Evaluating Agricultural Research and Productivity

Proceedings of a Symposium
Atlanta, Georgia
January 29-30, 1987

Miscellaneous Publication 52-1987
Minnesota Agricultural Experiment Station
University of Minnesota

THE AGRICULTURAL KNOWLEDGE PRODUCTION FUNCTION: AN EMPIRICAL LOOK

Philip G. Pardey*

INTRODUCTION

Economic analysis of the process of technical change has often involved macro-level studies of its causes and consequences. Relatively little attention has been given to the more fundamental knowledge generation process itself. This stems in large part from the real difficulties of obtaining appropriate indicators of research output.

The view that there exists a systematic relationship between research expenditures and knowledge increments has been taken up by numerous authors including Evenson (1968), Minasian (1969), Pakes (1978), Griliches (1979), and Kamien and Schwartz (1982). It follows naturally from the perception that, in general, science progresses by a sequence of marginal improvements rather than a series of discrete and essentially sporadic breakthroughs (see Burke [1978]).

Recent studies by Pakes and Griliches (1980), Hausman et al. (1981), and Hall et al. (1984) have sought direct estimates of the research input-output relationship for research performed by private firms in the non-agricultural sector. To date there appears to be no similar analysis of the public sector agricultural research process. The study reported here represents a first step in this direction. It develops some quantifiable indicators of agricultural knowledge production by the U.S. public sector research system and will also attempt to provide some clues as to the nature of the agricultural research spending-research output relationship.

I. MODEL SPECIFICATION

A stylized model of the relationship between the research inputs and knowledge output of the State Agricultural Experiment Stations (SAES) will be presented. It draws on the approach first sketched by Pakes (1978) and later used by Pakes and Griliches (1980) to study the patent-R&D expenditure relationship in the non-agricultural sector. The exploratory nature of this study dictates a rather parsimonious approach to modelling the knowledge production process so the model developed here represents a fairly simplified version of reality. Nevertheless it purports to be a useful framework in which to study, both the 'quality' of various publication measures as indicators of gross additions to the stock of knowledge, and the nature of the lagged relationship between research expenditures and publication output.

We begin with the notion of a quite simple knowledge production function (K.P.F.) whereby gross scientific knowledge increments, $\dot{K}^g_t$, are primarily a function of current and lagged research expenditures

$$(1) \qquad \dot{K}^g_t = g(\ c(L)R_t\ ,\ v)$$

where, $\dot{K}^g_t$ represents increments in the gross stock of scientific knowledge in time period t; $C(L)R_t$ a weighted sum of current and past research expenditures; and v a vector of other factors which contribute to $\dot{K}^g_t$. Cognizant of the panel nature of the data used here a more explicit version of (1) can be written as

$$(2) \qquad \dot{K}_{it} = \tilde{\theta} + \sum_{s=0}^{S} \tilde{\beta}_s R_{i,t-s} + \tilde{\mu}_i + \tilde{\lambda}_t + \tilde{\omega}_{it}$$

---

* Philip Pardey is jointly a Research Associate in the Department of Agricultural and Applied Economics, University of Minnesota, St. Paul and a Research Officer, International Service for National Agricultural Research, The Hague, Netherlands. Michelle Hallaway, Denise Farnwalt and Brian Stockland shared the burden of collecting the inordinately large amount of data required for this study. Errors of fact or judgement remain the author's.

where, $\tilde{\theta}$ represents an intercept term; $\dot{K}_{it}$ the gross scientific knowledge increment for state i in time period t; $\tilde{\mu}_i$ a state specific (time invariant) variable; $\tilde{\lambda}_t$ a time specific (state invariant) variable; $R_{i,t-s}$ state level deflated research expenditures lagged s time periods; and $\tilde{\omega}_{it}$ residual influences on $\dot{K}_{it}$ which are assumed to vary over both state and time periods.[1]

The $\tilde{\mu}_i$ variable represents state-specific differences in research efficiency which are assumed to be uncorrelated with time. This interpretation is analogous to similar variables purporting to measure managerial efficiency in more traditional production function studies. At a more fundamental level, the variable reflects differences in either the particular research agenda faced by each experiment station (i.e. their technological opportunities) and/or institutional factors which are conducive to a more-or-less productive research effort. While states clearly share a number of common agricultural production and distribution problems, there is certainly a significant class of problems which is state specific in nature given the varying requirements placed on agricultural production processes by location, specific economic, political and geo-climatic influences.

The $\tilde{\lambda}_t$ variable represents time specific shifts in the productivity of the research process. With much agricultural inquiry being of a downstream nature, discoveries in the complementary 'core' sciences at the upstream end of the spectrum, along with non-quantified improvements in scientific instrumentation and research hardware will act, inter-alia, to enhance the research productivity of the agricultural sciences over time. It is expected that this time specific variable would, in general, be positively related to $\dot{K}$.

Lastly, the $\tilde{\omega}_{it}$ term is taken to reflect the inherently stochastic nature of the research process or, alternatively, captures the combined influences of omitted time-varying factors specific to the i^{th} state. These three terms, $\tilde{\mu}_i$, $\tilde{\lambda}_t$ and $\tilde{\omega}_{it}$ jointly represent the variables captured by the v term in equation (1).

Given the unobservable nature of $\dot{K}$, we require a suitable indicator in order to make the model operational. State specific increments to the gross stock of scientific knowledge may be directly embodied in a variety of observable outputs. These include publications in scientific journals, patented and non-patented output such as new mechanical innovations and processes or new biological material and finally, other publications such as books, station bulletins, newsletters and the like. Of course not all new knowledge is embodied in these indicators. A certain number of findings are extended beyond the laboratory bench via direct contact with potential users either through telephone contact, public media releases or on-farm visits.

Here aggregate publication performance is used to directly proxy the quantity of agricultural knowledge produced by each of the experiment stations. Given the institutional and incentive structure under which SAES researchers operate it seems reasonable that publications more completely capture the knowledge output of the stations than alternative output proxies such as patents[2]. Publications afford researchers a means of establishing intellectual property rights over their work which will ultimately affect their salary scale, promotion rate, and tenure status. The ability to make some adjustments for variations in scientific quality also add substantially to its appeal as a direct measure of scientific output. This represents a significant advantage over studies of the patent-R&D expenditure relationship where the ability to make plausible adjustments for quality variations is not as readily available.

The quality adjustment made in this study give rise to two alternative measures of research output, namely

(1) raw publication counts, and
(2) constant quality publication counts.

The indicator function in which P_{it} represents the (constant quality) publication output of state i in period t is given by

$$(3) \qquad P_{it} = \tilde{\tilde{\theta}} + \alpha \dot{K}_{it} + \tilde{\tilde{\epsilon}}_{it}$$

where the error term is decomposed into three components such that

$$(4) \qquad \tilde{\tilde{\epsilon}}_{it} = \tilde{\tilde{\mu}}_i + \tilde{\tilde{\lambda}}_t + \tilde{\tilde{\omega}}_{it}$$

The $\tilde{\tilde{\mu}}_i$ variable represents state specific differences in the average propensity to publish. The number of publications realized from a particular level of research activity represents the joint influence of the institutional environment, as it influences the rewards accruing to those who publish, the (aggregated) utility functions of those who do the research, and the publishing traditions of the particular scientific disciplines represented at each experiment station.

For instance, SAES may vary in the emphasis they place on (non-refereed) publication mechanisms for extending the results of their research projects, depending in part on the demands of their clientele groups. This reflects the derived demand aspects of publication output. Furthermore, SAES administrators may vary in the emphasis they place on publication output as an indicator of the research productivity of their research staff. To the degree this affects their reward structure it acts as an incentive or disincentive to generate publications and thereby affects the supply of publication output (see Hansen et al. [1978])[3]. As a qualification to these influences it is likely that researchers effectively operate in a national market so that state level differences in publication incentives may not have a strong or measurable impact on the propensity to publish.

The $\tilde{\tilde{\lambda}}_t$ variable captures influences on the propensity to publish which change over time. Finally the $\tilde{\tilde{\omega}}_{it}$ variable reflects variation in the propensity to publish not accounted for by the time or state specific effects. In this study the total publication output for each state is derived from the publication performance of a stratified random sample of researchers. Consequently $\tilde{\tilde{\omega}}_{it}$ also captures sampling errors in the measurement of P_{it}.

Imposing orthogonality on the $\dot{K}$ and $\tilde{\tilde{\epsilon}}_{it}$ variables we will assume that the $\tilde{\tilde{\mu}}_i$, $\tilde{\tilde{\lambda}}_t$ and $\tilde{\tilde{\omega}}_{it}$ components of $\tilde{\tilde{\epsilon}}_{it}$ are also uncorrelated with the determinants of $\dot{K}$ given by (2). Substituting equations (2) and (4) into (3) gives the reduced form equation relating lagged research expenditures to publication output such that

$$(5) \qquad P_{it} = \theta + \sum_{s=0}^{S} \beta_s R_{i,t-s} + \mu_i + \lambda_t + \omega_{it}$$

where

$$\theta = (\tilde{\tilde{\theta}} + \alpha\tilde{\theta})$$

$$\mu_i = (\tilde{\tilde{\mu}}_i + \alpha\tilde{\mu}_i)$$

$$\lambda_t = (\tilde{\tilde{\lambda}}_t + \alpha\tilde{\lambda}_t)$$

$$\beta_s = \alpha\tilde{\beta}_s$$

$$\omega_{it} = (\tilde{\tilde{\omega}}_{it} + \alpha\tilde{\omega}_{it})$$

$$i=1,2,\ldots..N \; ; \; t=1,2,\ldots..T$$

The state specific term μ_i represents the weighted sum of both knowledge production and subsequent publication performance influences which are specific to the various states. Likewise the λ_t variable represents the weighted sum of these same influences which are of a time specific nature. When estimating equation (5) it is clear that the response of $\dot{K}$ to a unit change in research expenditure, R, cannot be identified given the information contained in this model. Nevertheless, the form of the distributed lag linking $\dot{K}$ to R can be investigated by normalizing the estimated lag coefficients to obtain $\tilde{\beta}_s / \Sigma\tilde{\beta}_s = \beta_s / \Sigma\beta_s$ for all s.

The research production process, represented by equation (5) or variants thereof constitutes the primary focus of the subsequent empirics.

II. DATA AND ESTIMATION ISSUES

We begin with a look at the publication-based proxy of SAES knowledge output and its associated quality adjusters. Given the paucity of previous work in this area, some attention will be given to the conceptual issues involved along with the mechanics of variable construction.

A. The Quantity Dimension of Research Output

The aggregate publication performance of each station was estimated on the basis of the average research performance of a random sub-sample of station researchers. First the researcher population of each experiment station, for the two fiscal years 1970/71 and 1974/75, was established by reference to the appropriate Cooperative State Research Service (CSRS) listing of Professional Workers in State Agricultural Experiment Stations and Other Cooperating State Institutions. All individuals who, on the basis of their location (i.e. branch versus main station), degree, appointment and/or professional status, could be reasonably classified as support or auxiliary staff were excluded. A population of around 10,000 researchers was identified, from which a stratified random sample consisting of around twenty percent of the population was chosen[4].

Table (1) indicates that overall the samples closely conform to the research disciplines represented in the SAES. Just over 50 percent of these researchers are associated with one of the plant science disciplines while 30 percent are spread amongst the animal sciences. Given the constant turnover of research personnel at any particular SAES, the 1970 to 1973 publication performance of each station was estimated using the 1970/71 sample of researchers (i.e. Sample I) whilst the 1974 and 1975 publication record used the 1974/75 researcher sample (i.e. Sample II). Both these samples constitute independent draws from the population of SAES researchers and represent 23.2 and 22.2 percent of the 1970/71 and 1974/75 population respectively.

The scientific publication output of each sample researcher for the appropriate years was obtained by a manual search of the respective Source Indexes of the Science Citation Index (SCI) compiled by the Institute for Scientific Information[5]. The overall data set involved separate records for 16,050 publications. This data was then integrated with biographical information concerning source authors (i.e. station and discipline affiliation, degree, appointment and professional status, etc.).

Some summary publication statistics are presented in Tables I.1 and I.2, Appendix I. Nearly two-thirds of the publications are full-length articles. Most of the remaining publications are abstracts of papers presented at professional meetings or shorter notes. From Table I.2 we observe that around 88 percent of the publications are jointly authored with the majority having one or two co-authors. Given this high percentage of co-authored articles we proceeded to calculate both the average number of articles and average number of prorated articles per sample researcher per year for each station. The raw and prorated averages were scaled by the appropriate population number of researchers per station to yield estimates of the total publication output for each station for the 6 years 1970-1975[6].

B. The Quality Dimension of Research Output

One of the most frequently encountered criticisms of direct indexes of scientific output is that the unit of measurement does not adequately account for likely quality differences. Two notions of quality are possible. In an economic context quality can be taken to mean the relationship between research benefits and the measured level of resources committed to that research. Hence, higher quality research generates a larger present value benefit stream for a given level of measured research inputs. Scientific quality according to Evenson and Wright (1982) is measured by conformity to standards established by scientific work at the frontier of the discipline. More generally the scientific quality of a researcher (or a body of research) is assessed with respect to the relative significance of the individual's

Table (1) Researcher Statistics - Number of Researchers per Discipline; Sample and Population Averages [1]

Discipline [2]	1970-73 Sample I	1974-75 Sample II	1970-75 Sample	1970-75 Popln.
	Average Number of Researchers			
Agronomy	539 (24.1)	553 (24.4)	546 (24.2)	2012 (20.2)
Entomology	172 (7.7)	195 (8.6)	184 (8.2)	880 (8.8)
Forestry	184 (8.2)	177 (7.8)	181 (8.0)	919 (9.2)
Horticulture	138 (6.2)	146 (6.4)	142 (6.3)	663 (6.7)
Plant Diseases	158 (7.1)	152 (6.7)	155 (6.9)	725 (7.3)
TOTAL PLANT SCIENCE	1191 (53.2)	1223 (53.9)	1207 (53.6)	5199 (52.3)
Animal Science	286 (12.8)	257 (11.3)	272 (12.1)	1019 (10.2)
Dairy	55 (2.5)	47 (2.1)	51 (2.3)	226 (2.3)
Fisheries	25 (1.1)	29 (1.3)	27 (1.2)	213 (2.1)
Poultry	43 (1.9)	42 (1.9)	43 (1.9)	197 (2.0)
Vet. Medicine	260 (11.6)	236 (10.4)	248 (11.0)	1325 (13.3)
TOTAL ANIMAL SCIENCE	669 (29.9)	611 (27.0)	640 (28.4)	2980 (30.0)
Agric. and Envir. Science	44 (2.0)	64 (2.8)	54 (2.4)	219 (2.2)
Core Sciences and Statistics	225 (10.0)	250 (11.0)	238 (10.6)	896 (9.0)
Genetics and Plant Breeding	34 (1.5)	13 (0.6)	24 (1.1)	118 (1.2)
Nutrition and Food Science	73 (3.3)	102 (4.5)	88 (3.9)	524 (5.3)
Other	3 (0.2)	4 (0.2)	4 (0.2)	13 (0.1)
TOTAL	2239	2267	2253	9949[3]

(1) Figures in parentheses are percentages of the respective SAES totals.

(2) Agronomy includes plant science, soil science and water and range science. Forestry includes wildlife. Horticulture includes pomology, vegetable crops, viticulture and oenology. Plant diseases include nematology and plant pathology. Animal science includes animal husbandry. Fisheries includes aquaculture. Vet. Medicine includes animal diseases and microbiology. Environmental science includes landscape architecture. Core sciences includes biochemistry, biophysics, biology, botany, chemistry and zoology.

Scientists were generally allocated to research disciplines on the basis of the discipline classification reported in the listing of "Professional Workers in SAES and other Cooperating Institutions". Over time and between station inconsistencies were resolved by reference to the research specialization which was also recorded for each researcher. Multiple counts (i.e. where the same researcher is listed more than once per state, say, at the main as well as sub-station[s]) were eliminated by cross-matching the discipline listing with an alphabetical listing also recorded in the "Professional Workers" publication.

(3) This figure excludes 424 researchers who were also listed at various substations but could not be allocated to a specific discipline. Also excluded are social science and agricultural engineering researchers. Adding all these researchers to this figure would increase the overall population size to 12,270.

contribution to their field. One method commonly used to establish significance entails a peer group review procedure (see Shaw [1967]). This subjective evaluation technique is limited by the problems of standardization of evaluation criteria and the individual biases and information base of the evaluators. Furthermore such exercises are essentially nonreplicable.

An alternative procedure is to measure the subsequent citation performance of a piece of research. The maintained hypothesis is that on-average the cited work is a useful input in the production of current research. Thus citation performance is a quantifiable measure of the impact of published work on the future knowledge (publication) output of the profession. This line of argument has been used in the recent economics literature by McDowell (1982) and Davis and Papanek (1984).

One of the major advantages of a citation based measure of scientific quality is its ability to capture not only the spatial impact of a piece of research (i.e. across disciplines) but also the temporal dimension of its influence. It seems reasonable to suggest that, ceteris paribus, articles which are cited more recently have a greater impact on the research process than those which are cited at a later point in time. The earlier an article is cited the greater its indirect impact on future research if the citing article is in turn cited by other researchers. Thus, in a present value context the citation profile and not simply the cumulative citation performance of a body of research is the important determinant of its overall impact on subsequent research.

As the present study involved observations on 16,050 publications it was well beyond our resources to map the citation performance of each article over a number of years. Rather than simply measure the cumulative citation rate of each article, and so miss the temporal dimension of its citation performance, it was decided to standardize the citation performance of each article on a particular year - in this case two - following its publication date.[7].

C. Research Inputs

To estimate equation (5) requires an accurate measure of the resources used by the SAES for research endeavors. Many previous augmented production function studies have used a fairly restricted set of research expenditure estimates. However, given the variety of construction and mismeasurment errors which were identified in the sources commonly used to date (for instance Latimer [1964], Cline [1975], and Davis [1979]) in this study we opted to construct a research input estimate based entirely on original sources[8].

Research expenditures for each SAES were split to factor level. Capital investments were further subdivided into plant and equipment plus land and building purchases, while 'labor', or more specifically non-capital expenditures included salaries, fringe benefits, and operating expenses. While all 'labor' related costs are appropriately expensed in the year of purchase, this is not so for capital. Capital purchases or investments represent gross additions to the capital stock at any point in time and what is required (see Griliches [1960], and Yotopolous [1967]) is a measure of the current productive service flow of the capital assets, which is viewed as an estimate of the capital resources used in the current period.

Non-capital expenditures were deflated using an index of average university salaries, since this component was primarily salaries of research and support staff. Land and building expenditures were deflated using the Handy-Whitman index of public construction prices. The implicit price deflator for State and Local Government Purchases of Goods and Services was used to deflate plant and equipment expenditures[9]. After deflating, the total real service flows derived from plant and equipment, and land and buildings were calculated. Land and buildings were assumed to have a 25 year service life, and plant and equipment a 10 year service life. Additionally, both service flow profiles were proxied by a One-Hoss Shay assumption with zero salvage value. These two series were summed to get a measure of the total service flows arising from capital investments. Together with the deflated non-capital expenditures they provide a reasonably accurate measure of the real resources committed by the SAES to the knowledge production process over the 1963 to 1975 period.

D. Estimation Issues

When pooling time-series, cross-section data the most contentious issues concern the nature of the state and time specific variables and the related, but separate issue concerning the relationship between these effects variables and the other regressors (Hsiao [1986]). Treating μ_i and λ_t as fixed variables, the resulting covariance or fixed effects model can be estimated by including these variables as separate time and state specific dummy variables and applying OLS. This least squares dummy variable (LSDV) estimation technique not only gives unbiased and efficient estimates of β_k but also gives direct estimates of each time and state specific effect which are unique up to a normalization. Thus, under a fixed effects regime we can obtain BLU estimates of the slope coefficients by applying OLS to the transformed equation[10],

$$(6) \qquad (P_{it} - \bar{P}_{i.}) = \sum_{k=1}^{S} \beta_k (R_{kit} - \bar{R}_{ki.}) + \omega_{it} - 1/T\sum_{t=1}^{T} \omega_{it}$$

It is clear from this equation that we utilize only the variation within each state (i.e. the over-time variation about state means) to derive what is called the within estimates. By completely ignoring the between state variation we consequently eliminate a major portion of the variation among both the explained and explanatory variables if the between effects variation is large.

To incorparate the between effects information into the estimation procedure we can treat the time and state specific effects as random variables, giving rise to the variance components or random effects model. This generates an error structure which is no longer independent and identically distributed so that a generalized least squares (GLS) procedure is required in order to obtain unbiased and efficient estimates of the β's[11]. Using the method suggested by Fuller and Battese (1969) we can obtain GLS estimates by applying OLS to the transformed equation,

$$(7) \qquad P_{it} - \gamma\bar{P}_{i.} = (1 - \gamma)\,\bar{\theta} + \sum_{k=1}^{K} \beta_k (R_{kit} - \gamma\bar{R}_{ki.}) + \omega_{it} - \gamma\bar{\omega}_{i.}$$

where $\gamma = 1 - \sigma_\omega/\sigma_1$ and $v_{it} = \omega_{it} - \gamma\bar{\omega}_{i.}$ is i.i.d.

As Mundlak (1978) observed, this GLS procedure ignores the consequences of the correlation which may exist between the effects and the explanatory variables. In the context of our model, this is equivalent to assuming that differences in research efficiencies are correlated (on average) with interstate differences in research expenditures. Pakes and Griliches (1980) argue that in general, at the firm level, differences in $\tilde{\mu}_i$ are transmitted to differences in average research expenditures, $\bar{R}_{i..}$, with more efficient research departments being allocated more research funds. This being the case we could then form the auxiliary regression

$$(8) \qquad \tilde{\mu}_i = \sum_{s=0}^{S} \phi_s \bar{R}_{i.,-s} + \tilde{\mu}_i^*$$

where

$$\bar{R}_{i.,0} = T^{-1}\sum_{t=1}^{T} R_{i,t}, \qquad \bar{R}_{i.,-1} = T^{-1} \sum_{t=0}^{T-1} R_{i,t-1} \qquad \text{etc.}$$

substituting equation (8) into (2), (3), and (4) we can rewrite (5) as

$$(9) \qquad P_{it} = \theta + \sum_{s=0}^{S} + \beta_s R_{i,t-s} + \sum_{s=0}^{S} \phi_s \bar{R}_{i..,-s} + \mu_i + \lambda_t + \omega_{it}$$

where $\quad \tilde{\mu}_i = (\tilde{\mu}_i + \alpha\tilde{\mu}_i^*), \quad \phi_s = \alpha\tilde{\phi}_s$

and all other variables are defined as before.

In contrast to the private sector research process it is not clear, a priori, that the research efficiency-research expenditure relationship described by equation (8) holds in the case of public sector agricultural research. This ambiguity arises in part from the mechanisms whereby the benefits from public sector agricultural research are appropriated. In the Pakes-Griliches case, private firms directly capture (part of) the returns from research activity presumably via patents, new product or process developments or licensing agreements. Assuming profit maximizing behavior, Mundlak and Hoch (1965) have shown that the unobservable part of the K.P.F. is transmitted to a research input demand equation given by the first order profit maximizing conditions.

However, in the case of publicly funded agricultural research much of the benefit is not directly appropriated by those institutions and/or individuals undertaking research, but instead is filtered through a political mechanism. Several authors (Guttman [1978], Rose-Ackerman and Evenson [1982], and Hadwiger [1982]) have suggested, that in the case of publicly funded agricultural research, political rather than just economic efficiency criteria influence the allocation of research resources. Within this context they identified a variety of factors correlated with increased appropriations to the SAES. Whether or not these factors are systematically related to the relative research efficiency of the SAES is not readily apparent so that the feedback mechanism described in the profit maximizing case above may not be operative here.

One approach which can resolve the issue, for the purposes of estimation, is simply to proceed with the LSDV approach and make inferential statements concerning the estimated β's conditional on the realized values of μ_i and λ_t in the sample. These LSDV estimators are, conditional on the μ_i and λ_t in the sample, best, linear, and unbiased. Because this approach makes no specific assumptions about the distribution of μ_i and λ_t it can be used for a wider range of problems. Nevertheless, if the restrictive distributional assumption of the variance component model is correct, then using this additional information will result in a more efficient estimator. As the superiority of the variance component over the covariance model is jeopardized in the presence of correlation between the effects and explanatory variables, this issue will also be explored in the empirical work to follow.

An unbiased estimate of σ_1, and σ_w can be derived from

$$\hat{\theta}_1^2/T = \hat{\nu}'\hat{\nu}/N\text{-}K$$

where $\hat{\nu}'\hat{\nu}$ is the sum of squared residuals from applying OLS to the between equation (i.e. where all variables enter as only state, over-time means) and

$$\hat{\sigma}_\omega^2 = \hat{\omega}'\hat{\omega} \;/\; N(T\text{-}1) \;-\; K'$$

$$\text{with } K' = K\text{-}1$$

where $\hat{\omega}$ is the estimated residuals from applying OLS to equation (6).

Finally, an estimator for $\hat{\sigma}_\mu^2$ can be obtained from

$$\hat{\sigma}_\mu^2 = (\hat{\sigma}_1^2 - \hat{\sigma}_\omega^2)/T$$

Notice there is no guarantee that $\hat{\sigma}_{\mu}^{2}$ is positive. A negative value may well be an indication that the random effects model is misspecified. For instance the maintained hypothesis of constant β over time and/or states may be in error or it could well be that the independence assumption $E(\mu_i | X) = 0$ is violated.

III. RESULTS

The average publication performance per researcher (per station) per year is given by the PUB and PROPUB variables in Table (2). Averaging across all disciplines and SAES gives a yearly publication output measure (PUB) of around 1.2 which is approximately halved to 0.53 if publications are prorated (PROPUB) according to the number of coauthors per publication[12]. Given the relatively small proportion of sole authored publications (averaging only 12.4 percent over the sample) this is to be expected.

The two research quality weights used in this study are the TOTCIT and NETCIT variables respectively. Both measure the average yearly citation performance of publications two years following their publication date. The TOTCIT variable captures the total number of citations for this period and the NETCIT variable nets out self-citations by both source and coauthors. Despite the large range in both these figures, their relatively low coefficient of variation suggests that the average citation performance of quite a few stations is at the lower end of the data range. Both measures show that for at least one of the sample years, 1970-75, none of the publications from Delaware were cited whilst Wisconsin achieved the highest total and net citation rate of 3.84 and 3.29 respectively.

It was argued earlier that prorated publication output, weighted by net citations, is an appropriate indicator of the overall performance of the SAES. Nevertheless an analysis of the relationship between these various measures, afforded by the correlation matrix in Table I.3, Appendix I, is instructive. A $\rho = 0.947$ indicates a strong positive relationship between the two quantity measures PUB and PROPUB with an even stronger relationship, $\rho = 0.984$, holding between the quality measures TOTCIT and NETCIT. In contrast, the relationship between the various quantity and quality measures is far less definitive. These results show that the systematic variance ratio from a regression of either quantity measure on either quality measure ranges from 0.126 to 0.140. Thus, at the station level, there is a positive but reasonably loose association between research quality and quantity.

The simple correlation coefficients in Table (I.3) also suggest the results from estimating equation (5) will be sensitive to the form of the dependent variable. Nevertheless on statistical grounds, the four quality adjusted research output indicators appear to be good proxies for each other. On conceptual grounds, the PRONET variable is the most appealing and will form the benchmark in the regression analysis to follow.

Prior information concerning the appropriate form of the relationship between publication output and research expenditures described by equation (5) is sparse. To keep a specification search within manageable proportions, we restricted our choice to linear, log linear, semi log and double log forms. The regressors were log and linear current expenditures on labor and estimated capital service flow, additive time and state dummies and multiplicative size dummies. Taking medium sized stations as the reference group, the size dummies allowed the expenditure coefficients potentially to vary for small (<100 researchers), medium and large (>500 researchers) stations. The specification with log P was preferred over the linear P model[13]. By applying a series of standard Fisher F-tests the data also suggests that a double log model is preferred over a log linear specification, size effects (at least as they impact the slope coefficients) are degenerate, and significant trend and state effects are present.

One of the stated goals of this study is to inquire into the strength of the systematic relationship between SAES research expenditures and publication based indices of research output. Various measures of research input and output were presented earlier and, given the panel nature of the data set, it is possible to partition their variance into several dimensions. This is done in Table 3.

Table (2). Descriptive Statistics for Selected Publication Based Agricultural Research Output Measures.[a]

Variable	Mean	Standard Deviation	Minimum	Maximum
Population	255.4	162.96	44.0	727.0
Per Researcher:				
Pub	1.159	0.471	0.160	3.019
Propub	0.527	0.211	0.093	1.242
Totcit	0.960	0.570	0	3.835
Netcit	0.790	0.580	0	3.288
Per Station:[b]				
Pub	330.08	238.37	8.80	1501.7
Propub	151.24	134.25	5.312	698.4
Pub x Net[c]	321.24	452.31	0.0009	3004.4
Pro x Net[c]	145.96	204.29	0.0005	1244.1

(a) These figures are derived from a panel set consisting of observations on 48 states (Alaska and Hawaii omitted) for the period 1970-75 inclusive.

(b) Per station figures are simply the per researcher figures weighted by the appropriate researcher population figure.

(c) When used as a quality weight, the zero citation count was arbitrarily set at 0.0001. This allowed logarithmic values to be calculated.

In all cases the between (states) dominates the within (states) variance with the various output measures showing relatively more variation in the within or over time dimension than the expenditure figures. Given the relatively large spread of station sizes and the short nature of the panel (8 years), this is not unexpected. Weighting the raw publication count measures (PUB and PROPUB) by the net citation count increases the variation proportionately more in the within than the between dimension.

Given the earlier assertion that the error term, $\tilde{\omega}_{it}$, in the knowledge production function (equation [2]) is orthogonal to $\tilde{\tilde{\omega}}_{it}$, the error term from the indicator function (equation [3]), we can write

$$R^2_{P.\dot{K}} = 1 - \text{var}(\tilde{\tilde{\omega}}) / \text{var}(P)$$

and

$$R^2_{P.\Sigma R} = 1 - (\alpha^2 \text{var}(\tilde{\omega}) + \text{var}(\tilde{\tilde{\omega}})) / \text{var}(P)$$

It follows that the coefficient of determination from a regression of P on current and lagged R gives the lower bound of the systematic-to-total variance ratio of P as a measure of knowledge increments $\dot{K}$. In this sense, the $R^2_{P.\Sigma R}$ statistic gives the 'proxy error' which follows from using (weighted) publications as an indicator of gross additions to the stock of knowledge.

From the results presented in Table (4), we observe that, in the total or OLS relationship of quality adjusted publication counts on summed research capital and labor expenditures, no more than 47 percent of the variation in P can be attributed to its 'error' as a measure of knowledge stock increments. Of course the actual 'proxy error' may be far less than this figure if the noise in the knowledge production function were known to dominate the overall error term ω_{it} in equation (5).

Partitioning the total relationship into its between and within components we observe, using the complete sample, that the 'proxy error' drops to a maximum of 25 percent in the between dimension. Thus, the on-average knowledge increment-lagged expenditure pattern for each experiment station appears to be captured fairly completely using quality adjusted publications as a proxy for knowledge increments. However, for the publication-expenditure relationship in the within dimension R^2's of the order of 0.07 were recorded. From the data available to us, we cannot identify the source of this large random component in yearly deviations of each SAES from its average level of operations. It would arise if the publication process were subject to a great deal of instability over time. Alternatively it may be that the knowledge production process is such that, within the range of our data at least, small fluctuations over time in the research expenditures of a particular station are not systematically translated to changes in research output ($\dot{K}$) for particular years.

Qualitatively these results are surprisingly similar to those obtained by Pakes and Griliches in their 1980 study of the firm level patent-R&D expenditure relationship. Using raw patent data over an eight year period from 1969-1975, and research expenditure data over the 1963-75 period for 121 medium and large U.S. corporations, they estimated the double log relationship between patents and current and (five years) lagged R&D expenditures. They obtained R^2's for the total regressions ranging from 0.74 to 0.95, for the between dimension ranging from 0.77 to 0.97, and for the within dimension ranging from 0.11 to 0.49[14].

In all dimensions their R^2's were somewhat higher than those obtained from this study. This may simply result from the bigger data set used in their investigation. However, it may also arise in part from the use of raw patent counts to proxy research output. In our study using the raw publication count (PROPUB) rather than the citation adjusted measure (PRONET) in general caused the R^2's to increase. It may also be that the biologically orientated research of the SAES is inherently more noisy than the research and development projects undertaken by the corporations in the Pakes-Griliches sample[15]. (i.e. the var (ω) term is relatively larger for biological as opposed to industrial research). Alternatively patenting activity, being

Table (3) Partitioning of Sample Moments for Various Agricultural Research Output and Expenditure Measures (N = 48, T = 6)[a]

	Between Variance[b]	Within Variance[c]	Ratio of Within to 'Total' Variance
Pub	5.479	0.050	0.009
Propub	5.465	0.061	0.011
Pubnet	11.097	0.650	0.055
Pronet	11.100	0.630	0.054
Labor	3.102	0.007	0.002
Capital	3.594	0.017	0.005
Labor + Capital	3.342	0.006	0.002

(a) All (state-level) research output and expenditure variables are measured in their natural log form.

(b) Given by $\frac{T\Sigma(X_{i.} - X_{..})}{N-1}$

(c) Given by $\frac{T\Sigma(X_{it} - X_{i.})}{N(T-1)}$

Table (4) Regression of Citation Adjusted Publication Output (PRONET) on Agricultural Research Expenditure[a].

Variable	OLS	WITHIN	BETWEEN	EGLS[b]
R_0	-0.5994 (1.0050)[c]	0.0252 (0.8050)	-5.4798 (10.735)	0.0204 (0.8473)
R_{-1}	-0.0075 (1.2889)	0.0789 (0.8986)	-5.4824 (21.584)	0.3080 (0.9920)
R_{-2}	-0.1180 (1.2709)	-0.9056 (0.8992)	8.3317 (16.079)	-0.3187 (0.9835)
R_{-3}	1.0557 (1.2559)	-0.1406 0.8817	24.977 (14.755)	0.6920 (0.9640)
R_{-4}	0.2871 (1.2745)	-0.4862 (0.8931)	-28.982 (19.648)	0.3462 (0.9754)
R_{-5}	-0.0557 (1.2994)	-0.3574 (0.8868)	20.689 (18.662)	-0.0226 (0.9938)
R_{-6}	-0.0823 (1.3153)	-0.1552 (0.9166)	-33.306 (17.776)	0.3767 (1.0132)
R_{-7}	0.9123 (1.0260)	-1.5631 (0.8741)	20.828 (9.4215)	0.0902 (0.8635)
Trend	0.0377 (0.0384)	177.37 (43.095)	-	0.0473 (0.0301)
ΣR_{-i}	1.5567 (0.0896)	-3.5041 (1.3761)	1.5740 (0.2005)	1.4922 (0.1546)
R^2	.5315	.0660	.7521	.2737
NT	288	288	48	288
'Mean' Lag	3.76	4.83	4.47	3.41
χ^2				16.395

(a) All variables measured in natural logs.

(b) For the estimated GLS procedure, all variables, Z_{it}, are transformed as $Z_{it} - \hat{\gamma} Z_{i.}$ where

$\hat{\gamma} = 1 - \hat{\sigma}_\omega/\hat{\sigma}_{1.}$ Here $\hat{\sigma}_1^2 = 3.3162$, $\hat{\sigma}_\omega^2 = 0.6611$ and $\hat{\sigma}_\mu^2 = 0.4509$.

(c) Standard errors in parentheses.

Table (5) Regression of Citation Adjusted Publication Output (PRONET) on Agricultural Research Expenditure[a].

Variable	OLS	WITHIN	BETWEEN	EGLS[b]
R_0	0.3367 (0.8707)[c]	0.4875 (0.5724)	-6.9013 (10.961)	0.3042 (0.6123)
R_{-1}	-0.7926 (1.1479)	-0.6302 (0.6566)	-1.8776 (23.569)	-0.6316) (0.7096)
R_{-2}	0.2694 (1.1465)	-0.1402 (0.6558)	15.791 (19.031)	-0.0894 (0.7124)
R_{-3}	0.9849 (1.1819)	1.0121 (0.6733)	1.9235 (16.713)	1.0693 (0.7330)
R_{-4}	0.5729 (1.1248)	0.5282 (0.6249)	11.091 (21.574)	0.5689 (0.6929)
R_{-5}	-0.6560 (1.140)	-0.8193 (0.6388)	13.839 (18.441)	-0.7484 (0.7044)
R_{-6}	0.0359 (1.170)	0.5086 (0.6562)	-29.706 (17.344)	0.5151 (0.7263)
R_{-7}	1.2883 (0.8727)	0.1079 (0.5453	19.401 (9.4456)	0.3270 (0.5913)
ΣR_{-i}	1.3623 (0.0809)	1.0548 (0.5923)	1.3784 (0.0474)	1.315 (0.1498)
R^2	.5320	.0352	.6877	.2421
NT	264	264	44	264
'Mean' Lag	3.87	3.30	4.64	3.62

(a) Four outlier states (Delaware, Maine, Nevada and New Mexico) and trend variable omitted. All variables measured in natural logs.

(b) See note (b) Table (4.4). Here $\hat{\sigma}_1^2 = 3.0410$, $\hat{\sigma}_\omega^2 = 0.2640$ and $\hat{\sigma}_\mu^2 = 0.4628$.

(c) Standard errors in parentheses.

the outcome of a corporate decision making process, may be less sporadic in nature (particularly in the over time or within dimension) than the publishing process, which represents in part the aggregate publishing propensities of individual researchers within each of the SAES.

A primary objective of this study is to gain some insight into the nature of the lagged relationship between research inputs and outputs. Following the procedure suggested by Hatanaka and Wallace (1980), we estimated the distributed lag relationship between research inputs and research outputs in a form-free manner. This is an appropriate approach to take if we are interested in capturing certain features of a lag distribution, such as the long run lag (i.e. the sum of the lag coefficients), the mean lag, and the variance of the lag, with few ad hoc constraints being imposed on the lag distribution.

The results in Tables (4) and (I.4) in Appendix I, indicate that the individual lag coefficients are estimated with a low degree of precision. However, the sum of the lag coefficients, measuring the long run expenditure response of (quality adjusted) publication output, is estimated quite precisely. As expected, the OLS and between estimates are very close. They show that a 1 percent once and for all increase in real research expenditures leads to around 1.6 percent increase in constant quality research (publication) output. The output response for unadjusted research output is somewhat lower at around 1.2. Thus the use of quality unadjusted publication counts underestimates the research output response resulting from increased research expenditures by approximately 25 percent. This result holds across most of the specifications reported here.

The EGLS long run elasticity estimate appears to be dominated by the between variation and is at odds with the within estimate. Given the truncated nature of the panel, this is not surprising but it was decided to investigate several reasons which could account for this discrepancy. The first concerns misspecification of the random effects model. A violation of the orthogonality assumption concerning the state effects variable μ_i and the X_{it}'s (i.e. $E(\mu_i | X_{it}) \neq 0$) causes the random effects estimator to be biased and inconsistent, while having no impact on the fixed effect estimator. Hausman (1978) suggests a natural test of the null hypothesis of independent μ_i's is to consider the difference between the two estimators, $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{RE}$. If no misspecification is present then $\hat{q}$ should statistically be near zero[16].

The relevant χ^2 statistics for a Hausman test presented in Tables (4) and (I.4) suggest there is no significant misspecification in the random effects model for either the PRONET or PROPUB case. The maintained hypothesis that the X_{it}'s and the μ_i variable are orthogonal is not rejected. This contrasts with the Pakes-Griliches study of the private sector patent production process where firm specific effects were correlated with research expenditures. The funding mechanism regarding public sector agricultural research is such that the direct or indirect link between these state specific effects and the level of research expenditures appears more tenuous than in the case of private sector funding.

For an alternative explanation we observe that the within estimate of the expenditure coefficients for the PRONET model is extremely sensitive to the inclusion or omission of a trend variable. The deviation of slowly changing research expenditures around an over time mean is not only small but appears to be highly collinear with a simple trend variable. The spread between the within and EGLS estimates drops sharply when the trend variable is omitted[17]. In contrast, the OLS and EGLS estimates of the long run response of research output are relatively insensitive to changes in the trend variable specification.

Table (5) shows the difference between the EGLS and within result for the PRONET model is further reduced by the omission of four outlier states - Delaware, Maine, Nevada and New Mexico. For at least one of the six years in the sample, these states exhibit a calculated error $|P_{it} - \hat{P}_{it}|$ greater than twice the standard deviation of the estimate. Omitting them from the sample causes the sum of squares residuals to drop by approximately 62 percent. The precise cause of this 'deviant' research performance is not clear but could be related to the relatively small size of these stations. In particular, there may exist an aggregation effect whereby the summed research output of a small number of researchers is more volatile than for a somewhat larger research organization. Unfortunately, corroborative evidence on such a size effect is difficult to come by.

Using the absolute value of the estimated lag coefficients we can construct point estimates of the 'mean' of the normalized lag distributions, m_i, such that,

$$\hat{m}_i = \hat{\mu}_i / \hat{\mu}_o$$

$$\text{where } \hat{\mu}_i = \sum_{s=0}^{K} s^i \, |\beta_i| \qquad i = 0,1,\ldots\ldots,k.$$

Various estimates of the 'mean' lag were given in the preceding tables and are averaged in Table (6) along with comparative figures from some other studies. The gestation lag represents the average lag between project inception and completion, while the time from project completion to commercial application is given by the application lag.

Averaging the Rapoport (1971) and Wagner (1968) figures gives a mean gestation lag of around 1.34 years which is close to the Pakes-Griliches (1980) estimate of 1.6 years. The quality unadjusted estimate from this study, which is closest to the output measures used by these comparative studies, suggests that the mean gestation lag for public sector biologically orientated research is around 0.7 to 1.0 years longer than the private sector manufacturing oriented research. The different nature of both the research problems and the institutional environment could account for this difference.

Moreover, the mean gestation lag between research expenditures and quality adjusted research output is consistently longer than the quality unadjusted output measures. The summary figures in Table (6) show that for the case of public sector agricultural research, the quality adjusted 'mean' lag is approximately six months (19 percent) longer than the quality unadjusted figure. These results suggest that in failing to standardize the units by which research output has been quantified, previous studies (such as the Pakes Griliches [1980] study which simply used raw patent counts) have significantly underestimated the mean lag between project inception and project completion.

IV. SUMMARY

These initial results on the agricultural research expenditure-research output relationship are quite encouraging. Although they convey relatively little information about the precise shape of the lag distribution, we have been able to obtain a significant relationship between lagged research expenditures and (weighted) publication output, even after controlling for unspecified state specific effects. Summary measures such as the 'mean' gestation lag and long run expenditure response were used to characterize this relationship. These measures were informative and plausible in the light of comparative studies in the private, non-agricultural research sector. The empirical implications of using raw versus quality adjusted publication output variables were also explored and found to be of significance.

For our data at least, the relationship between research expenditures and research output within states over time appeared quite tenuous. Short term fluctuations in research expenditures showed little systematic influence on research output. The on-average or longer run differences in research expenditures between the states does appear to influence research performance in a fairly systematic manner.

The summary measures do suggest that a significant lag between research inputs and outputs exist although Hall et al. (1984), summarizing extensive empirical work on the patent-R&D relationship for the private sector, highlight the substantial difficulties in trying to pin down this relationship. Key issues involve possible simultaneity between patent output and R&D expenditures, lag truncation biases due to the relationship between pre and in-sample R&D expenditures, and a lack of independence between unspecified state-specific effects and in-sample R&D expenditures. In the present study we tested formally for the failure of this independence assumption and got results which suggest that unspecified state effects were not correlated with in-sample research expenditures. Nevertheless, in both the PROPUB and PRONET models these unspecified state effects appeared to account for significant differences in the research performance of the SAES, even after controlling for differences in research spending.

Table (6) Estimates of the 'Mean' R&D Lag (in years).

	R&D Gestation	Application	Total
Rapoport[a]			
Chemicals	1.48	0.24	1.72
Machinery	2.09	0.31	2.40
Electronics	0.82	0.35	1.17
Wagner[a]			
Durables	1.15	1.47	2.62
Nondurables	1.14	1.03	2.17
Pakes-Griliches[a]			
All Manufacturing	1.16	n.a.	n.a.
Pardey[b]			
Agriculture			
(i) Quality Adjusted	3.36	n.a.	n.a.
(ii) Quality Unadjusted	2.83	n.a.	n.a.

(a) From Pakes and Schankerman (1978) calculated from data contained in Rapoport (1971) and Wagner (1968).

(b) Represents an average of the 'mean' lags from the specifications presented in Table (5) and an equivalent set of PROPUB regressions minus 6 months, an approximation of the average publication lag from project completion to publication.

FOOTNOTES

1. For notational simplicity the gross superscript, g, will be suppressed forthwith.

2. Evenson and Wright (1982) express similar sentiments.

3. They show that research output as proxied by publications significantly enhances the earnings of academic economists. See also Wright (1983) and Pakes and Nitzan (1983) for discussions on the economics of invention incentives in the private sector.

4. The goal is to estimate the total (i.e. population) number of publications per station, $\hat{P}_i = N_i \hat{p}_i$ where n_i and N_i represent the sample and population size for state i respectively and $\hat{p}_i$ is an estimate of the average number of publications per researcher in state i. However, the variance of $\hat{P}_i$ is estimated by $V(\hat{P}_i) = N_i^2 s_i^2/n_i [(N_i - n_i)/N_i]$ after applying the finite sample correction factor to the estimate of p_i. (Here $s_i^2 = \Sigma(x_{ij} - \bar{x}_i)^2/n_i - 1$, x_{ij} being the observed number of publications for researcher j in state i, and $\bar{x}_i = \hat{p}_i$, s_i^2 is assumed to be relatively constant across different i's.) Thus, in order to transmit homoskedastic sampling error to the error term in a regression of $\hat{P}_i$, we need to determine n_i by selecting a constant, k, defined as: $k = N_i^2/n_i(1 - n_i/N_i)$ such that $\Sigma n_i/\Sigma N_i = 0.20$. However, choice of k such that Σn_i approximately equals twenty percent of ΣN_i implied unrealistically small sample sizes ($n_i < 1$) for the smaller stations. A practical alternative was to (iteratively) choose k^* such that $k^* = 1/n_i(1 - n_i/N_i)$. If the underlying population variance was in fact constant across all stations, then this procedure would introduce heteroskedastic sampling error into a regression with total publications per station as the dependent variable. We will return to this issue in Section III.

5. There are alternative sources - the annual Cooperative State Research Service (CSRS) Funds for Research at State Agricultural Experiment Stations reports, and listings from the United States Department of Agriculture's Current Research Information System (CRIS) - but they were both deemed too incomplete and unreliable for our purposes. Moreover, it was not possible to develop a quality index for the CSRS listing, while the fixed length format of the CRIS records means that it under-reports the publication output of highly productive or large projects. Limiting the coverage to scientific publications also eliminates a potentially serious upward bias which may result from using a broader class of publications. For example, many station bulletins etc. simply 'repackage' the knowledge produced by the station (and already reported in scientific articles) for a non-scientific audience.

6. The pro-rated measure is empirically more appealing a priori because it removes the implicit double counting which is likely when scaling the unadjusted per researcher figure to a station level figure. The population figure used here for scaling was inclusive of social science and agricultural engineering researchers.

7. The proportion of articles published by a sub-sample of 150 researchers, receiving either one or two citations, peaks in year t-2. This year also records the second lowest proportion of non-cited articles and the highest number of total citations.

8. See Pardey (1986) chapter 4 for more details.

9. See Pardey, Craig and Hallaway (1987) for more deflator details. Here non-capital expenditures were calculated as Total Expenses - .8 (Fees, Sales and Miscellaneous)-Equipment - (Land and Buildings).

10. For the moment we assume the time effect is being proxied by a simple linear trend variable. This is tested in Section III. Indirect least squares intercept estimates can be recovered by substitution.

11. With this approach we estimate, instead of the (N-1) μ's and (T-1)λ's of the covariance model, only two parameters for each effect, namely their mean and variance.

12. These averages are in line with those obtained for other studies. Shaw (1967) recorded a mean publication per year count of 1.68 based on the complete publication records of approximately 3,000 scientists in ARS-USDA through to January 1965. Limiting the count to peer reviewed articles (as does this study) Salisbury (1980, Table 9) obtained an annual publication output, averaged across all Illinois SAES scientists for the 1948-78 period, of 1.08.

13. A Box-Cox procedure (where all variables, z, are transformed $(z^{\lambda} - 1)/\lambda$) which artificially nested the linear and log P models was tried, with $\lambda = 0.270$ maximizing the value of the resultant likelihood function. We have no rationale for accepting this value other than its statistical difference from $\lambda = 1$ or 0. However, application of a Glejser (1969) test suggested that the linear P model suffered from heteroskedasticity in the error term ω, possibly induced by our sampling procedure as discussed earlier. Logging the dependent variable appears to remove the problem.

14. These figures dropped even further to range from 0.06 to 0.47 after partialling out time.

15. Their sample included 38 firms in the chemical drugs and medicine industry, 13 in machinery, 10 in office computing and accounting machinery, 8 in electronics components and communication, 11 in professional and scientific instruments and 41 in other manufacturing.

16. From results presented in Hausman, we can write $V(\hat{q}) = V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})$ and form the specification test statistic $m = \hat{q}'M(\hat{q})^{-1}\hat{q} \stackrel{A}{\sim} \chi^2_K$ where $\hat{M}(\hat{q}) = (X'Q_1X)^{-1} - (X'\hat{\Omega}^{-1}X)^{-1}$, $K =$ the number of unknown parameters in β when no misspecification is present, Q_1 the matrix such that $Q_1X = X - X_{i.}$, and $\hat{\Omega}$ the estimated covariance matrix from a random state specific effects specification of equation (5). (In fact $M(\hat{q}) = 1/T[V(\hat{q})]$.) The equivalent test in a regression format is to perform OLS on the augmented equation $Y_{GLS} = \beta X_{GLS} + \alpha X_{FE} + v$ where Y_{GLS}, X_{GLS}, and X_{FE} are the appropriately transformed variables, and test whether $\alpha = 0$ by comparing the estimated variance from the random effects specification to the estimated variance from the augmented specification.

17. In an extended discussion on the veracity of within estimates Mairesse (1978) cites this collinearity problem, along with the fact that within deviations are not generally large and may be severely affected by measurement error, as good reasons for resorting to the between estimates.

REFERENCES

Burke, J., Connections, Little, Brown and Co., Boston, 1978.

Cline, P.L., "Sources of Productivity Change in United States Agriculture," Unpublished Ph.D. Thesis, Oklahoma State University, Stillwater, 1975.

Davis, P. and G.F. Papenek, "Faculty Ratings of Major Economics Departments by Citations," American Economic Review, Vol. 74, 1984, pp. 225-230.

Davis, J.S., "Stability of the Research Production Function Coefficient for U.S. Agriculture," Unpublished Ph.D. Thesis, University of Minnesota, St. Paul, Minnesota, 1979.

Evenson, R.E., "The Contribution of Agricultural Research and Extension to Agricultural Production," Unpublished Ph.D. Thesis, University of Chicago, Illinois, 1968.

Evenson, R.E. and B. Wright, "An Evaluation of Methods for Examining the Quality of Agricultural Research," in An Assessment of the United States Food and Agricultural Research System, Office of Technology Assessment, Vol. II, Part C, April 1982.

Fuller, W.A. and G.E. Battese, "Estimation of Linear Models with Crossed-Error Structure," Journal of Econometrics, Vol. 2, 1974, pp. 67-78.

Glejser, H., "A New Test for Heteroscedasticity," Journal of the American Statistical Association, Vol. 64, 1969, pp. 316-323.

Griliches, Z., "Measuring Inputs in Agriculture: A Critical Survey," Journal of Farm Economics, Vol. 42, 1960, pp. 1411-1433.

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

Guttman, J., "Interest Groups and the Demand for Agricultural Research," Journal of Political Economy, Vol. 86, 1978, pp. 467-484.

Hadwiger, D.F., The Politics of Agricultural Research, University of Nebraska Press, Lincoln, 1982.

Hall, B.H., Z. Griliches and J.A. Hausman, "Patents and R and D: Is There a Lag?," Working Paper No. 1454, National Bureau of Economic Research, September 1984.

Hansen, W.L., B.A. Weisbrod and R.P. Strauss, "Modeling the Earnings and Research Productivity of Academic Economists," Journal of Political Economy, Vol. 86, 1978, pp. 729-741.

Hatanaka, M. and T.D. Wallace, "Multicollinearity and the Estimation of Low-Order Moments in Stable Lag Distributions," in J. Kmenta and J. Ramsay, eds., Evaluation of Econometric Models, Academic Press, New York, 1980.

Hausman, J.A., "Specification Tests in Econometrics," Econometrica, Vol. 46, 1978, pp. 1251-1272.

Hausman, J.A., B. Hall and Z. Griliches, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," Technical Paper No. 17, National Bureau of Economics Research, August 1981.

Hsiao, C. Analysis of Panel Data. Cambridge University Press, Cambridge, 1986.

Institute for Scientific Information, Science Citation Index, Philadelphia, yearly volumes, 1970-1975.

Kamien, M.I. and N.L. Schwartz, Market Structure and Innovation, Cambridge University Press, Cambridge, 1982.

Latimer, R.G., "Some Economic Aspects of Agricultural Research and Education in the United States," Unpublished Ph.D. Thesis, Purdue University, West Lafayette, Indiana, 1964.

Mairesse, J., "New Estimates of Embodied and Disembodied Technical Progress," Annales de l'INSEE, No.30-31, 1978, pp.681-720.

McDowell, J.M., "Obsolescence of Knowledge and Career Publication Profiles: Some Evidence of Differences Among Fields in Costs of Interrupted Careers," American Economic Review, Vol. 72, 1982, pp. 752-768.

Mundlak, Y., "On the Pooling of Time Series and Cross Section Data," Econometrica, Vol. 46, 1978, pp. 69-85.

Mundlak, Y. and I. Hoch, "Consequence of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," Econometrica, Vol. 33, 1965, pp. 814-828.

Pakes, A.S., "Economic Incentives in the Production and Transmission of Knowledge: An Empirical Analysis," Unpublished Ph.D. Thesis, Harvard University, Cambridge, 1978.

Pakes, A. and M Schankerman, "The Rate of Obsolescence of Knowledge, Research Gestation Lags, and the Private Rate of Return to Research Resources," Discussion Paper No. 659, Harvard Institute of Economic Research, Harvard University, Cambridge, October 1978.

Pakes, A. and Z. Griliches, "Patents and R&D at the Firm Level: A First Report," Economic Letters, Vol. 5, 1980, pp. 377-381.

Pakes, A. and S. Nitzan, "Optimum Contracts for Research Personnel, Research Employment, and the Establishment of "Rival" Enterprises," Journal of Labor Economics, Vol. 1, 1983, pp. 345-365.

Pardey, P.G., "Public Sector Production of Agricultural Knowledge," Unpublished Ph.D. Thesis, University of Minnesota, St. Paul, Minnesota, 1986.

Pardey, P.G., B. Craig and M. Hallaway, "U.S. Public Sector Agricultural Research Deflators: 1890-1985," Department of Agricultural and Applied Economics Staff Paper P87-25. St. Paul: University of Minnesota, August 1987.

Rapoport, J., "The Anatomy of the Product-Innovation Process: Cost and Time", in E. Mansfield et.al., Research and Innovation in the Modern Corporation, Norton, New York, 1971, pp. 110-135.

Rose-Akerman, S. and R.E. Evenson, "The Political Economy of Agricultural Research and Extension: Grants, Votes and Reapportionment," American Journal of Agricultural Economics, Vol. 67, No. 1, February 1985, pp. 1-14.

Rosenberg, N., Perspectives on Technology, Cambridge University Press, Cambridge, 1976.

Salisbury, G.W., "Research Productivity of the State Agricultural Experiment Station System: Measured by Scientific Publication Output," Agricultural Experiment Station Bulletin No. 762, University of Illinois, Urbana-Champaign, July 1980.

Shaw, B.T., "The Use of Quality and Quantity of Publication as Criteria for Evaluating Scientists," Miscellaneous Publication No. 1041, United States Department of Agriculture, Washington D.C., 1967.

U.S. Department of Agriculture, Agricultural Research Service, Report on the State Agricultural Experiment Stations, Washington D.C., yearly issues, 1955-1959.

U.S. Department of Agriculture, Cooperative State Research Service, Professional Workers in State Agricultural Experiment Stations and Other Cooperating Institutions, Agricultural Handbook No. 305, Washington, D.C., various years.

U.S. Department of Agriculture, Cooperative State Research Service, Funds for Research at State Agricultural Experiment Stations and Other Cooperative State Institutions, Washington, D.C., yearly issues, 1960-1975.

U.S. Department of Agriculture, Office of Experiment Stations, Report on the State Agricultural Experiment Stations, Washington, D.C., yearly issues, 1938-1954.

Wagner, L.U., "Problems in Estimating Research and Development Investment and Stock," American Statistical Association, Proceedings of the Business and Economics Statistical Section, 1968, pp. 189-198.

Wright, B.D., "The Economics of Invention Incentives: Patents, Prizes, and Research Contracts," American Economic Review, Vol. 83, 1983, pp. 691-707.

Yotopoulos, P.A., "From Stock to Flow Capital Inputs for Agricultural Production Functions: A Microanalytic Approach," Journal of Farm Economics, Vol. 49, 1967, pp. 476-491.

Appendix I

Table (I.1) Publication Statistics - Type of Publication, Period Averages.[a]

Publication Type[b]	Publications per Year 1970-73 average	1974-75 average	1970-75 average
Article	1807 (64.9)	1709 (66.6)	1758 (65.7)
Biographical	0	1	1
Correction	7 (0.3)	7 (0.3)	7 (0.3)
Discussion	3 (0.1)	2 (0.1)	3 (0.1)
Editorial	12 (0.4)	9 (0.4)	11 (0.4)
Letter	18 (0.6)	14 (0.5)	16 (0.6)
Meeting	679 (24.4)	600 (23.4)	640 (23.9)
Note	244 (8.8)	206 (8.0)	225 (8.4)
Review	14 (0.5)	21 (0.8)	18 (0.7)
TOTAL	2783	2567	2675

(a) Figures in parentheses are percentages.

(b) The publication type categories are self explanatory. They are the same categories as used by ISI. (See ISI, 1982, p. 14).

Table (I.2) Publication Statistics - Coauthor Frequency per Publication, Period Averages.[a]

	Publications per Year		
Number of Coauthors	1970-73 average	1974-75 average	1970-75 average
0	386 (13.9)	275 (10.7)	331 (12.4)
1	1036 (37.2)	910 (35.4)	973 (36.4)
2	759 (27.3)	749 (29.2)	754 (28.2)
3	400 (14.4)	393 (15.3)	397 (14.8)
4	140 (5.0)	156 (6.1)	148 (5.5)
5	44 (1.6)	52 (2.0)	48 (1.8)
6	20 (0.7)	35 (1.4)	28 (1.0)

(a) Figures in parentheses are percentages.

Table (I.3) Simple Correlation Matrix for Various Agricultural Research Output and Quality Indicators. [a]

1.	Pub	1.000							
2	Propub	0.947	1.000						
3	Totcit	0.370	0.349	1.000					
4	Netcit	0.374	0.355	0.984	1.000				
5	Pub x Tot	0.705	0.662	0.849	0.846	1.000			
6	Pub x Net	0.689	0.647	0.847	0.860	0.994	1.000		
7	Pro x Tot	0.697	0.696	0.854	0.853	0.989	0.985	1.000	
8	Pro x Net	0.679	0.678	0.851	0.867	0.982	0.989	0.994	1.000
		1	2	3	4	5	6	7	8

(a) All variables are measured on an average researcher per state basis.

Table (I.4) Regression of Raw Publication Output (PROPUB) on Agricultural Research Expenditure[a].

Variable	OLS	WITHIN	BETWEEN	EGLS[b]
R_0	-0.1718 (0.4682)[c]	0.1103 (0.2527)	-2.8121 (6.4844)	0.0345 (0.2770)
R_{-1}	-0.4966 (0.6004)	0.3850 (0.2821)	1.1064 (13.038)	0.5597 (0.3143)
R_{-2}	-0.1319 (0.5921)	-0.4261 (0.2823)	-1.3339 (9.7122)	-0.1623 (0.3115)
R_{-3}	0.5089 (0.5851)	-0.1050 (0.2768)	15.544 (8.9129)	0.2178 (0.3039)
R_{-4}	0.2160 (0.5937)	0.0125 (0.2804)	-15.974 (11.868)	0.3138 (0.3079)
R_{-5}	0.0058 (0.6053)	-0.0784 (0.2784)	8.5244 (11.273)	0.0429 (0.3131)
R_{-6}	-0.1290 (0.6127)	-0.2141 (0.2878)	-10.703 (10.738)	0.0321 (0.3211)
R_{-7}	0.3713 (0.4779)	-0.3965 (0.2745)	6.8120 (5.6910)	0.1114 (0.2829)
Trend	-0.0432	10.066 (13.531)	-	-0.0397 (0.0098)
ΣR_{-i}	1.1659	-0.9329	1.1638	1.0900
R^2	.7480	.0466	.8163	.3529
NT	288	288	48	288
'Mean' Lag	3.23	3.50	4.28	2.69
χ^2				16.621

(a) All variables measured in natural logs.

(b) See note (b) Table (4). Here $\hat{\sigma}_1^2 = 1.2102$, $\hat{\sigma}_\omega^2 = 0.0602$ and $\hat{\sigma}_\mu^2 = 0.1917$.

(c) Standard errors in parentheses.