of complexity associated with extension information as compared to the level and rate of change in the level of complexity of information persons at a given level of education would be capable of "digesting," and (2) the covariance between farm operator education and the distribution of extension effort. If "they" do not talk to you then what they are saying does not really matter.

[22]But the bias introduced by excluding operator education in imputing cost may be relatively more important on smaller farms since operator labor accounts for a much larger share of all costs. Therefore, this omission may have resulted in a negative bias in measured scale economies.

context of this analysis, it is not possible to say whether larger farms were lucky in the sense that they happened to be producing those things whose relative price increased using inputs whose relative price fell or whether they correctly anticipated price movements and adjusted product and input mix more efficiently than smaller farms.

The evidence is that in addition to price movements, comparative advantage increased for firms most dependent upon intermediate inputs and firms most dependent upon the traditional land and labor inputs. That is, there is evidence of non-neutral technical change that enhanced the productivity of intermediate inputs and of land and labor relative to the productivity of the non-traditional fertilizer and machinery inputs. This evidence is contained in the "significantly" positive coefficients for variables I(b) and I(c). Here, the evidence is not uniquely in favor of larger farms because smaller farms tend to be more dependent upon land and labor and larger farms are more dependent upon intermediate inputs. In fact, the compositional effects are predominantly favorable to larger farms since, at the national average, the elasticity of expenditures on land and labor with respect to expenditures on all primary inputs is 0.96, indicating that the composition of expenditures on primary inputs does not change rapidly as size varies. In 18 of 48 states, larger farms are actually more land-labor intensive than smaller farms.

Although the state specific research-input composition interaction variables do have positive coefficient estimates, they are not significantly different from zero at ordinary confidence levels. While there is evidence of non-neutral technical change, the evidence that this change is locally produced is unconvincing.

Another important feature of the evidence contained in Table 7 is that regional dummy variables are not significant. That is, given the compositional and informational effects, residual errors are not systematically related to region. This is reasuring since regional dummies, in this context, are little more than summary measures of ignorance.

The evidence concerning the importnace of variables in the information complex is much less convincing than for the compositional variables. It is true that five of the six coefficients are at least twice as large as their standard errors while the remaining "t-ratio" is 1.8. And it is also true that the interaction coefficients have the correct sign (there is no hypothesis concerning the signs of the own coefficients). But it is obvious from the definition of the variables that they are highly colinear. This cannot be avoided since interaction is the crux of the information hypothesis.

In any case the test of the information hypothesis is on the <u>joint</u> significance of the variables rather than upon their individual significance. That these variables are colinear is best seen by comparing equations 2 and 3. When the interaction variables are deleted the "t-ratios" on each of the three own coefficients are less than one: if the variables in the information complex have a story to tell, they tell it through interaction. Nonetheless, in comparing equation 2, in which all informational variables are included, to equation 4, in which they are deleted, the "test" of their joint significance gives $F(6,38) = 1.98$ with an associated critical value of 2.35 at the 0.05 level.

Interestingly enough, if regional dummy variables are included in equations 2 and 4, the corresponding test statistic is $F(6,36) - 2.50$, with 0.5 critical level, 2.36. Also, in the gross revenue formulation, the corresponding test gives F(6,37) = 3.35 with 2.35 as the associated (.05) critical value (Appendix A). Thus, the evidence pertaining to the information hypothesis is mixed. I favor the assumption that information is significant and offer equation 2, TAble 5, as my best estimate of factors affecting returns to scale in U.S. agriculture, 1959.

Accounting for Scale Economies

At the beginning of this paper, I noted that Griliches estimates that slightly more than one-half or the productivity increments achieved in U. S. agriculture between 1940-1960 are the results of increasing farm size alongside economies of scale. If these economies are purely internal engineering phenomena, then little more could be said, but the evidence presented here is that they may themselves be the product of price and technology changes which are not neutral between firms relying on different product and input mix. Furthermore, the dynamics of technical change appear to have created scale economies in the use of information. Nonetheless, a question remains concerning the sheer magnitude of these effects.

Define scale bias as the measured scale economy less one[23] at the sample mean; for the value-added formulation, scale bias equals 1.14. Table 6 reports the distribution of measured scale bias among its sources as measured by equation 2, Table 5. Approximately 75 per cent of the scale bias is accounted for with 55 percent being the result of composition differences and 20 percent attributed to variables in the information

---

[23]Scale bias is simply (minus) the elasticity of average cost with respect to total cost.

Table 6

Factors Affecting Measured Scale Bias in U.S. Agriculture, 1959

Composition		
Price changes	27.0%	
Land and labor share differentials	-6.7	
Intermediate input share differentials	34.2	
		54.5%
Information		
Partial effect of:		
Education	3.9%	
Research	.7	
Extension	.7	
Interaction	15.0	
		20.3%
Explained Total		74.8%

Notes to Table 6

Computed from equation 2, Table 5. The partial effect of education, research and extension are each evaluated at the sample mean, then multiplied by the average value of the indicated variable. For example, the partial effect of research = (-.006 + .1039 Ed. - .1448 (Ed. x Ext.)) Research. The interaction effect is computed by computing the change at the point of means in measured scale economies if all information variables are assigned the value, zero, and then subtracting the partial effects.

The effects of price changes is computed by determining how much the mean scale bias would fall if the pecuniary bias were 0 instead of .214, given the regression coefficient (1.435). Similarly, the effects of dissimilarity of input mix vis a vis land-labor and intermediate inputs is computed by assigning elasticities of unity for the average at the point of means, .96 and 1.68, respectively. These elasticities would be unity if firms were scalar replicas of each other.

complex. Price changes appear to have been very significant in augmenting the comparative advantage of large farms as have been residual effects (probably technical change) that have increased the comparative advantage of large farms. For the land-labor and intermediate input variables, a larger expenditure share is positively related to the comparative advantage of larger farms, so that the effect for land and labor at the sample mean is negative: on average, the expenditure elasticity for land and labor is less than one (recall that it exceeds one in 18 of 48 states).

At the point of means, the partial effects of education, research, and extension are very small. Three-fourths of the scale bias explained by these factors is associated with interaction between them, which supports the information hypothesis.

Summary

The evidence presented here is that empirically measured returns to scale in U.S. agriculture have been produced externally to the individual firm by price and productivity changes that favor input and product compositions of larger farms and by a rapid pace of technical change with associated scale economies in the use of information.

The Census data show that larger farms derive proportionately more of their revenue from livestock products whereas smaller farms concentrate more upon cash grains. The increasing price of livestock products relative to cash grains during the 1950's created comparative advantage for livestock farms, which being larger also created scale economies. On the input side, it is also true that larger farms depend more heavily upon farm produced or intermediate inputs than do smaller farms. Since this largely reflects the increasing concentration in livestock production, the empirical analysis is couched in value added terms with costs of intermediate inputs being subtracted from revenue rather than added to cost.

Residual or primary inputs are land, labor, machinery and fertilizer and on balance larger farms are more dependent upon non-traditional machinery and fertilizer inputs while smaller farms focus more upon traditional land and labor inputs. Since the price of land and labor has increased relative to fertilizer and machinery, these changes have been differentially favorable to larger farms. In fact, at the sample mean, my estimate is that roughly 27 percent of the measured scale bias is associated with product and input price changes that occurred during the decade prior to 1959.

But price changes do not tell the whole story of comparative advantage originating in differences in product and input mix. The evidence is that given price changes, firms devoting a larger share of expenditures to intermediate products and to traditional land and labor inputs have gained relative to others.

Here the picture is blurred because larger farms depend much more upon intermediate inputs and smaller farms are more dependent upon land and labor. But at the sample mean, the net effect has favored larger farms. ABout 34 percent of the measured scale bias appears to be associated with the rising relative importance of intermediate inputs as costs for primary inputs increase while the measured bias is reduced by only 7 percent as a result of the increasing advantage of the traditional inputs.

Another characteristic of larger farms is that their operators are more schooled and in a technically dynamic state in which learning becomes an important input into the profit calculus, the added operator schooling appears to have enhanced the comparative advantage of larger farms. While it is true that these effects are partly dampened by negative extension-education interaction, information effects still account for around 20 percent of the measured scale bias. This information bias is created by positive interaction between operator education-scale differentials and research expenditures and is reduced by interaction with extension. The negative education-extension interaction suggests that vis a vis allocative efficiency, education and extension are substitutes.

III. Additional Empirical Evidence

In this section results are reported for three empirical studies linking education to allocative effeciency. Huffman ( ) focuses on a single dimension of allocative efficiency: the rate of farmer response to price changes for a single input. Khaldi and Fane estimate the effects of farmer education on input mix. In each of these, a production function is first estimated and from that function optimal input quantities are imputed. The discrepancy between estimated optimal and observed input use is then related to education.

The Speed of Response to Price Change

In a recently completed dissertation, Wallace Huffman analises relationships between changes in the price of nitrogen fertilizer and the speed of farmer response to these changes. Growth in using commercial fertilizers has been an important source of advance in farm production. From 1950-54 to 1964 the quantity of nitrogen fertilizer annually applied to all crops tripled in the United States and in the five states Huffman considers (Illinois, Indiana, Iowa, Minnesota and Ohio) nitrogen use increased five-fold during this period. Between 1959 and 1964, the price of nitrogen fertilizers fell from 10 to 15 percent, while corn prices increased 12 percent. Clearly, there was a growing incentive to expand the use of this input.

With Census of agriculture data for 1959 and 1964 county aggregates, Huffman estimated corn production per acre as a function of nitrogen fertilizer per acre, yield in 1949 -- to account for differences in soil productivity, average temperature and length of growing season, normal rainfall and current deviation from normal rainfall. From this estimated production function the implied marginal product of nitrogen together with the observed price of nitrogen relative to the corn price is used to impute profit maximizing nitrogen levels. Prior to 1959 the relative price of nitrogen to corn had been farily stable and after 1960, declined rapidly. Huffman estimated that farms were much closer to an economic equiliberium vis a vis nitrogen use in 1959 than in 1964 as indicated in Table 1. Most counties used insufficient nitrogen in both years,

Table 1

Maximum and Minimum County Average Rates of Nitrogen Use for Major Corn Producing Counties in Five Midwestern States, 1959 and 1964

	1959		1964	
	Actual Use	Estimated Optimum Use	Actual Use	Estimated Optimum Use
	(pounds per acre)			
Maximum	38.9	38.3	87.5	106.0
Minimum	6.4	9.9	11.4	38.3

Source: Huffman ( ), page 37.

but the gap was wider in 1964 than in 1959.

With a partial adjustment model of the form

$Z_t - Z_{t-1} = \lambda(Z_t^* - Z_{t-1})$, for which λ is the proportion of the gap between nitrogen use in 1959 and the 1964 optimum that is closed by 1964. Huffman argued this speed of adjustment measure, λ, should be related to farm operator education, average amount of contact between farm operators and county extension agents, farm size, and the deviation in 1959 between actual and optimum use. Specifically, he posited the following relationship:

$$\lambda = \exp. \{\alpha_o + \alpha_1 \text{ Ed} + \alpha_2 \text{ ExT} + \alpha_3 \text{ ExT}\cdot\text{Ed} + \alpha_4 \text{ ACRES} + \alpha_5 \Delta_{59} + u\}$$

with Ed, average farm operator education presumably speeding adjustments;

ExT, contact between extension agents and farmers easing information gathering and speeding adjustment;

Acres, the average number of acres in a corn farm serves as the "scale" incentive to make accurate decisions; and

Δ_{59}, is the descrepancy between use in 1959 and the imputed optimum.

The rational for the interaction term, Ed·ExT, is that education and extension are substitutes. Results from ordinary least squares estimates of this relationship are summarized in Table 2.

Table 2

Regression Estimates of the Partial Adjustment Coefficient

Coefficients (t-statistics)					
$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_0$
2.20 (3.80)	1.96 (2.25)	-1.07 (-2.11)	.006 (2.50)	-.01 (-2.0)	-5.24 (-5.16)

Dependent variable = $\log_e((Z_{64} - Z_{59})/(Z^*_{64} - Z_{59}))$

$R^2 = 0.25$ #Observations = 122 county aggregates

Source: Huffman ( ), page 47.

These results are impressive by any criterion. Education and extension are individually and jointly "significant". Both serve to speed adjustment toward new equiliberia and although they are substitues ($\hat{\alpha}_3<0$) each has a positive marginal effect within observed ranges of the data. The estimate, $\hat{\alpha}_4 > 0$, is the larger farms adjust more quickly than smaller ones. This, of course, is as it should be and points to scale economies in using information - the cost of $\lambda \neq 1$ is higher for larger farms. The estimated effect of the 1959 discrepancy between actual and optimum use is that firms operating further from the 1959 optimum responded more slowly to change between 1959 and 1964.

To mitigate the inference that the proportionate adjustment factor could exceed unity in the simple exponential form, Huffman also estimated a logistic function,

$\lambda = 1/(1 + \exp(X'\beta))$. Statistical results are essentially denticle, the only exception being that the t-statistic for the extension and extension-education interaction terms falls to between 1.8 and 1.9.

Huffman extimates that at the sample mean, an extra year of schooling would have increased farm profits by $94. in 1964 through this single dimension of nitrogen fertilizer use.

Least Cost Input Combinations

The procedures of Khaldi and Fane are essentially identicle. Using 1964 United States Census of Agriculture data each estimated a variant of an aggregate agricultural production function. For each observation, factor prices and the estimated production function are used to impute least-cost input combinations necessary to produce observed output. The resource wastage implied by the excess of observed cost over imputed minimum cost is related to education and other variables. The results are generally suggestive that education enhances allocative ability.

Khaldi's estimates refer to farm average quantities for state aggregates. The "states" are the thirty-nine individual states and state groups used by Griliches ( ). Farm production, defined as sales per farm is assumed to be a function of (1) land: (2) machinery; (3) labor; (4) fertilizers; (5) an aggregate for livestock services including interest charges on livestock inventories, livestock purchases and feed purchases. Machinery, fertilizers, and the livestock variable are value measures. Land and labor are alternatively expressed in value of the service flow and in physical quantities: acres and man-years, respectively. The estimated function is Cobb-Douglas and some estimates include environmental variables -- farm operator education and publicly sponsored agricultural research -- while other specifications include only the five material inputs.

Khaldi uses the alternative production function estimates to compute least-cost input mix corresponding to each level of observed output. The logarithm (base e) of observed cost relative to estimated minimum necessary cost is taken as an estimate of the fractional resource wastage. Khaldi argued that if education enhances allocative efficiency, there will be an inverse relationship between his measure of inefficiency and average levels of farm operator schooling. He argues further that since research introduces new inputs and procedures that are not costlessly evaluated, there will be a direct relation between the rate at which new things are introduced, approximated by the state specific levels of agricultural research and the measure of allocative inefficiency.

The second state of Khaldi's procedure is the regression of the logarithm of observed cost relative to estimated minimum cost on average farm operator schooling, approximated by the proportion of farm operators who had attended college, and state levels of agricultural research. He "explains" 10 to 15 percent of the between state variance of the inefficiency measure in this regression. The education effect is negative, increased schooling reduces inefficiency, and the estimated "t-statistic" varies from -2.0 to -2.3 with 36 degrees of freedom. The research effect is more ambiguous. Coefficient estimates are positive but are only slightly larger than estimated standard errors.

Khaldi then disaggregates his inefficiency measure, calculating for each of the five material inputs, the squared percentage discrepancy between observed expenditures on that input and his estimate of what expenditures would be if total cost were minimized. In subsequent regressions in which these factor specific inefficiency indexes were dependent variables, results generally confirm the aggregate result: education enhances allocative efficiency. For four of five material inputs (land is the exception) Khaldi estimated an inverse relationship between education and errors in factor use. He did not offer a point estimate of the value of schooling in this allocative dimension.

Khaldi's results are sensitive neither to inclusion of environmental variables in the production function estimates or to specification of material inputs in value or physical units.

Fane's production function estimates are farm averages for counties within four midwestern states; Indiana, Illinois, Iowa and Missouri. Output is sales per farm and the estimated production function is Cobb-Douglas. Material inputs include: livestock (10% of inventory value plus purchases of livestock and feed), expenses for seeds and fertilizers, machinery (17% of acquisition costs plus expense for gas, oil and machine hire including custom work), land and buildings (6% of value), expense rithm of average years of schooling (including college) for farm operators, and the logarithm of slaes per farm in the county.

When all inputs are considered the elasticity of cost per farm, holding constant the estimate of minimum necessary cost, with respect to operator education is -1.3 in 1964 (t-statistic = - 5.0). There were insufficient data for use family labor in 1959 to permit a similar comparison. When all inputs other than family labor were considered, there was no evidence that education affected allocative efficiency in 1959[24] and only weak support for the hypothesis in 1964.[25] In other comparisons, for which the variable input set was restricted alternatively to (1) livestock, seeds and fertilizers, and hired labor, and (2) all of these inputs plus machinery, education was estimated as significant in increasing cost efficiency. For 1964, Fane estimated the cost-efficiency marginal value product of one year of schooling to be roughly $100 at the sample mean.

Fane conclusions vis a vis scale economies in the use of information are especially interesting. Although the point estimate varies the central tendency of the coefficient is around 0.5 for livestock, seeds and fertilizers, and machinery and is 0.08 for the 1964 input aggregate. That is, Fane estimates that ceteris paribus, i.e., holding farm operator education constant, a one percent increase in output increases cost by proportionally more than is implied by cost minimization criteria. I am mildly skepticle of this result. It is, of course,

---

[24]The education coefficient is positive and the standard error is more than three times as large as the coefficient.

[25]The estimated coefficient is negative, but the standard error is only slightly smaller in absolute value than the coefficient.

possible that complexity of the decision process increases with scale, especially if the production process is not homothetic so that size expansion is not simply scalar magnification. But because the cost of incorrect choice increases with firm size the informational incentives are for an inverse relation between size and allocative inefficiency. Education is the only informational input which data permit Fane to hold constant, so that unless education is strongly complememtary to other sources of information, we might expect that the relation between efficiency and size, net of education, to be direct.

Nonetheless, if Fane's estimates are correct, then this tendency toward increasing inefficiency with larger sclae reinforces scale economies in the use of information. Fane's estimate is that

$$\frac{EC/C_*}{ES} = -1.34;$$

S refers to farm operator schooling, C indicates actual cost and C_* is cost when cost are minimized. Fane also estimates

$$\frac{Ey}{EC_*} = 1.38 \text{ (y is output)}$$

and

$$\frac{EC/C_*}{Ey} = 0.08.$$

The cost-allocative product of schooling is

$$MP_S = -\frac{\partial C}{\partial S}$$

which in the Cobb-Douglas formulation of Fane,

$$MP_S = 1.34\frac{C}{S}.$$

In the scale neutral model discribed earlier

$$\frac{EMP_s}{EC_*} = 1.$$

For Fane's model

$$\frac{EMP_s}{EC_*} = 1 + \frac{Ey}{EC_*} \quad \frac{E(C/C_*)}{Ey} = 1.11.$$

Notice that Fane's estimates do not imply that inefficiency increases with farm size when indirecteffects of the correlation between operator education and farm size is considered along with the direct effect. The net relationship is,

$$\frac{EC/C_*}{Ey} = \left.\frac{EC/C_*}{Ey}\right|_S + \left.\frac{EC/C_*}{ES}\right|_y \frac{ES}{EC_*} \quad \frac{EC_*}{Ey}$$

$$= 0.08 + (-1.34)\ (0.112)\ (1/1.38) = -0.029.$$

The estimate, $\frac{ES}{EC_*}$, is the national mean taken from Table , of the section on sclae economies. Evidently, larger farms are more allocatively efficient, this because the indirect effect of rising average schooling levels with farm size dominates Fane's estimate of the direct effect.

IV. Summary of the Empirical Results

The evidence of linkages between allocative efficiency and education of farm operators if not formidable is at least highly suggestive. There is also evidence of scale economies in using information.

In my analysis of relative wages in U.S. agriculture, roughly one-third of the wage discrepancy between persons who had and those who had not attended college is attributed to the learning-adjustment opportunities created by agricultural research. In comparing state difference in measured scale economies, I attributed 20 percent of the pro-scale bias to the information complex linking education, extension and research.

Huffman offers convincing evidence that more educated farm operators respond more quickly to input price change. He finds, further, that adjustment is faster on larger farms and that education and extension are substitute sources of information.

Khaldi and Fane offer evidence that input combinations are more nearly optimal when schooling levels are higher.

In each of these analyses, the feeding machanism between education and productivity is structural change. What then, of the return to schooling in static situations. Here, we must turn to evidence from other studies. I have described the results of Chundru for India. There is additional evidence for other countries, but in most cases the data do not permit firm conclusions. Craig Chi-yen Wu ( ) in an analysis of farm productivity in Taiwan, reports that education is more profitable on general farms where productivity advance is most rapid than on rice farms that prior to the survey, reportedly experienced fairly stable technology. George Patrick and Earl Kehrberg ( ) attempted to estimate the productivity of farmer schooling for five areas of Brazil. Their evidence is suggestive that schooling in more productive in areas using more modern techniques and that the rate of change in technique is more important than the level in determining return to schooling. Thomas Haller ( ) has preformed a similar analysis for Columbia.

142

REFERENCES

[1] Evenson, Robert E., " The Contribution of Agricultural Research and Extension to Agricultural Production," unpublished Ph.D. dissertation, Department of Economics, University of Chicago, 1968.

[2] Fettig, Lyle P., "Adjusting Farm Tractor Prices for Quality Changes, 1950-1962," Journal of Farm Economics, Vol. XLV (Aug. 1963).

[3] Friedman, Milton, Price Theory, a Provisional Text, Aldine Press, Chicago, 1962.

[4] Griliches, Zvi, "The Sources of Measured Productivity Growth: U.S. Agriculture, 1940-1960," Journal of Political Economy, Vol. LXXI (Aug. 1963).

[5] Griliches, Zvi, "Research Expenditures, Education, and the Aggregate Agricultural Production Function," American Economic Review, Vol. LIV (Dec. 1964).

[6] Huber, Paul B., "Disguised Unemployment in U.S. Agriculture, 1959," unpublished Ph.D. dissertation, Department of Economics, Yale University, 1970.

[7] Kislev, Yoav, "Estimating a Production Function from 1959 U.S. Census of Agriculture Data," unpublished Ph.D. dissertation, Department of Economics, University of Chicago, 1965.

[8] Kislev, Yoav, "Overestimates of Returns to Scale in Agriculture— A Case of Synchronized Aggregation," Journal of Farm Economics, Vol. 48, No. 4, Nov. 1966.

[9] Madden, Patrick J., "Economies of Size in Farming," Agricultural Economics Report No. 107, Economic Research Service, U.S. Department of Agriculture, Feb. 1967.

[10] Nelson, Richard R., "Uncertainty, Learning, and the Economics of Parallel Research and Development Efforts," Review of Economics and Statistics, Vol. 43, Nov. 1961.

[11] Timmer, C. Peter, "On Measuring Technical Efficiency," Food Research Institute Studies in Agricultural Economics, Trade and Development 9, No. 2 (1970): 99-171.

[12] Samuelson, Paul A., Foundations of Economic Analysis, Harvard University Press, 1947.

[13] Tweeten, Luther G., "Theories Explaining the Persistence of Low Resources Returns in a Growing Farm Economy," American Journal of Agricultural Economics, Vol. 51, No. 4, Nov. 1969.

[14] Tweeten, Luther and Schreiner, Dean, "Economic Impact of Public Policy and Technology on Marginal Farms and on the Nonfarm Rural Population," in Benefits and Burdens of Rural Development (E.O. Heady, ed.), Iowa State University Center for Agricultural and Economic Development, Iowa State University Press, Ames, Iowa, 1970.

[15] U.S. Bureau of the Census, United States Census of Agriculture, various years.

[16] U.S. Department of Agriculture, Agricultural Statistics, various issues.

[17] U.S. Department of Agriculture, Economic Research Service. Farm Income: State Estimates 1949-1966, FIS-207 supplement, 1967.

[18] U.S. Department of Agriculture, Agricultural Marketing Service, Crop Reporting Board. Farm Labor, 1949, 1954, and 1959.

[19] Welch, Finis, "Education in Production," Journal of Political Economy, Vol. 78, No. 1, Jan. 1970.

Reference List to Part A

Following the derivation of the stadardized schooling distribution for each state, the number of schooling classes is reduced from the eight census classes to three: 1–4 years, 12, and 4 or more years of college following a procedure described in Welch (1969*a*). The procedure is used in recognition of the fact that 99+ percent of the total wage variation between states and across schooling classes for the five classes, 0, 5–7, 8, 9–11, and 13–15 years of schooling is reflected in the variance of the three remaining classes, 1–4, 12, and 16+ years. Table 6 summarizes results for regressions of wages for the five excluded classes on the remaining three. Let the coefficient falling on the *i*th row and the *j*th column of table 6 be a_{ij}, and let N_i represent the age-adjusted number of persons in each schooling class. Then

$$N_i^* = N_i + \sum_{j=1}^{5} N_j a_{ij}$$

(i = 1–4, 12, and 16+) defines the estimated number of persons for each of the three classes.

References

Becker, Gary S. "Underinvestment in College Education?" *A.E.R.* (May 1960).

———. *Human Capital.* New York: Nat. Bur. Econ. Res., 1964.

Chaudhri, D. P. "Education and Agricultural Productivity in India." Ph.D. dissertation, Univ. Delhi, 1968.

Coleman, James S., et al. *Equality of Educational Opportunity.* Washington: U.S. Office of Education, 1966.

Evenson, Robert. "The Contribution of Agricultural Research to Production." *J. Farm Econ.* (December 1967)

———. "The Contribution of Agricultural Research and Extension to Agricultural Production." Ph.D. dissertation, Univ. Chicago, 1968.

Graybill, Franklin A. *An Introduction to Linear Statistical Models.* Vol. 1. New York: McGraw-Hill, 1961.

Griliches, Zvi. "Estimates of the Aggregate Agricultural Production Function from Cross-sectional Data." *J. Farm Econ.*, vol. 45 (May 1963). (*a*)

———. "The Sources of Measured Productivity Growth: U.S. Agriculture, 1940–1960." *J.P.E.* 71 (August 1963): 331–46. (*b*)

———. "Research Expenditures, Education, and the Aggregate Agricultural Production Function." *A.E.R.* 54 (December 1964): 961–74.

———. "Notes on the Role of Education in Production Functions and Growth Accounting." Unpublished paper, 1968.

Kislev, Voav. "Estimating a Production Function from U.S. Census of Agriculture Data." Ph.D. dissertation, Univ. Chicago, 1965.

Klivov-Malul, Ruth. *The Profitability of Investment in Education in Israel.* Jerusalem: Maurice Falk Inst. Econ. Res. Israel, April 1966.

Mundlak, Y., and Razin, A. "On Multistage Production Functions, the Theory of Aggregation and Technical Change." CMSBE report no. 6716, Chicago, 1967.

Nelson, R. R., and Phelps, E. S. "Investment in Humans, Technological Diffusion, and Economic Growth." *A.E.R.* (May 1966), pp. 69–75.

Schultz, T. W. *Transforming Traditional Agriculture.* New Haven, Conn.: Yale Univ. Press, 1964.

Solow, R. M. "The Production Function and the Theory of Capital." *Rev. Econ. Studies*, vol. 2 (1956).

Thomas, J. Allan. "Efficiency in Education: A Study of Mean Test Scores in a Sample of Senior High Schools." Ph.D. dissertation, Stanford Univ., 1962.

U.S. Bureau of the Census. *U.S. Census of Population: 1960.* Washington: Government Printing Office, 1963.

U.S. Commission on Civil Rights. "Equal Opportunity in Farm Programs." A report of the U.S. Commission on Civil Rights, 1965.

Welch, Finis. "Measurement of the Quality of Schooling." *A.E.R.* 56 (May 1966): 379–92.

———. "Linear Synthesis of Skill Distributions." *J. Human Resources* (Summer 1969). (*a*)

———. "Some Aspects of Structural Change and the Distributional Effects of Technical Change and Farm Programs in Agriculture." Unpublished paper, 1969. (*b*)

Appendix A

<u>The Census of Population</u> provides data for persons with 0, 1-4, 5-7, 8, 9-11, 12, 13-15, and 16 or more years of school completed. In this Appendix, I describe the computation technique used to derive wages representative of each schooling class in each state and, correspondingly, the "number" of persons in each class.

<u>Wages</u>. The wage variable refers to total income for persons in the rural farm population in 1959. Although data for earnings which exclude transfers and income from property not managed directly are preferable, they are not available. The U.S. Census (1963) provides for each state the joint incom-schooling (tables 138), age-income (tables 134), and age-schooling (tables 103) distributions for males 25 years old and over. These three distributions are used to compute the cross-products matrix required for a regression of thelogarithm of income on two classes of dummy variables, the eight schooling classes and six age classes (25-34, 35-44, 44-54, 55-64, 65-74, and 75 and over). Income estimates are interval midpoints for the $1,000 intervals from $0 to $7,000; for the interval $7,000-$9,999, $8,200 is used, and for the open-ended interval $10,000 and over, the mean is estimated from a Pareto distribution. With this cross-products matrix, the regression coefficients are computed using the standard linear regression formula. For each schooling class, the antilog of the predicted log of income for persons 45-54 years old is multiplied by the ratio of the arithmetic to the geometric mean of income for the class. This is so that estimates refer to mean rather than geometric mean values of income. In the joint age-schooling distributions, it is not possible to identify persons without

income so that all persons, with and without income, are included in the income prodictions. To correct for this error, a regression equation is estimated in which the dependent variable is the proportion of persons in an age or schooling class with income in 1959 and the independent variables are the same age and schooling "dummies." With the estimate of this equation, the probability of having income is computed for each schooling class (conditional on age = 45-54), and that probability is divided into the income estimate. The resulting wage is interpreted as the wage representative of a schooling class for males 45-54 years old with income in 1959.

Number of persons. The number of persons in each schooling class is computed in terms of male, 45-54 years old, earning units. This number is calculated as the total income of all persons in a schooling class including females 25 years old and over and young persons ages 14-24, from census tables 134. The income estimates for the closed intervals are the same as those used in the estimation of wages, but for the open-ended intervals, means are estimated usi-g separate approximations to Pareto distributions.

Following the derivation of the standardized schooling distribution for each state, the number of schooling classes is reduced from the eight census classes to three: 1-4 years, 12, and 4 or more years of college following a procedure described in Welch (1969a). The procedure is used in recognition of the fact that 99+ percent of the total wage variation between states and across schooling classes for the five classes, 0, 5-7, 8, 9-11, and 13-15 years of school-years. Table 6 summarizes results for regressions of wages for the five excluded the ith row and the ith column of table 6 be a_{ij}, and let N_i represent the age-adjusted number of persons in each schooling class. Then

$$N_i^* = N_i + \sum_{j=1}^{5} N_j a_{ij}$$

(i = 1-4, 12, and 16+) defines the estimated number of persons for each of the three classes.

Table 7

Estimated Linear Relationships Among Wage Rates (Annual Incomes) of the Eight Schooling Classes for the United States, 1959

Wage Rates Taken is Independent Variables	Wage Rates Taken as Dependent Variables, Their Coefficients and Standard Errors				
	W_o	W_{5-7}	W_8	W_{9-11}	W_{13-15}
Regression no	1c.	2c.	3c.	4c.	5c.
W_{1-4}	1.147 (.082)	.633 (.034)	.456 (.045)	.210 (.026)	-.084 (.069)
W_{12}	-.170 (.050)	.322 (.021)	.529 (.027)	.758 (.016)	.588 (.130)
W_{16+}					.431 (.075)
R^2	.991	.984	.980	.995	.981

Source. The basic data are derived as described in the above, except in this case wages refer to the total population instead of rural farm only. There are 58 observations including the 48 states of the conterminous United States with a separation of 10 Southern states into 10 white and 10 nonwhite states.

Note: Subscripts on the wage variables indicate years of school completed.

Table 8

Estimates of Factors Affecting the Productivity of More Relative to Less Schooled Persons in U.S. Agriculture, 1959; 39 "States" Only

Independent Variables	Dependent Variables W_{16+}/W_{12} (1)	W_{16+}/W_{1-4} (2)	W_{12}/W_{1-4} (3)	(4)
1. Functional illiterates	.065 (.033)	.456 (.038)	.391 (.040)	.362 (.039)
2. High School Graduates	.008 (.082)	-.471 (.098)	-.479 (.101)	-.362 (.039)
3. College Graduates	.081 (.064)	.107 (.094)	.026 (.077)	
4. Nonlabor inputs	-.154 (.033)	-.092 (.040)	.062 (.041)	
5. Research expenditures ($00)/farm	.045 (.090)	.170 (.107)	.125 (.110)	
6. Days per farm by Extension personnel	-.166 (.072)	-.173 (.086)	-.007 (.088)	
7. Intercept	1.992	2.250	0.258	0.745
R^2	.517	.844	.760	.701
Residual sum of squares (degrees of freedom)	.264 (33)	.369 (33)	.389 (33)	.482 (37)

Note: Subscripts indicate years of school completed. Standard errors of the coefficient estimates are in parentheses. For notes, see Table 1.

Table 9

Estimates of Factors Determining the Productivity of College Graduates Relative to Laborers with Conventional Skill in U.S. Agriculture, 1959; 39 "States" Only

Independent Variables	Dependent Variables, the Relative Wage (1)	(2)	(3)
1. The aggregate of functional illiterates and high school graduates	.743 (.055)	.741 (.056)	.756 (.078)
2. College graduates	-.420 (.079)	-.423 (.078)	-.756 (.078)
3. Nonlabor inputs	-.323 (.055)	-.318 (.053)	
4. Research expenditures ($00)/farm	.557 (.138)	.535 (.135)	.720 (.184)
5. Days per farm by extension personnel	-.052 (.138)		
6. Intercept	1.173	1.102	-2.334
R^2	.865	.864	.724

Note: Standard errors are in parentheses. For notes, see Table 1.

Appendix B

The Gross Revenue Formulation

Table 7 gives results for the gross value formulation. In it there are significant regional effects not captured in the other variables. The direct pecuniary bias is not significant, while other compositional variables are significant. In particular, in regression equation 2, the expenditure elasticity variables interact positively with research expenditure.[26] In the information complex, all interaction variables have the expected sign with respective t-ratios for variables II(d), II(e), and II(f) equal to 3.1, -2.6, and -1.5. Further, in comparison with equation 4 in which interaction effects are omitted and equation 3 which includes them, the own effects of the education bias, research, and extension are less impressive in the absence of interaction (only the education coefficient retains a t value near 2).[27] As in the value added function, there is a wuestion concerning the joint significance of all six variables. In equations 2 and 3 they individually exhibit significance at the 0.05 level with the exception of variable II(f). Yet to test the joint hypothesis that all six coefficients on these variables are zero, simultaneously, equations 3 and 5 are compared. The computed $F(6,37) = 3.35$ as compared to $F.05 = 2.34$. Therefore these

---

[26] In comparison with equation 3, the computed "F" statistic for the research input composition interaction variables is $F(2,35) = 4.33$ while $F(.05) = 3.27$.

[27] The computed "F" value comparing equations 3 and 4 is $F(3,37) = 2.85$ which is marginally significant at the 0.05 level.

Table 10

Regression Estimates of Factors Affecting Measured Returns to Sclae in U.S. Agriculture, 1959

Gross Value Formulation; Coefficients (Standard Errors)

Variable	Mean	Regression Number 1	2	3	4	5
I. Composition Effects						
a) Direct pecuniary bias	.008	-.498 (.366)				
b) Expenditure elasticity for land and labor	.817	.884 9.210)	.904 (.212)	1.052 (.166)	1.148 (.172)	1.043 (.173)
c) Expenditure elasticity for intermediate inputs	1.473	.399 (.161)	.272 (.132)	.543 (.103)	.560 (.106)	.520 (.112)
d) b x research expenditures		.011 (.011)	.012 (.011)			
e) c x research expenditures		.020 (.007)	.022 (.008)			
II. Information Complex						
a) Education bias	.088	1.656 (.894)	1.793 (.899)	2.523 (.930)	.848 (.440)	
b) Research expenditures	13.27	-.045 (.019)	-.049 (.019)	-.046 (.002)	-.001 (.001)	
c) Days on farms by extension personnel	.307	.495 (.311)	.641 (.295)	.744 (.306)	-.063 (.075)	
d) a x b		.119 (.045)	.137 (.044)	.098 (.044)		
e) a x c		-6.249 (2.675)	-6.990 (2.651)	-7.565 (2.798)		
f) a x b x c		-.056 (.065)	-.093 (.060)	-.104 (.065)		
III. Regional Dummies (South)		+	+	+	+	+
Intercept		.279	.408	-.170	-.074	.102
Dependent variable	1.68					
R^2		.862	.854	.818	.776	.719
RSSQ		.0848	.0894	.1116	.1374	.1727
d.f.		34	35	37	40	43

Notes to Table 10

The only difference between variables reported here and in Table 7 is that in the gross value formulation, intermediate inputs are included in both cost and revenue. Accordingly, the within-state regressions to estimate pecuniary effects, expenditure elasticities, and the education bias use a different definition of cost.

variables appear significant, and in conjunction with the evidence of equation 4 it appears that the interaction effects, which contain the information-scale hypothesis, are required to make "sense" of the individual variables II(a), II(b) and II(c).

Actually, the evidence summarized on Tables 5 and 7 are similar in most respects. The two expenditure measures are "significantly" positive and the information variables, II(a)-II(f), have the same signs and similar computed "t-ratios." The most important difference is that the pecuniary effect which is not significant in the gross revenue formulation literally swamps the value added form. Also, neither the regional dummy variables nor the research-expenditure elasticity variables are significant in the value added form.