The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Econometric Assessment of Research Programs: A Bayesian Approach

Lin Qin and Steven Buccola

Oregon State University

Selected Paper prepared for presentation at the Agricultural and Applied Economics Association's 2012 Annual Meeting, Seattle, Washington, August 12 – 14, 2012

Lin Qin (qinl@onid.orst.edu) is a Graduate Research Assistant and Steven Buccola (sbuccola@oregonstate.edu) a professor in the Department of Agricultural and Resource Economics at Oregon State University.

This research was funded in part by the AquaFish CRSP under USAID CA/LWA No. EPP-A-00-06-00012-00 and by US and Host-Country partners. AquaFish CRSP accession number 1400. Opinions expressed herein are the authors' and do not necessarily reflect the views of the AquaFish CRSP or the US Agency for International Development.

Copyright 2012 by Lin Qin and Steven Buccola. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided this copyright notice appears on all such copies.

Econometric Assessment of Research Programs: A Bayesian Approach

Abstract

Effective research-project assessment typically is impeded by project variety. In particular, bibliometric approaches to science assessment tend to offer little information about the content of the projects examined. We introduce here a new approach – based on Bayesian theory – of econometrically evaluating the factors affecting scientific discovery, and use the method to assess a biological research program comprised of numerous heterogeneous projects. Our knowledge metric not only flexibly accommodates project variety but accounts for information in "failed" as well as "successful" studies. Using a mean-absolute-deviation utility functional form to measure new scientific knowledge, we decompose knowledge gain into a mean-surprise and statistical-accuracy effect. The two effects are econometrically examined independently, and then combined into the net knowledge production function. Research FTE and distance to study site have statistically significant but moderate effects on the amount by which research shifts the prediction of scientific outcome. However, scientist education powerfully improves the research study's predictive accuracy or precision, a one-percent boost in the average investigator's formal schooling improving precision by 4.3 percent. Largely on the basis of that precision effect, increasing returns to research project scale are evident.

Keywords: Bayes Rule, knowledge production, project assessment, science assessment

Econometric Assessment of Research Programs: A Bayesian Approach

Research administrators are continually asked to assess not only the significance of research project outcomes but their costs, which largely are determined by the effectiveness with which projects are designed and managed. The great volume and variety of projects under an administrator's purview, however, typically frustrate such assessment. The volume of projects discourages a study-by-study analysis of their findings. The variety of projects complicates the linkages an administrator might wish to establish between the knowledge gained in a project and the ways in which resources have been expended on it.

The most common way to try to circumvent such volume and variety problems is to regress studies' publication or patent counts against study aspects, such as study expenditure and topic area, that might influence those counts (Pardey 1989, Jaffe 1989, Hall, Jaffe, and Trajtenberg 2000). However, this bibliometric approach is fraught with difficulties. Publication, patent, and citation rates themselves contain little clue about the associated research discoveries themselves. They substantially lag the study's completion and – still more – the expenditures of its research effort. They fail to include many media that researchers employ to communicate their findings. And they are ill-suited to reflect the *amount* of knowledge a given study represents. Perhaps more importantly, publication, citation, or patent-rate datasets usually offer little corresponding information about study research designs, management practices, personnel or capital resources, or environmental conditions. Yet it is in the correspondences between such research inputs and outputs that the administrator is precisely interested.

To provide an improved vehicle for research program assessment, we introduce here a new approach – based on Bayesian theory – of econometrically evaluating the factors affecting scientific discovery. We apply the method to the assessment of a biological research program consisting of numerous and – the better for testing the approach's robustness – rather heterogeneous studies.

A brief description of that program provides context for the challenges facing any rigorous yet feasible research evaluation method. It consists of twenty-four aquacultural research investigations conducted under U.S. government sponsorship between 2007 and 2009 in 11 countries. The investigations are highly heterogeneous. Objectives range from examining fish production efficiency, to water quality, human health, species development, and marketing. Methods include both controlled experiments and statistical surveys. Experiment-based studies involve a variety of not only research treatments but of outcomes or findings in a given treatment. A feed-formulation treatment, for example, can generate data on survival rate, feed conversion, final body weight, and flesh quality. Statistical studies may inquire about an exporter's preferred fish length, species, and quality, or the water quality in an estuary. The investigations were conducted in a variety of technological and cultural settings across the globe. The knowledge metric employed in assessing these projects must be flexible enough to accommodate such variety, yet permit pooling into a single econometric model of the associations between knowledge-gained and resource-expended. In particular, it must overcome the data and conceptual problems of distinguishing program influences from other factors affecting the fish-farm or training setting. Much of our effort therefore was devoted to dealing with, and taking advantage of, cross-study heterogeneity.

Bayesian Measure of Scientific Knowledge

A knowledge metric accounting adequately for the studies' discoveries must satisfy at least three requirements. It must be: (a) comparable across the studies investigated; (b) *ex ante* in nature; and (c) reflective of the knowledge provided by both research "failures" and "successes." The first or cross-study comparability criterion can be achieved by expressing the metric in, for example, outcome percentage changes rather than level changes. Criterion (b) is essential for the metric's usefulness in research planning as well as evaluation, and requires we cast the metric in terms of the information the study is *expected* to generate. Criterion (c) is important because some studies fail in the sense that the hypothesized improvement – a better water purification scheme or feed ration, say – does not materialize. The failure does not, however, imply that expenditures have been wasted: the disappointment was valuable in pointing to more fruitful research directions (CGIAR Science Council 2009). In other words, the study was a success insofar as it updated our knowledge of the probabilities of management outcomes, expressed as a shifting or narrowing of the probability distribution of those outcomes in the presence of alternative treatments. Our research discovery assessment will consist in comparing these outcome distribution changes with the investigation inputs – such as expenditures, human capital, and effort – that have made the changes possible.

Production or Distance Function Approach

To compare outputs with inputs in this way, we regard each evaluated study as a production unit that uses inputs like money and personnel to produce discoveries (Buccola, Ervin, and Yang 2009; Xia and Buccola 2005). Every experiment-type study examines several alternative treatments, and every survey-type study examines several survey respondent sub-groups. Furthermore, a given experimental treatment gives rise to a multitude of findings or outcomes such as weight gain and feed/gain ratio, and a given survey involves a variety of questions. For each of these two reasons, we are able to examine a large number of research input-output combinations, allowing strong statistical inferences about research program management success.

A production function approach to research discovery assessment may be represented as

(1) $$K = f(\mathbf{X}) = f(X_1, X_2, \ldots, X_I)$$

where K is a measure of the knowledge discovered in the investigation and $\mathbf{X} = (X_1, X_2, \ldots, X_I)$ is the vector of the study's I inputs. Equation (1) becomes a distance function upon appropriate restrictions. We focus on the manner in which the before-study ("prior") and after-study ("posterior") probabilities of outcomes K – like mortality rate – are obtained. Well-developed methods are available for eliciting investigators' prior probabilities of their discoveries (Stael von Holstein 1970). The corresponding posterior probabilities are obtained from the investigation's statistical results, in the form either of analysis-of-variance results or statistical means and variances. Bayesian methods are employed to update such probabilities as the investigation proceeds (Schimmelpfennig and Norton 2003). If Y is the percentage improvement in a study outcome such as pond quality, and Z the experimental performance (sample information) of the pond-cleaning technology the researchers are studying, the likelihood that outcome Z will occur depends on study inputs $\mathbf{X}$ and on random events ε; that is, $Z = Z(\mathbf{X}, \varepsilon)$. Bayes Theorem says the probability the investigator assigns to a particular pond-quality improvement is, once the experiments have been completed,

(2) $$p[Y \mid Z(\mathbf{X},\varepsilon)] \propto p(Y) \cdot p[Z(\mathbf{X},\varepsilon) \mid Y]$$

where $p(Y)$ is the scientist's prior estimate of the chances that pond-quality Y will occur.

Equation (2) provides the very research knowledge measure K we seek. To see this, suppose an oyster producer is faced – in the presence of her present oyster management practices $\mathbf{X}$ – with a decision d about how many oysters to promise to deliver next month at quality grade A. If she later delivers them at lower than the promised grade, her quality reputation will suffer;

if at higher than the promised grade, she will not be prepared to market them at the proportionately higher price. Her utility, that is, rises with her accuracy in predicting quality grade. Oyster management research improves such prediction accuracy, so that her marketing decisions can be based on posterior probability $p(Y \mid Z)$ rather than prior probability $p(Y)$. That gain is reflected in the *value of sample (research-produced information* **Z** (Winkler 1972), namely the very knowledge measure we seek:

$$(3) \qquad K = VSI(d, Z) = E''\{U[d'' \mid Z(\mathbf{X}, \varepsilon)]\} - E''\{U[d' \mid Z(\mathbf{X}, \varepsilon)]\} = f(\mathbf{X}, \varepsilon)$$

Here, d' is the optimal decision in the presence of prior information only, d'' is the optimal decision in the presence of both the sample **Z** and prior information, ε represents random unobservable study inputs, $\mathbf{Z}(\mathbf{X})$ indicates Z's dependence on research inputs **X**, and *f* is the knowledge production function. Equation (3) shows that the value of sample information created by the research study is the disutility the fish farmer suffers if deprived of the research study.

Functional Form

A number of functional forms are available for specifying fish-farmer utility *U* in (3). We adopt here the mean absolute deviation (MAD) form $U(d, Y) = -|Y - d|$ in which utility is proportionate to the absolute difference between a random outcome and its prediction (Robert 2001). In the present context it carries two assumptions: (a) the farmer loses as much utility when fish quality turns out to be a given amount *below* his quality prediction as it does when it turns out to be a given amount *above*; and (b) that loss is proportionate to the difference between the predicted and actual quality. The realism of these assumptions might be greater for some outcomes than for others.

With the MAD functional form, knowledge equation (3) becomes

(4) $$K = \left(\sum_i |Y_i - M_{PR}| \right)_{PO} - \left(\sum_i |Y_i - M_{PO}| \right)_{PO} = f(\mathbf{X}, \varepsilon)$$

where M_{PR} is the mean of the prior probability distribution $p(Y)$ of *Y*, and M_{PO} is the mean of the posterior distribution $p[Z(\mathbf{X},\varepsilon) | Y]$. The first middle term in equation (5) is the mean absolute deviation of a sample fish-quality observation from the pre-research quality prediction, namely $d' = M_{PR}$. The second middle term is the mean absolute deviation of an observation from the post-research prediction $d'' = M_{PO}$ (*Y*'s sample mean). In the present analysis, an AquaFish investigation's knowledge output is modeled as the difference between these two farmer risks, each computed from the post-research probability distribution of quality outcomes.

Numerical decomposition shows (4) can, with a high degree ($R^2 \approx 0.97$) of accuracy, be expressed as

(5) $$\ln K = A + a_1 \ln\left(|M_{PR} - M_{PO}|\right) + a_2 \ln\left(STD_{PO} \right)$$

where A, a_1, and a_2 are estimated constants and STD_{PO} is the standard deviation of the post-research fish quality distribution. That is, a research study's information contribution depends separately on: (i) the absolute difference between the pre- and post-research estimate of expected output (the study-produced absolute shift in the output expectation and hence the study's *mean surprise*), and (ii) the post-research sample estimate of the standard deviation of outcomes (the study's predictive success and hence *accuracy*). This decomposition allows us much flexibility in examining the impacts of study inputs **X** on knowledge output *K*, for it permits us to distinguish between how those inputs affect the shift in the outcome forecast as well as the accuracy of that forecast. Mean surprise and accuracy are, that is, separate aspects of a study's forecast effectiveness. Mean surprise tells us how close the average dart is to the bulls eye; accuracy tells us how close the darts are to one another.

Expressions $\ln\left(|M_{PR} - M_{PO}|\right)$ and $\ln\left(STD_{PO}\right)$ on the right-hand-side of (5) were estimated – for each research treatment, outcome, and survey question – by taking 50 random draws of each the two parenthesized expressions in the middle of (4) and using them to compute M_{PR}, M_{PO}, and STD_{PO}. Because outcomes Y_i in the studies we examined are expressed in a variety of units (kilograms per hectare, micrograms of oxygen per liter of water, and so forth), we divide each by the mean of the 50 outcomes drawn. Each outcome thus is equivalently expressed as its percent deviation from its own posterior mean. Thus also, this normalization reduces posterior mean M_{PO} to 1.0 in each experimental treatment, outcome, and survey question, so that M_{PR} is expressed as a proportion of M_{PO}.

Estimation Methods

Data required for constructing knowledge output measures *K* and input measures **X** were obtained from the key study scientists. The scientist in an experiment-type investigation is asked – for each treatment and each outcome of that treatment – to state:

(i) her prior or pre-study probability p_L, p_M, p_H that each of three selected (*L*, *M*, *H*) *experimental outcome* levels would be attained; and (ii) the corresponding posterior probabilities as represented by the outcome's mean and standard deviation in the ANOVA results. The scientist in a survey-type investigation is asked – for each major survey question – to state: (i) his prior probability p_L, p_M, p_H that he would obtain each of three selected (*L*, *M*, *H*) *survey answer* levels; and (ii) the corresponding posterior probabilities as represented by the mean and standard deviation of the respondent's answers to that question.

Knowledge Inputs

Study inputs included researchers' salaries and wages, travel, research materials such as feeds and medicines, training materials, student-workers' tuitions, and publication expenses. At a more aggregate level, information also was available on materials and equipment and administrative overhead. Human capital – the knowledge and skills embedded in the investigators themselves – is to some degree reflected in researchers' salaries. But such reflections are imperfect and usefully supplemented by, for example, information about the research team's academic rank and experience. More expenditure or human capital would provide greater scope for solving the research problem and hence presumably boost knowledge production *K*.

Experimental approaches may be more or less difficult to plan and manage than survey studies are, so our expectations of experimental controls' impacts on knowledge production are somewhat ambiguous also. One might, however, at least expect controlled experiments to bring lower sample variances than survey studies do, since experimental controls are designed for very purpose of reducing random noise. A study's outcome dimensions likely differ among one another in understandability or accessibility – fish weight-gain perhaps being more difficult to measure than water oxygen level. Similarly, an investigation's topic area category may influence research difficulty insofar as some topics may be less understood and more expensive to address than are others. But we do not have strong prior expectations for such outcome-wise or topic-area-wise effects.

Finally, public infrastructure can influence knowledge output in many ways. Distance and the difficulty of travel to study site consume resources that otherwise could be expended on knowledge-generation at the site. Climate, culture, or other geographic factors can have their own impacts on research success, for example when traditions among a surveyed group of scientists influence the types of information they are willing to reveal. Data collection ran from October 2010 to December 2011.

Model Specifications

Indexing knowledge-production model (1) in terms of the types of research inputs discussed above gives

(6) $$K_{ijk} = f(\mathbf{X}) = f(\mathbf{E}_i, \mathbf{H}_i, \mathbf{T}_{ijk}, \mathbf{I}_i)$$

in which

$\mathbf{E}_i$ is the vector of expenditures on the i^{th} investigation, in dollars per biennium;

$\mathbf{H}_i$ is the vector of human capital variables in the i^{th} investigation;

$\mathbf{T}_{ijk}$ is the vector of research problem types in the j^{th} treatment and k^{th} outcome of the i^{th} investigation;

$\mathbf{I}_i$ is the vector of public infrastructure variables in the i^{th} investigation.

The correspondingly indexed form of equation (5) is

(7) $$\ln K_{ijk} = A + a_1 \ln\left(|M_{PR} - M_{PO}|\right)_{ijk} + a_2 \ln\left(STD_{PO}\right)_{ijk}$$

in which

(8) $$\ln\left(|M_{PR} - M_{PO}|\right)_{ijk} = f_1(\mathbf{X}_{ijk}, e_{1,ijk})$$

(9) $$\ln\left(STD\right)_{ijk} = f_2(\mathbf{X}_{ijk}, e_{2,ijk})$$

We therefore are able to estimate the latter two equations separately, and so can observe how research inputs affect research distribution shift or *mean surprise* (8) separately from how they affect research *accuracy* (9). We then combine them, *via* equation (5) and its α_1 and α_2

parameter estimates, to conclude how research inputs affect total knowledge output – the expected value of the research program's sample information.

Results

Sample size was 415. In the typical study, research treatment, and outcome, the researcher's prior outcome expectation was 0.224 (22.4%) higher or lower than the mean outcome in the subsequent experiment or survey. That is, the researcher's expectations tended to be 22% "off" what eventually happened, creating a 22% mean surprise. The associated standard deviation of 0.278 is higher or lower than the mean, implying substantial variation in how far a researcher's prior expectations missed the eventual mark. The distribution of these absolute surprises was skewed strongly to the right: mean surprises tended to cluster just above zero, the successively larger ones being continuously less frequent.

The shape of the sample distribution of posterior standard deviations – measuring the spread or inaccuracy of research outcomes around their experimental or survey means – was similar to the distribution of mean shifts or surprises. Average experimental or survey outcome was 0.38 (38 percent) above or below its own mean. The variation of this research accuracy across studies, treatments, and types of outcome was, at 0.653, nearly twice (1.718) as high as the sample-mean accuracy. And the distribution of these accuracies across studies and outcomes was again strongly positively skewed. They bunched just to the right of zero – where standard deviations are low – the successively larger ones (with higher standard deviations and hence more inaccurate) being continuously less frequent. Because our measure of knowledge-gained is a weighted sum of mean absolute surprise and variance, these coefficients of variation are quite adequate for examining the factors affecting scientific knowledge creation.

The average research team consisted of 13 individuals, the average investigator 33 years old with 17 years of education. Sixty-eight percent of research outcomes pertained to fish production (growth, feeding, and mortality), 18% to water quality, and 14% to fish marketing strategies. Average distance to the experiment or survey site was 843 kilometers. Two-thirds of scientists' transportation to research site involved largely auto or bus, and one-third largely walking. Sample variation of most of these research inputs is adequate for regression inference. For example, coefficients of variation of research outcome dimensions are mostly above one. Few of these variables are pairwise-correlated to the degree that would create inference problems.

Factors Affecting Output: Mean Surprise

Estimates of research surprise equation (8), using the 2007 – 2009 sample of 415 observations, are shown in table 1. Twenty-eight outlier observations were dropped from the original 443 observations on account of their excessive influence on regression estimates. Column (2) gives the coefficient estimates and column (4) the *t*-statistics. Because both research surprise and the continuous research inputs are expressed in logs, coefficients of these inputs are "output elasticities," the percent increase in surprise generated by a one-percent rise in the research input. The coefficient of the analytical-approach variable in table 1 instead is the percent difference in mean surprise between the indicated group and the base group.

In the interests of space, we report here only the results for the team-size, team-education, distance-to-study-site, and analytical-approach variables. Expanding a study's size by increasing the number of its full-time-equivalent investigators statistically significantly shifts the location of the research outcome's probability distribution. In particular, a one-percent *FTE* boost lifts mean surprise by 0.206 percent. The bigger the team, that is, the greater the mean surprise expected.

On the other hand, the team's aggregate education does not significantly affect mean surprise at all (t = - 0.264). Perhaps it is the team leader's rather than junior members' skills that materially influence a distribution shift. In ways, that is, that outcome accuracy does not, the gap between the prior and posterior expectation of a study outcome represents a fundamental shift in scientific understanding of the issue at hand. That gap depends on insights into how best to frame the research question and controls, insights that only the research leader ordinarily can provide. Finally, average distance to research site does affect the mean surprise a study generates: a one-percent greater travel distance reduces surprise by 0.107 percent, all else constant. This effect, too, is strongly significant (t = - 4.38). And, although modest in magnitude, it is consistent with our expectation that travel time cuts down on the time and energy available to the leader and team for creativity and innovation.

Studies consisting of controlled experiments bring, all else equal, an average 105 percent less mean surprise than do studies consisting of statistical surveys. This result is rather expected. The possibility of analyzing a problem with experimental controls is normally encountered where the scientist is relatively familiar with the problem's stochastic environment. Such familiarity in turn implies one would not normally expect the research to greatly change the investigator's expectations of a study outcome. Survey methods, in contrast, usually are more exploratory, ones investigators use when the environment is relatively unknown. Surveyors' expectations therefore would tend to be weak, a situation conducive to marked changes in them as the study proceeds.

Research mean-surprise model R-square is 0.26. We are able to explain about one-quarter of the variation across treatments, outcomes, and studies in research mean surprise. The modesty of this proportion is to be expected. The investigator insights needed to fundamentally shift our understanding of a research problem are relatively ineffable and not easily explained with measurable inputs. That is not, as we shall see next, true for the research accuracy dimension.

Factors Affecting Output: Statistical Accuracy

Consider now (table 2) the corresponding estimates of research accuracy equation (9), showing research inputs' effects on the standard deviations (imprecisions) of the experimental or survey-question outcomes. Factors with negative effects in table 2 are those contributing positively to precision and thus, by way of equation (7), to scientific knowledge.

The scale of the research investigation – represented by team *FTE* – has an unexpected precision effect, although its weak statistical confidence ($t = 1.32$) serves to temper it. A one percent *FTE* rise brings 0.11% lower predictive accuracy in an investigation's experiments and surveys. Negative output elasticities of this sort typically are explained in the economics literature as a crowding effect. Depending on current team size and team-leader management skill, boosting team size can complicate the leader's management burden more than it frees his time for strategic thinking. Why this phenomenon holds here for the production of outcome accuracy and not (as above) for the production of mean surprise is unclear.

The research team's education has an especially powerful impact – indeed the only one in this study with an elasticity exceeding unity – on the accuracy of the team's research findings. A one-percent rise in the team-members' average formal education reduces outcome standard deviation by 4.26%. The intuitive nature of this result is clear: better-educated teams are better able to maintain the experimental controls which restrict research outcomes' random variations. At the ground level, these controls usually are in the hands of the team's most junior members, whose educational preparation therefore becomes crucial for experimental accuracy. In earlier runs, collaborator (for example cooperating fisher) education had no detectable influence on accuracy, perhaps because it is the collaborator's experience rather than education that better predicts her usefulness.

The influence on research accuracy of mean distance to study site in table 2 is consistent with distance's influence on research mean surprise: greater travel distance leads to greater sampling variance and hence to lower research accuracy. Physical distance presumably inhibits site visits and hence study monitoring. On the other hand the effect is, along with modest statistical significance, rather small. A one-percent rise in travel distance impairs accuracy by about 0.03 percent.

The scientist's choice between an experimental and survey research design has a much more profound research-accuracy implication than distance has. Experimental studies in the present dataset provide, all other factors constant, 75% lower outcome variances and thus more accurate study results than do survey investigations. That effect has extraordinarily high statistical significance. It also is highly intuitive: as we have noted above, experimental controls are imposed for the very purpose of reducing study outcome variances below those achievable when only statistical controls are employed.

Research accuracy model R^2 is 0.44. We are able to explain about one-half the statistical-accuracy variation in our sample of scientific studies, treatments, and outcomes.

Implications for Knowledge Production

We now aggregate the table 1 and 2 estimates, by way of equation (7), to see how research inputs influence knowledge production itself. Results of our equation (7) regression on the sample data are:

$$\text{(10)} \quad \ln K \; = \; -1.188 \; + \; 1.76 \; \ln\left(|M_{PR} - M_{PO}|\right) \; - \; 0.71 \; \ln\left(STD_{PO}\right) \quad R^2 \; = \; 0.97$$
$$\qquad\qquad (-30.97) \qquad\qquad (122.08) \qquad\qquad\qquad\qquad (-50.79)$$

in which the left-hand term is the log of knowledge output, the second right-hand term the log of the mean difference between prior expectation sand mean study findings, and the final term the log of the standard deviation of those findings. Numbers in parentheses are *t*-statistics. Mean surprise is a positive element, and standard deviation (the negative of accuracy) a negative element, of knowledge. Given the MAD utility functional form (4) assumed, mean surprise is – at the margin – substantially more important to knowledge-creation than accuracy is. In particular, equation (10) shows mean surprise's knowledge weight is 1.76 / 0.71 = 2.5 times greater than accuracy's knowledge weight. Using the 1.76 and -0.71 weights together with tables 1 and 2, we compute in table 3 each research input's net knowledge effect. For purposes of the table 3 calculations, we have set to zero any input effect whose estimated absolute *t*-value in tables 1 and 2 is below unity.

By way of its effect on mean surprise alone, the impact on net knowledge of expanding research-team *FTE* in table 3 is (1.76) (0.206) = 0.363. That is, by virtue of its boost to mean surprise, expanding team research time by one percent enhances knowledge by 0.36%. By virtue of the expansion's impact on research findings' sample variance, it reduces knowledge by (-0.71) (0.112) = - 0.079%. Thus, a one-percent team expansion lifts new knowledge by a net (0.363 – 0.079) = 0.284%. That is, scientist effort has a positive but modest total effect on scientific knowledge despite its questionably positive effect
(table 2) on research inaccuracy.

Because team education's influence on mean surprise is highly nonsignificant (table 1), its impact on net knowledge is identical to its impact on statistical accuracy alone, weighted by accuracy's incremental weight (- 0.71) in knowledge creation. Thus, a one percent education improvement lifts knowledge creation in table 3 by (- 0.71) (- 4.263) = 3.026%. Finally, by virtue of its effect on mean surprise, reducing the physical distance from station to study site by one percent lifts knowledge-gained – that is, exacerbating the distance reduces knowledge-gained – by (1.76) (0.107) = 0.189%. By way of its impact on sample variance, distance reduction lifts knowledge-gained by (- 0.71) (- 0.033) = 0.023%. Its net knowledge boost is therefore (0.189 + 0.023) = 0.212%. Distance's mean-surprise and accuracy effects here work in the same direction: reducing travel distance helps create both more mean distribution shift and more research accuracy.

Returns to Scale

It is useful to view research inputs' scientific knowledge effects in terms of returns to scale. Because team-size, education, and travel-distance-reduction are continuous inputs in a research enterprise, we can use them to compute the returns to scale in the production of mean research surprise. In particular, the table 1 sum of the mean-surprise elasticities with respect to these three factors is (ignoring education's statistically highly nonsignificant effect), 0.206 + 0 +0.107 = 0.313. Returns-to-scale below unity imply productive efficiency declines rapidly as enterprise size is expanded. Thus, scaling the team's labor time and education upward and travel distance downward would, while bringing more average surprise, do so with rapidly declining input-efficiency. Such decline reflects, we reason, the poor replicability of the lead scientist's time and creative talent, which presumably are responsible for most of a research study's mean surprise.

In contrast, scale returns in the production of research accuracy are, from table 2 and again reversing signs to signify that a standard deviation decline is an accuracy improvement, -0.112 + 4.263 + 0.033 = 4.184. This result implies increasing returns to scale: augmenting team *FTE* and education and reducing mean travel distance in the same proportion would boost statistical accuracy by four times that proportion. At least in the range of study sizes examined here, there appears, that is, to be no limit to the increased statistical accuracy achievable by expanding study sizes, and with that expansion the treatment and survey replication that improves statistical fit.

Finally, weighting the mean-surprise elasticities by their 1.76 knowledge weight and accuracy elasticities by their – 0.71 weight gives a net return-to-scale of 0.284 + 3.026 + 0.212 = 3.522. Returns-to-scale therefore appear to be strongly increasing, implying increasing productive efficiency as investigation size grows. Such high scale returns are remarkable because the research outcome dimensions and topic-area categories for which we control in the present analysis serve to restrict what the investigator examines, imposing a constraint and hence limiting the possibilities for efficient scaling-up of research effort. However, the increasing returns are due solely to research inputs' very strong accuracy effects, compensating heavily for their rather moderate impacts on scientific mean surprise.

References

Buccola, S.T., D. Ervin, and H. Yang. 2009. “Research Choice and Finance in University Bioscience.” *Southern Economic Journal* 75: 1238 - 1255

CGIAR Science Council. 2009. *Defining and Refining Good Practice in Ex-Post Impact Assessment – Synthesis Report.* CGIAR Science Council Secretariat: Rome, Italy.

Hall, B., A. Jaffe, and M. Trajtenberg. 2000. “Market Value and Patent Citations: A First Look.” NBER Working Paper No. 7741.

Jaffe, A. 1989. “Real Effects of Academic Research.” *American Economic Review* 79: 984-99.

Pardey, P.G. 1989. “The Agricultural Knowledge Production Function: An Empirical Look.” *Review of Economics and Statistics* 71:453-61.

Robert, C.P. 2001. *The Bayesian Choice, 2nd Ed.* New York: Springer-Verlag.

Schimmelpfennig, D. E. and G.W. Norton. 2003. “What is the Value of Agricultural Economics Research?” *American Journal of Agricultural Economics*. 85: 81 – 94.

Stael von Holstein, C.A. 1970. *Assessment and Evaluation of Subjective Probability Distributions.* The Economics Research Institute, Stockholm School of Economics

Winkler, R.L. 1972. *An Introduction to Bayesian Inference and Decision*. New York: Holt, Rinehart, Winston.

Xia, Y., and S.T. Buccola. 2005. “University Life Science Programs and Agricultural Biotechnology.” *American Journal of Agricultural Economics* 87: 229 – 24

Table 1. Research-Input Effects on the Absolute Difference Between Prior and Posterior Mean Finding (Mean Surprise), Selected Aquacultural Research Studies, 2007 – 2009.

Research Input	**Estimate**	**Standard Error**	***t*-value**
Intercept	0.171	3.790	0.045
Continuous Variables			
Team FTE	0.206	0.097	2.115
Team Mean Education	-0.353	1.333	-0.264
Mean Distance to Study Site	-0.107	0.025	-4.380
Analytical Approach			
Experiment *vs* Survey (Base Group: Statistical Surveys)	-1.051	0.248	-4.244

Notes

Dependent variable: Absolute difference between prior expectation and posterior sample mean of experimental finding or survey response

Residual standard error: 1.091
Sample size: 415
Multiple R-square: 0.26

Table 2. Research-Input Effects on the Standard Deviation (Inaccuracy) of Research Findings, Selected Aquacultural Research Studies, 2007 – 2009.

Research Input	**Estimate**	**Standard Error**	***t*-value**
Intercept	11.543	3.302	3.496
<u>Continuous Variables</u>			
Team FTE	0.112	0.085	1.319
Team Mean Education	-4.263	1.161	-3.671
Mean Distance to Study Site	0.033	0.021	1.565
<u>Analytical Approach</u>			
Experiment *vs* Survey	-0.752	0.216	-3.485
(Base Group: Statistical Surveys)			

<u>Notes</u>

Dependent Variable: Standard deviation of experimental finding or survey response
Residual standard error: 0.9508
Sample size: 415
Multiple R-square: 0.44

Table 3. Decomposition of Research Inputs' Net Knowledge Effects, Selected Aquacultural Research Studies, 2007 – 2009.

Research Input	Knowledge Contribution *via* Research Mean Surprise	Knowledge Contribution *via* Research Accuracy	Total Knowledge Contribution
<u>Continuous Variables</u>			
Team FTE	0.363	-0.079	0.284
Team Mean Education	0	3.026	3.026
Mean Distance to Study Site*	0.189	0.023	0.212
<u>Analytical Approach</u>			
Experiments *vs* Surveys (Base Group: Statistical Surveys)	-1.854	0.534	-1.32

Notes: Contributions in the first column are elasticities in table 1 multiplied by mean surprise's marginal positive contribution (1.76) to scientific knowledge. Contributions in the second column are elasticities in table 2 multiplied by standard deviation's marginal negative contribution (- 0.71) to scientific knowledge. Numbers for the continuous inputs are percentage changes induced by a one-percent change in the indicated input. Those for categorical variables are percent changes associated with switching from the base group to the group indicated.

* Mean distance effects in table 3 refer to a *reduction* in mean distance.