WORKING PAPER SERIES

The Role of Data and Program Code Archives in the Future of Economic Research

Richard G. Anderson
William H. Greene
B. D. McCullough
and
H. D. Vinod

_____

The Role of Data & Program Code Archives
in the Future of Economic Research

Richard G. Anderson
Federal Reserve Bank of St. Louis

William H. Greene
New York University

B. D. McCullough
Drexel University

H. D. Vinod
Fordham University

July 2005 (revised)

Abstract

This essay examines the role of data and program-code archives in making economic research "replicable." Replication of published results is recognized as an essential part of the scientific method. Yet, historically, both the "demand for" and "supply of" replicable results in economics has been minimal. "Respect for the scientific method" is not sufficient to motivate either economists or editors of professional journals to ensure the replicability of published results. We enumerate the costs and benefits of mandatory data and code archives, and argue that the benefits far exceed the costs. Progress has been made since the gloomy assessment of Dewald, Thursby and Anderson some twenty years ago in the *American Economic Review*, but much remains to be done before empirical economics ceases to be a "dismal science" when judged by the replicability of its published results. JEL Classification: B4, C8

1. The Supply of, and Demand for, Replication

In theory, economic research is *scientific* research—that is, economic theory suggests empirically falsifiable hypotheses and investigators seek to confront those hypotheses with data. Often, observers of science regard all researchers as if they are single-minded seekers of truth. As a general result, this is unlikely to be true in other than the simplest representative-agent models. But, as we discuss, the publication process embeds problems of strategic information transmission, knowledge hoarding, bounded rationality, and lying for strategic advantage (including claiming falsely that empirical results are robust and reproducible). To our knowledge, these issues have not heretofore been addressed together.

Many analysts have noted that actual researchers seek not only truth but also fame and fortune, that is, researchers seek individual rewards. For most academic researchers, publication is essential to their livelihood. Economic models with heterogeneous agents, applied to economists themselves, suggest that utility maximizing researchers will rationally choose outcomes other than those chosen by a social planner whose goal is to maximize scientific progress per dollar spent. Further, information-theoretic models suggest that, unless the preferences of authors, editors and readers are fully aligned, authors perceive a strategic advantage to withholding some information. Early commentators include Mayer (1980) and Kane (1984); formal models have been presented by Feigenbaum and Levy (1993, 1996) and Levy and Peart (2001).[1] These models suggest that researchers will exercise a level of care in their research that is less than that which would be chosen by an omnipotent social planner, and that researchers will be reluctant to share data and programs if such reluctance delays (or makes impossible) attempts to replicate their published results.[2] These professional practices of economists are not consistent with uniform (and optimal, from a social planners' viewpoint) truth-seeking behavior.

Ideally, investigators should be willing to share their data and programs so as to encourage other investigators to replicate and/or expand on their results.[3] Such behavior allows science to move forward in a Kuhn-style linear fashion, with each generation seeing further from the shoulders of the previous generation.[4] At a minimum, the results of an endeavor—if it is to be labeled "scientific"—should be replicable, i.e., another

---

[1] Levy and Peart (2001) discuss the role of truth-seeking behavior in representative-agent models where every agent is a truth-seeker and in models with heterogeneous agents where individual rational choice suggests otherwise. Two other interesting papers that use a principal agent framework are Hoffman and Just (2000) and Kilpatrick (1998). In both papers, replication solves a principal agent problem in which researchers lack incentives sufficient to exercise the optimal level of care regarding false results, that is, the issue is to align individual agent's incentives with the socially optimal outcome. In the models, false results are revealed only via replication; hence, the researcher's incentive against replication. The social planner, however, requires that the researcher share all data.

[2] The numerous comments included in the same volume as Feigenbaum and Levy (1993) review most of the arguments regarding replication/verification of published studies.

[3] In this analysis, the term "replication" refers to reproducing an author(s)' results using the same dataset and software. Other authors have referred to the same concept as "verification" (Hyman, 1972) and "certification" (Leamer, 1993).

[4] Kuhn (1972) characterizes science as such a linear process. Others have objected, noting occasional backsliding due to the bandwagon effects of false but seductive ideas that divert scientists on to false avenues of research.

researcher using the same methods should be able to reach the same result. In the case of applied economics using econometric software, this means that another researcher using the same data and the same computer software should achieve the same results. Yet, it is well-known that the likelihood of replication by a reader, or subsequent researcher, without the original authors' programs and data is near zero.

How confident would you be of the published results in an empirical economics article if:

- o No one had ever attempted to replicate the results?

- o Replication had been attempted—and failed?

- o The original authors could not reproduce their own results?

One might think that the solution to the problem is simply to require that all empirical research be reproducible—but it is not quite so simple. Consider some more answers to the above question:

- o If the results had been replicated using the original software package?

- o If the results had been replicated using the original software package, but a newer version of the same package produced different results?

- o If the results had been replicated using the original software package, but a different software package produced different results?

- o If two would-be replicators, both using the original software package, found that one could replicate and the other could not?

So, what is the "replicability" policy of most current professional journals? Most journals have no such policy. A few journals have "replication policies" whereby authors "must" provide data and code upon request, and even fewer journals have mandatory data/code archives, where the researcher must deposit his data and code prior to publication. There is no journal, however, that can provide tangible evidence that its published results are, generally speaking, replicable. To date, every systematic attempt to investigate this question has concluded that replicable economic research is the exception and not the rule.

The implicit replication standard of economics journals is to assume (or, perhaps, pretend?) that the quality of the research underlying its articles is of sufficiently high quality that another researcher could, if desired, replicate the articles' results. In fact, modern empirical economic research is too complex for such a simple assertion—and has been for a half-century or more. Few journals would even attempt to publish a description of all an article's data sources and every programming step. But, without knowledge of these details, results frequently cannot be replicated or, at times, even fully

understood. [5] Recognizing this fact, it is apparent that much of the discussion on replication has been misguided because it treats the article itself as if it were the sole contribution to scholarship -- it is not. We assert that Jon Claerbout's insight for computer science, slightly modified, also applies to the field of economics[6]:

> An applied economics article is only the advertising for the data and code that produced the published results.

Can such "advertising" be misleading? And, if so, does there exist a mechanism for enforcing "truth in advertising"? As we discuss further in the next section, publication in professional journals may be interpreted as a strategic information problem. The published article's results and conclusions are but an advertisement for the substantive research—how those results were obtained. But, authors may perceive a strategic advantage to publishing the conclusions while withholding the underlying data and program code. To the extent that the data and program code are useful for future research, some authors may wish to prevent authors from "catching up" in the research race. If the research is valid, then hoarding this material—the core of the article's contribution to knowledge—is harmful to the progress of science because it does not permit tests for replication, robustness, and refutation. On the other hand, to the extent that the data and code would reveal the written research to be incorrect, this hoarding protects the dishonest or incompetent researcher.

Examples abound. When Breusch and Gray (2004) reanalyzed and replicated the work of Chapman, *et al.* (2001), they found "strikingly different results." As another example, the book by Giles and Tedds (2002) estimated the size of the Canadian underground economy to be 3.4% of GDP in 1976, rising to 15.6% in 1995. The policy implications of such a result are staggering. Yet, when this work was reanalyzed and replicated, it was found that "the overall level of their estimates is a result of numerical accidents" (Breusch, 2005).

The experiences of economists who have sought to replicate published research suggest that Panglossian faith in economists' commitment to the scientific method is

---

[5] The global-warming debate provides an illustration outside economics. In an important article, Mann, Bradley and Hughes (1998) presented evidence of temperature warming during the twentieth century, relative to the previous several centuries. Their article became prominent when one of its charts (a hockey-stick shaped scatter plot, with a "shaft" consisting of historical data and a "blade" consisting of upward-sloping twentieth century data) was featured prominently in the 2001 report of the U.N. Intergovernmental Panel on Climate Change (the Kyoto treaty). As expected, high visibility invites replication and tests of robustness. In a series of papers, McIntyre and McKitrick (2003, 2005a, 2005b) have chronicled their difficulties in obtaining the data and program code; the publishing journal, *Nature,* did not archive the data and code. After some delay, the authors provided the data (see Mann et al., 2004) but have declined, at least as of this writing, to furnish their statistical estimation programs despite their statement that the statistical method is the principal contribution of their article, specifically, to "…take a new statistical approach to reconstructing global patterns of annual temperature back to the beginning of the fifteenth century, based on calibration of multiproxy data networks by the dominant patterns of temperature variability in the instrumental record." (Mann et al. 1998, p. 779). McIntyre and McKitrick's examination suggests that Mann et al.'s statistical procedure (a calibrated principal components estimator) lacks power and robustness; specifically, that the procedure induces hockey-stick shapes even when the true data generating process has none.
[6] Jon Claerbout, noted geophysicist and computer scientist, is The Cecil and Ida Green Professor at Stanford (see Buckheit and Donoho, 1995).

unwarranted: The incentive structure of publish-or-perish is just too strongly skewed toward irreproducibility. Numerous authors have explored this nexus. Dewald et al. (1986) discussed the rational choices of journal editors who fear reducing the attractiveness of their journals to authors by requiring disclosure of data and programs. Diamond (1996) proposed a more general model of the behavior of economists as scientists. Mirowski and Sklivas (1991) and Feigenbaum and Levy (1993) constructed formal models that suggest, in equilibrium, the number of replications is near zero.

In these models of economists' behavior, the incentives are straightforward. Researchers receive a stream of rewards for new knowledge that begins with publication and eventually tapers to near zero. A replication that demonstrates the results to be false immediately ends the reward stream; if the replication further uncovers malicious or unprofessional behavior (e.g., fraud), negative rewards flow to the researcher. So long as withholding data and program code does not reduce the post-publication stream of rewards (and disclosure of data and program code does not increase it), researchers will rationally choose not to disclose data and programs.[7] Such models largely explain the well-known proclivity of academic researchers in many disciplines to keep secret their data and programs (*e.g.*, Fienberg et al., 1985; Bornstein, 2001; Boruch and Cordray, 1985; Bailair, 2003).

At least insofar as the accuracy of published results is concerned, applied economics is a "poor relation" to theoretical economics. Theoretical economic results, generally speaking, are sounder than empirical economic results. The reason for this is simple: the process by which the researcher obtained the result is transparent and amenable to verification. Frequently referees check to make sure that theorems are correct and, if the referee has not vouchsafed every part of the article, the interested reader can do so. Not so with empirical economics, where the process of obtaining a result is far from transparent, and myriad details not described in the text can be found only in the code. Empirical economics, by actively discouraging replication, does not incorporate the self-correcting mechanism of the scientific method -- there is no process whereby bad results can be removed from the cumulative body of knowledge.

Previous studies by two of the present authors have found that, left to themselves, many economists (and econometricians) do not understand the difference between *algebraic* calculations and *numerical* calculations; Altman (2003) contains a number of excellent discussions of the issues. Perhaps the most frequent (and egregious) example is calculating the coefficient vector of an ordinary least squares regression. The algebraic formula, $\hat{b} = \left(X'X \right)^{-1} X' y$, is well-known to have poor numerical properties (McCullough and Vinod, 1999). It is nonetheless commonplace for GAUSS and Matlab code written by economists to implement the algebraic formula rather than the QR decomposition. But, without access to an author's code, how can a reader know how the article's results were obtained?

---

[7] The model of Feigenbaum and Levy (1993), in which rewards to researchers are driven by citations, also suggests that the divergence between the search for truth and rational individual choice will be largest for younger researchers (such as those without academic tenure), who will be less inclined to search for errors than older researchers and less inclined to devote scarce time to documenting their work.

In counterpoint to the arguments advanced above, some researchers have argued that replicating or verifying published results is unnecessary. These arguments follow two lines. First, some have argued that most published scientific results are of no interest or importance and, hence, replication of such results should neither be eased nor encouraged. The arguments of Collins (1993) are dated but remain of value because he makes his argument starkly and because similar views continue to be advanced:

> … no-one cares about the large majority of scientific results—whether they are right or wrong makes no difference to anyone. A reasonable guess, based on some old surveys, is that about 50 percent of scientific papers are not read by anyone except the author, referees, and editor, while studies of the citation indices suggest that about 90 percent of published papers are never cited. (Collins, p. 233)

To economists, Collins' argument seems odd indeed. Intelligent people operating in competitive markets seldom knowingly squander real resources, and significant resources were used to produce, review and publish such articles. Further, presumably knowledgeable editors and referees agreed the articles contained new knowledge. And, as Collins later notes, it is impossible to know ex ante which published papers later will become prominent. In his comment, Collins suggests that the experiments of Dewald et al. (1986) held little value because few, if any, of the *JMCB* articles examined mattered for the progress of economic science. One can only imagine the ire of the authors, bristled by the mild request that they submit data and program code, now being told that their contributions mattered not at all to economics! Further, Collins ignores the result in Dewald et al. (1986) that authors are far more likely to have at-hand an article's data and programs at the time of publication than several years thereafter. Archiving data and program code hedges all researchers against the uncertain future importance of published papers.

The second line of argument, not orthogonal to that advanced by Collins and others, is that the competition of ideas eventually will bring to the forefront all "important" results; typical of this line of argument is Hamermesh (1997). Such results, it is argued, will be verified, replicated, and repeated until a synthesis emerges, with or without archives of data and program code. This argument is appealing because it not only absolves the researcher from his scientific responsibility to produce replicable results, it also suggests the possibility that even though no individual study is replicable, "on balance" the intellectual marketplace will get the "average" outcome correct. This is unlikely due to sample-selection publication bias. More sophisticated analyses, such as Bornstein (1991) and Feigenbaum and Levy (1996), recognize that researchers have significant control over the set of results submitted for publication. A sample-selection bias arises because results that are perceived as "significant" generate a larger stream of rewards to the researcher than do others. This bias does not vanish even in repeated trials by multiple researchers. The median of a set of biased individual estimates is not an unbiased estimate of the true median, and the meta-estimate of the median will coincide with the true median on a set of measure zero. Replication from the authors' data and program code promises not only verification of the original results, but also an opportunity to explore the result's robustness. Replication/verification/extension that

begins without the authors' original data and program code has no power to examine publication sample-selection biases.[8]

Yet, the importance of the sample-selection publication bias has been a matter of dispute. Some papers have proposed a "meta-data" approach to science, essentially an analysis-of-variance or response-surface analysis of how experimental results vary with observable design factors such as the chosen data set or the method of estimation; see Hunter and Schmidt (1996), Stanley (2001), or Hert et al. (2004). These studies ignore the publication-bias problem, however. Further, even if this meta-analysis epistemology eventually functions as intended, it may be horribly inefficient: It does not address the issue of research path-dependence, that is, of researchers being diverted down blind alleys by incorrect results, nor the highly unequal prominence of professional journals. How many results, published in lesser journals, does it takes to outweigh the results published in a single *American Economic Review* article? These issues, and others, can be rendered moot simply by requiring that authors make public their extant data and program code.

2. The Theory of Strategic Information Transmission and "CheapTalk"

Creating original research manuscripts for professional journals is craft work. Although often referred to as "knowledge workers," researchers might equally well be regarded as artisans, with creative tasks that include collecting data, writing code for statistical analysis or model simulation, and authoring the final manuscript.[9] Similar to other craftsmen, researchers' output contains intellectual property—not only the final manuscript, but also the data and programs developed during its creation. Yet, for the academic-type researchers with which we are concerned in this analysis, publishing in peer-reviewed journals is necessary to maintain professional viability (and, for those without academic tenure, employment). A strategic situation naturally arises in which the researcher feels compelled to reveal a sufficient amount of his material to elicit publication, while simultaneously seeking to retain for himself as much of the intellectual property as possible. There are few available models of such a process in the literature.

In this analysis, we interpret the publication process—consisting of submission, refereeing, and an editorial decision—as a problem of strategic information transmission, in the classic sense of Crawford and Sobel (1982). The process also may have aspects of bargaining, to the extent that an author receives a "revise and re-submit" request from an editor.[10] Further, the process may have aspects of a repeated game to the extent that the author during his career may submit multiple different manuscripts to the same editor. If

---

[8] There also is George Stigler's well-known dictum regarding product differentiation in intellectual research: both results that are too similar to, and too different from, previously published results will be difficult to publish. Hence, as a matter of rational choice among scientists, an accepted but incorrect result, unless proven false by replication, can persist in the literature for a significant period of time.

[9] Indeed, "polishing" the final manuscript is well-except vernacular.

[10] Crawford and Sobel (1982) note that their primary motivation stems from the theory of bargaining. The extensive subsequent citations to their article (approximately 300 in the Social Science Citation Index since publication, and 120 in the last five years) support the model's broad applicability to situations where partially "hiding your hand" (not revealing all information) influences a decision by others, such as our context here.

the repeated submissions resemble a long-memory process, then the backward-looking learning that arises from repeated contacts can lead to convergence of the preferences of agents (Blume and Arnold, 2004).

The Crawford–Sobel model contains two agents whose preferences differ by a measurable distance. One agent sends information to the other, who will make a decision. The outcome of the decision enters both utility functions. The better-informed agent is designated the sender, S, and the decision maker is designated the receiver, R. The sender has private information that is of value to both agents, but the sender wishes to share only enough valuable information to induce the desired decision by R. The preference function of R is defined before information is received from S, that is, R's preferences are independent of the information sent by S.[11] We interpret the better informed agent as the author of a specific manuscript. We interpret the receiver as the editor of a journal, who has preferences regarding the type of material desired for the journal and the minimum quality of article that will be published in the journal. We do not specify the process that forms the receiver's (editor's) "preferences," but his choices likely are constrained by either an editorial board or the association that owns the journal (if any). The sender's private information includes knowledge of the "quality" of the article's results, including the quality of the author's computer programs, the robustness of the results to minor variations in the data, and the degree of care exercised by the sender during the course of the research.

Crawford and Sobel construct a continuous measure of the degree of strategic interaction between the sender and received by allowing the sender, S, to partition his private information into subsets, that is, the sender partitions the support of the probability distribution of the variable that represents private information. His decision variable is which partition(s) to send to the decision maker, R. In the model, the "best" decision is made when S sends all the available information to R. How does this paradigm map into the interaction between an author and a journal's editor and referees? For a careful author with a robust result, sending more information (data and programs), when requested, cannot reduce the likelihood of publication. For a less careful author, sending more information increases the likelihood that the result will be discovered as fragile, incorrect, or fraudulent, whether before or after publication. (We discuss below that the flow of benefits to an author from a published article cease almost immediately when the results are discovered to be false.) Similarly, as we discuss in detail elsewhere in this analysis, the sender's incentive to re-check results for accuracy—multiple times, if necessary—depends on the likelihood of an error being uncovered and the professional penalty that is incurred for publishing flawed work. The sender's incentive is to withhold, conditional on obtaining publication, as much information as possible.

In the Crawford-Sobel model, the strategic withholding of information by the sender vanishes when the sender and receiver have the same preferences. For academic publishing, the convergence of preferences likely is an increasing function of the prominence of the authors. Even outside the realm of professional journals, it is well-recognized that many of the more prominent academic economists place their data and programs freely on their own web pages. Further, it is well-recognized that sharing such

data and programs often increases the prominence and visibility of the authors. In the limit, many of these authors become editors (or associate editors) of prominent journals—and their preferences perhaps converge, in the Crawford-Sobel sense. Recently, as we discuss elsewhere in this analysis, a number of journals have established data and/or program code archives. As we document below, considered as a function of time, the rate of establishment of program and data archives has been uneven with few until very recently. Why the uneven pattern? To proceed our later discussion, we note that the Crawford-Sobel analysis suggests an answer. The surge in the establishment of archives during the last few years, ten years after the release of the first Mosaic web browser and the widespread opening of the World Wide Web to researchers, perhaps reflects a shrinkage of the distance between the preferences of sender and receives, of authors and editors who both have found more open information sharing beneficial.[12]

The importance of the convergence of preferences for reducing strategic information withholding has been reinforced by recent empirical studies in the knowledge management literature. Wasko and Faraj (2005) examine knowledge sharing within a legal professional association. Members contributed information to the association's web site which was freely shared. Benefits to the members were primarily reputational, including being perceived as more clever and capable than others. Better known members tended to contribute more knowledge, ceteris paribus, as did members with more interactions with other members. Parallels could be drawn to prominent economists who act as editors of major journals but freely share their data and programs. Kankanhalli et al. (2005) repot a similar study of large-scale public sector organizations, interpreting the results via social capital theory. Their conclusions are similar to Wasko and Faraj. Finally, Lin et al. (2005) examine how knowledge users identify "experts," that is, how professionals who share information become known in an informal community as experts. In their sender-receiver model (an extension of Crawford-Sobel), information asymmetry induces "experts" (including, say, leading researchers) to withhold private information from the market, thereby increasing the barriers for others to assess the quality of available information (such as the results in published articles) and slowing the production of new knowledge. Their model suggests no solution to this problem beyond that of Crawford and Sobel: more closely align preferences between senders and receivers to reduce the strategic advantage of withholding of private information.

In summary: The publication process for professional journals in economics is a strategic information problem. Current practice in economics is to define "publication" as the event of a manuscript/article appearing in print. Best *scientific* practice, however, would suggest re-defining "publication" to include the underlying data and programs. Yet, to the extent that the programs and data are valuable for future research, economic theory suggests that authors will seek a strategic advantage by withholding such information, if possible. Such strategic information models do not have unique equilibria, and hence comparative statics exercises are subtle. Yet, the models suggest that the withholding of private information for strategic advantage is a decreasing function of the distance between the sender's and receiver's preferences; solving the withholding problem requires aligning the preferences of the sender and receiver.

---

[12] In part, this idea echo's Kuhn's theory of scientific revolution in which the rate of adoption of new ideas is an increasing function of the rate of retirement of older researchers.

3. The Past and Present

Two separate issues arise in replication of published articles: (1) obtaining the article's data and programs, and (2) using the data and programs to reproduce the article's results. Professional journals have most commonly addressed the first by adopting a "publication policy" that requires authors, *as a condition of publication*, to provide to readers, on request, the article's nonconfidential data and/or programs. Does such a policy work?

As an experiment to "stress-test" such policies at three journals—the *American Economic Review* (AER), the *International Journal of Industrial Organization* (IJIO), and the *Journal of International Economics* (JIE)—McCullough and Vinod (2003b) requested data and programs from the authors of the articles in the [then] most-recent issue. Despite the authors' pre-publication commitment to provide such materials, only one-third did so. The results are shown in Table 1.

Table 1: Do Authors Honor Replication Policies?

Journal	Number of authors asked to supply data and code	Number of authors that supplied data and code
IJIO	3	1
JIE	4	1
AER	8	4
Total	15	6

Note: The figures indicate only if the author(s) supplied data and code, not whether the supplied data and code actually reproduced the published results.

Source: McCullough and Vinod (2003b)

The second issue—replicating published results from authors' materials—has been examined in several studies. Dewald et al. (1986) examined 54 data sets for articles submitted to and/or published by the *Journal of Money, Credit and Banking* (*JMCB*) during 1982-1984. Adopting a mandatory data/code archive is no more a guarantee of replicable research than is a voluntary "replication policy." They judged that only 8 were sufficiently complete and well-documented so as to permit a straightforward replication attempt. In two studies, McCullough et al. (2005a, 2005b) examined the later archives from the *JMCB* and Federal Reserve Bank of St. Louis *Review*, respectively.[13] The archived materials reproduced the published results in only 5 to 7 percent of the sampled articles. These studies are summarized in Table 2.

Table 2: Do Archives Work?

---

[13] At the time of McCullough (2005a)'s study, all of the *JMCB* archives examined by Dewald et al. had been discarded by the *JMCB*. Hence, theirs is a set of newer articles with no overlap to the previous study.

Article	Empirical articles requiring sharing of data and code	Actually supplied data and code	Data and code replicate published results	Replicators lacked software to run code
Dewald et al. (1986)	na	54	2	na
McCullough et al (2005a)	193	69	14	7
McCullough et al (2005b)	236	78	9	19
Total	429	201	25	26

Similar conclusions have been found in other social sciences; for political science, see King (1995, 2003) and Ray and Valeriano (2003); for psychology, see for example the papers in Neuliep (2001); for discussion of the physical sciences, see for example Fienberg et al. (1985) and Bailer (2003).

The difficulties are further well-illustrated by the experience of the *American Economic Review*. At the time of publication of Dewald et al., the *AER* adopted a "replication policy" that required authors to retain an article's data and share non-confidential data with readers on request (Ashenfelter, et al, 1986). This policy was substantively weaker than the mandatory data/code archive recommended by Dewald et al. Subsequently, many other journals followed the lead of the *AER*, including *The Economic Record*, *Journal of International Economics*, *Journal of Human Resources*, *International Journal of Industrial Organization* and *Empirical Economics*. Non-compliance with such a policy was predictable—the policy offered little or no reward to authors for compliance, while continuing to expose them to the risk that a flaw might be uncovered in their work (Mirowski and Sklivas, 1991). The policy was tested, and found wanting, seventeen years later in McCullough and Vinod (2003a). In response, the *AER* implemented a mandatory data and program code archive (Bernanke, 2004).[14] This brings to eight the number of journals with data and/or code archives: *American Economic Review, Econometrica, Review of Economic Studies, Macroeconomic Dynamics*, the *Journal of Money, Credit and Banking*, the Federal Reserve Bank of St. Louis *Review* , the *Economic Journal*, the *Journal of Applied Econometrics*, and the *Journal of Business and Economic Statistics*. The first five require both data and code, while the last three require only data; the first three joined the club in the last year – significant momentum for the club.

Some authors have praised replication as an important aspect of economic research, even when replication efforts have produced corrections to published results. In acknowledging a programming error in a previous article, Feldstein (1982), p. 630, wrote:

> "They [Leimer and Lesnoy] set an admirable example of the tradition of replication on which all scientific work ultimately rests. As economic research increasingly involves large and complex computer programs to

---

[14] A cynic might point out that the *AER*'s 1986 replication policy statement appeared on the *first* page of its issue of the journal (just before the Dewald et al. article), while the 2004 announcement appears on the *last* page of its issue, following the notes and comments section. But perhaps the ultimate efficacy of the announcement is inversely proportional to the publicity the editor gives it; we hope so.

analyze microeconomics datasets or simulate models that cannot be solved analytically, replication studies like that of Leimer and Lesnoy should become increasingly important."

The reluctance of academic researchers to share data and programs has been widely discussed for more than twenty years, even when their research has been supported with public funds from, say, the National Science Foundation.[15] The historical record suggests that neither *AER*-style data-sharing "policies," nor laissez faire data archives, are adequate for replication (e.g., McCullough et al., 2005a, who analyzed the archive of the *Journal of Money, Credit and Banking*).   A journal's archive policy must include guidelines for authors: programs should include comments for replicators; the names of datasets should match the names used in programs, *all* data transformations should be described, etc. Such rules follow from the experiments with the archives of the *JMCB* and Federal Reserve Bank of St. Louis *Review* discussed in McCullough et al. (2005a, 2005b), who found replication often thwarted by authors' disorganized materials. In response to McCullough et al. (2005a), the *JMCB* is revising its archive rules. In response to McCullough et al. (2005b), we expect that the Federal Reserve Bank of St. Louis will revise its rules, too.[16]

> In its first issue, the editor of *Econometrica* (1933), Ragnar Frisch, noted the importance of publishing data such that readers could fully explore empirical results. Publication of data, however, was discontinued early in the journal's history. In a 1966 issue of the journal, Zellner and Thornber (1966) showed that the same code run on different computers could produce different answers. The journal's editorial board did nothing – did not even require authors to reveal the computer on which they had run their code. The journal arrived full-circle in late 2004 when *Econometrica* adopted one of the more stringent policies on availability of data and programs.[17]

> Yet while *Econometrica* went full-circle in one direction, *Journal of Political Economy* did so in the other direction. In the pages of the *JPE*, Feige

---

[15] On this subject, the excellent papers in Fienberg et al. (1985), are as relevant today as when they were written. Cecil and Griffen (1985) discuss legal issues regarding data sharing for publicly funded projects. Their analysis suggests that, contrary to popular opinion, data and programs produced by publicly funded projects are not automatically in the public domain as a result of public funding.

[16] The archive of the Federal Reserve Bank of St. Louis *Review* began in the fall of 1992 when the new research director, William Dewald, required that the results in each *Review* article be replicated in-house by a research analyst prior to publication. Initially, files were distributed on floppy disks and via the Research Division's dial-in computer bulletin board service. Beginning in 1995, the files were made available on the Division's Internet web pages via links placed next to each published article, and on the public archive for published articles maintained by the ICPSR at the University of Michigan. So far as we are aware, no other Federal Reserve Bank research publication has a public data/program archive. During 2004, the Federal Reserve Bank of St. Louis *Review* became the second, and only currently published, Federal Reserve Bank publication to be indexed in the *Social Science Citation Index* (the first was the Federal Reserve Bank of Boston's *New England Economic Review*, which ceased publication in 2004). Because inclusion in the Citation Index is based, in part, on the frequency with which a journal is cited in other journals, inclusion of the *Review* lends support to the arguments of Tauchen (1993) and Gleditsch et al. (2003) that articles published in journals with data and program code archives are more valuable to researchers, and hence more likely to be cited, than articles without.

[17] <http://www.econometricsociety.org/submissions.asp#4>, as of January 13, 2005.

argued for, as a minimum standard, "full reporting of procedures and data" by journals (Feige, 1975); this minimum standard would effectively require providing sufficient information to permit replication. In response, the editors of the *JPE* did not require full reporting of procedures and data, but instead instituted a "Confirmations and contradictions" section, whose purpose was to foster the use of alternative statistical tests of hypotheses. They noted, "Confirmations will require new data; contradictions will be most powerful when based on the same data" (JPE Editors, 1975). Replication might seem a natural by-product of such work; can one really understand the method by which the original results were produced without first replicating them? Mirowski and Sklivas (1991) analyzed the thirty six notes appearing in this section from 1976 through 1987. Only five involved actual replications, but one of which was successful. This twenty percent success rate at the *JPE* is better than what Dewald, et al. found at the *JMCB*, but it hardly inspires faith in the replicability of results published in the *JPE*. Updating these numbers, between 1988 and 1999, the section published 13 notes, only one of which contained a replication. Apparently *JPE* has allowed the section to die an ignominious death befitting the section's true relation to replication: it has been inactive since 1999. *JPE* has gone back into the fold with those journals that do not even bother to pay lip service to the idea that published results should be replicable.

Prospects for attaining the goal of publishing replicable research are bright for those journals and their editors, such as *Econometrica,* that care about replicability. Eight of the oldest and most prestigious economics journals are published by not-for-profit organizations and professional societies—the *AER*, *Econometrica*, *Review of Economics and Statistics*, *Quarterly Journal of Economics*, *Journal of Political Economy*, *The Economic Journal*, *Journal of Business and Economics Statistics,* and the *International Economic Review.* Currently, four of these owners (the *AER*, *Econometrica*, *JBES*, and *EJ*) have attached a data archive to their journal. The others should do so also.

4. The Benefits

A number of benefits will arise from widespread availability of data and program archives. Many of these benefits can appear only if journal archives include data and program code. Experience and evidence confirm that data-only archives are not adequate to enable replication of most published results (e.g., Racine, 2001; McCullough et. al., 2005a, 2005b). Data-only archives should be broadened to include both data and program code.

The costs of data and program code archives are so low that even modest benefits produce a large benefit/cost ratio. In the past, some editors argued that journals would incur a significant financial cost to operate an archive. The Inter-University Consortium for Political and Social Research (ICPRSR) estimated that during the mid-1980s, for example, it cost $25 to prepare and send a dataset. Today, in the "Internet age," the costs of operating a data and program archive are very low.[18] The cost to authors of

---

[18] The *JMCB* data archive constructed by Dewald et al. at Ohio State University between 1982 and 1984 consisted primarily of reels of mainframe (IBM) style computer tape and decks of punched cards. Occasionally, a lucky (and well-funded) researcher owned an IBM PC—especially after release of the IBM

producing materials for archives also is small. In previous studies, authors have stated that their cost of creating and contributing materials was minimal *when they knew in advance that submitting such materials was required.*[19] Further, the willingness and ability of authors to supply materials is higher when the research is fresh than when it has aged a year or two between completion and publication (e.g., McCullough and Vinod, 2003a). Today, the primary "cost" to a journal of a data and program archive is the increased caution focused by the archive on its authors, who wish to avoid errors leading to embarrassment.[20]

Errors do occur in empirical research, replication is the way to uncover them, and professional journal archives are an efficient way to disseminate corrections. The results in large-scale replication studies such as Dewald et al. (1986) and McCullough (2005a, 2005b), suggest that the frequency of inadvertent errors in published articles is not low. Further, some errors can be costly. Suppose, for example, that McCrary (2002) had not found the programming error reversing Levitt's (1997) result in the *American Economic Review* that increases in police substantially reduce crime. If a policymaker had acted on Levitt's finding and shifted funds from, say, low-income housing to police, social welfare would have been reduced. Some errors, and their corrections/corrigendums, also can be difficult to locate. Consider the case of Hansen and Seo's (2002) article in the *Journal of Econometrics*. The numerical values shown in the article's Table 4 are all incorrect (that is, each of the 72 reported p-values is incorrect), and over half of them suggest incorrect statistical inferences at the 5% level (in 40 of the 72 cases, the correct result and the published result are on different sides of 0.05). How do we know this? Hansen published the correct results and code on his personal website—but not in the journal nor on the journal's web site. All researchers occasionally make mistakes; Hansen is to be lauded for placing scientific integrity above the appearance of infallibility. But such corrections leave unanswered the responsibility of the journal that published the article. Corrections to published articles that are never disseminated, obviously, benefit few researchers. From a scientific perspective, erroneous results should be "purged" from the cumulative body of knowledge. To do so, what better method exists than via the data/program code archive of the journal that published the original research?

No one should ask journal editors and referees to ensure that articles are replicable— the cost in time and materials would be prohibitive. Rather, the issue is aligning principal-agent incentives such that authors themselves furnish materials that readily permit replication. Journals can do so by including pages that cite successful or

---

XT in 1982 and, later, the IBM PC AT in 1984—and shared data on 5-1/4 inch floppy disks. (Later, toward the end of the project, we placed most data files on floppy disks.) Such an archive had two drawbacks. First, it was costly (and tedious) to duplicate data on cards and mainframe magnetic tape. Second, as noted by Dewald et al., the archive lacked immediacy. A *JMCB* subscriber, while reading an article, could not simply click on a web link to view the data and programs but rather had to endure a delay of at least several days as data were duplicated and mailed. Dewald (1993) notes that the costs of the archive to the *JMCB* were minimal while he was editor.

[19] Dewald et al (1986); Anderson and Dewald (1994).

[20] The only journal of which we are aware that has an active replication section is the *Journal of Applied Econometrics* which, curiously, has a "data only" archive. Preferably, an archive also should include the output from the program. A difficulty for some researchers is that some Windows-style menu-driven software, such as EViews, does not produce a batch program that can be run to reproduce results. Indeed, many users of menu-driven software are at a loss to write programs.

unsuccessful replications. In response to McCullough et al. (2005a), the *Journal of Money, Credit and Banking* instituted just such a section. Major corrections might deserve several published pages but not every discrepancy between a replication and a published article is so important. A replication section might furnish no more than a URL to a web page containing the replicator's work; such a web page might be operated by the journal, by the replicator, or perhaps be an "e-journal" devoted solely to replication. We anticipate that a large number of archives will generate a larger number of replications and, in turn, demand for an e-journal composed mostly of replication results.

In addition to merely being good science, the use of data/code archives to produce verified/verifiable results will bring other benefits. Reinforcing a speculation in Dewald et al. (1986), Gleditsch et al. (2003) found that articles that provide data are *cited* twice as often as articles that do not provide data—and citations are valuable (Diamond, 1986). One can reasonably expect that the cite rate will be even higher for articles that have data and code.[21]

Broadly, data and program code archives will lead to more and better economic research, for several reasons.

o Research will be done more carefully. Little is likely to focus a researcher's attention on detail more than the knowledge that other researchers—graduate students, professors, and the like—will be able to initiate *their* research project by examining the accuracy of *your* project.

o Research will (finally) be more self-correcting. Even though there will be fewer errors because researchers will be more careful, archives do not ensure that researchers are perfect. Archives do ensure that there is a reasonable chance that errors will be uncovered – as opposed to the status quo, wherein there is an almost zero probability that errors are uncovered.

o Research will progress more rapidly. It will no longer be necessary to reinvent the wheel just to extend or check the robustness of another researcher's results. New studies will better be able to extend previous studies, moving economics closer to Kuhn's image of linear progress.

o Archives will provide a rich source of examples to try new estimation and inference methods.

Additional, less-noted benefits will accrue through better software, for a number of reasons.

o It will be easier for developers to incorporate and standardize new developments. At times, an author's algorithm may be less than transparent in a published article. Subsequent implementations can lead to different software packages giving different numbers. McCullough and Renfro (1999) discuss such problems with

---

[21] As discussed in Anderson and Dewald (1994) and further below, subsequent to Dewald et al. a large group of journal editors seemed to disagree when offered the opportunity to require their authors to submit data to an NSF-sponsored archive.

Bollerslev's GARCH. Fiorentini, Calzolari and Panattoni (1996) proposed a benchmark for the problem, across packages; afterwards, software results became much less disparate (Brooks et al, 2001). Yet, problems remain. McCullough and Vinod (2003b) demonstrated that six packages gave six different answers to the "find the 1% VaR" exercise considered by Engle (2001) in his primer on GARCH.

o   More people solving the same problem with different packages can improve existing routines in software packages by increasing the interaction between users and developers, which is a fundamental component of how software changes over time (Greene, 2004). This can include uncovering bugs and comparing differing algorithms used to implement the same estimators. Developers cannot fix problems in software packages if their existence is not known. Zeileis and Kleiber (2004), for example, ported Bai and Perron's (2003) structural change code from GAUSS to R and were unable to replicate successfully. Their experiments uncovered a weakness in one of GAUSS's distribution functions, which Aptech Systems (developer of GAUSS) quickly fixed. In another example, Bruno and DeBonis (2004) compared several packages' approaches to panel data estimation. Investigating disparate output, they found that all the packages used different, but theoretically valid, estimators. A simulation exercise found that the estimators were not of equal accuracy. In a third example, Stokes (2004) used six packages to solve a problem for which it is known that a numerical solution does not exist, and found that five of the packages claimed to have found a solution.

o   More benchmarks will be produced. Producing a benchmark can be accomplished two different ways. One way is via very careful coding, with every step well-documented, as in the case of the Calzolari and Panattoni (1988) FIML Benchmark. Another way is when two independently coded programs produce the same answer. A recent example is Drukker and Gua (2003), who show that a panel data method in the package Stata produces the same result as an independently programmed result published in *Journal of Applied Econometrics*.

o   Users will better be able to determine the effect of new software releases. Data and program code archives provide a ready-made test bed for developers, and their customers, to check for version incompatibilities and, in some cases, to revisit past research.

Software packages already incorporate the easy-to-program procedures. New methods in the journals tend to be more difficult to program. Data and program archives provide developers a clearer view of algorithms, accelerating the pace at which software can incorporate new estimators.

5. The Future

The economic theory that we outlined in Section Two suggests that authors might perceive strategic advantages in not furnishing data and programs to journals so long as the preferences of authors, editors and readers are not fully aligned. We have suggested that a mechanism for aligning these preferences is mandatory data and program code archives. However, to do this will require leadership on the part of journal editors.

Journal of Money, Credit and Banking, Macroeconomic Dynamics, and Federal Reserve Bank of St. Louis *Review* have, for years, stood alone. In just the past year, the number of journals with mandatory data/code archives has doubled: *American Economic Review* led the way, followed in short order by *Econometrica* and *Review of Economic Studies*. Who will be next to stand for replicability of published results? Or will these six stand alone for several more years, as the majority of economic journals remain uncommitted to publishing replicable research?

As a profession and a science, where should economics seek to be in, say, five years? How does the profession, via collective action, move from a low-replicability equilibrium to a high-replicability one? In our view, professional journals must accept that leadership challenge. As a form of collective action, journals can assure that published research is sufficiently documented—including data and program archives—and that published articles' results are replicable. Absent such a journal structure, both economic models of rational choice—and the evidence—suggests that most authors will not voluntarily choose to incur the costs of creating archival files and documentation.

Archives of data and program code are the key. Outlets have existed in which to publish replication studies for more than twenty years—the *Quarterly Journal of Business and Economics* began encouraging submission of replication studies in 1984; see Mittelstaedt and Zorn (1984) and Kane (1984). Yet, the supply of replication studies depends on journals archiving authors' data and program code. Further, the results of replications, no matter where published, are difficult for subsequent researchers to locate unless they are noted on, or linked to, the journal's archive web page.

In the past, many journal editors seemed uninterested in taking steps toward ensuring that the results they published were replicable. Some editors, indeed, expressed the fear that requiring authors to submit data and programs would reduce the attractiveness of their journal, relative to others that did not have such a requirement. Any such negative impact is reduced to the extent that the availability of data and programs increases the value of a journal to its readers. Yet, as noted in Anderson and Dewald (1994) and discussed further below, even when the National Science Foundation offered journals a *free* archive, their editors refused to require that authors submit data and programs. Of course, adopting a "replication policy" also was free and, since in practice authors were not bound to honor the policy, it imposed an even lower cost on authors. Yet, most professional journals did not even choose to implement a non-binding policy—and the same is true today. Is this science?

Specific actions include:

1. Journals should implement mandatory data/code archives, and implement rules (*e.g.,* those enumerated by McCullough (2005b)) to ensure that the archives function properly.

2. Journals should include replication "announcement" sections. Such sections provide professional credit to researchers who undertake replications. Further, we suggest that all professional journals, on their web sites, establish a location for attaching to published articles a "successfully replicated" comment including a summary and supporting material. In some cases, replication efforts that explore

an article's robustness or accuracy might deserve additional page space. The *Quarterly Journal of Business and Economics* pioneered this in 1984, and more recently has been joined by the *Indian Journal of Economics and Business*, the latter with a broader editorial policy that does not require extension of the original article but encompasses reporting the results of a straightforward replication attempt. The *IJEB* also has made the replication articles available to all, subscribers or not, on their web site, and has placed supporting materials on the journal's web site.

3. Some journals must be willing to publish the results of failed replications when the primary journal will not. This "threat of entry" removes the incentive for any individual editor to seek to increase his journal's attractiveness (to authors, not readers) by making it known that he does not publish replication results. Current examples of journals willing to publish such replications include the *Journal of Economic and Social Measurement* and the *Indian Journal of Economics and Business*. Ideally, the replication sections would be available on the Internet to subscribers and non-subscribers alike.

Economic theory suggests, as discussed above, that the profession's current low-replicability equilibrium may be interpreted as a principle agent problem. If so, leadership by journals via data/program archives likely can lead the profession to a high-replicability equilibrium. At that point, the role of archives and replication as a check on error or fraud likely will diminish, while the frequency of replication as a starting point for new research will increase.[22] If so, we recommend that the economics profession—perhaps with leadership from the American Economic Association—create a single "one-stop-shop" archive for data and programs across journals.[23] In fact, such an archive already exists although it is lightly used. In 1995, the Inter-University Consortium for Political and Social Research (ICPSR)—on its own, without NSF financial support—established an open-access archive for the data and programs of any published article.[24] As of December 2004, approximately 250 articles were included. Two-thirds of the articles were economics-related. Omitting articles related to Michigan's Panel Study of Income Dynamics, *all* the economics articles were from the Federal Reserve Bank of St. Louis *Review*.

   In the case where the cooperation necessary to build a single archive is impossible, we recommend that each journal's archive use XML tags to identify its material. This simple practice will allow all archives to be linked via a VDC/DDI environment, and be

---

[22] Feigenbaum and Levy (1993) predict this dynamic shift in the scientific role of replication. Once replication/verification becomes commonplace, the quality of articles will endogenously improve such that most articles are easily replicated. Yet, the number of replications need not decrease even if the number of replications done as a check on error or fraud decreases because the frequency of replication as a starting point for new studies should *increase*.

[23] Such an AEA-sponsored archive seems a reasonable extension of the proposals of Goffe (2004). Vinod (2001) suggests that such an archive—say, the AEA-sponsored "Resources for Economists" at http://netec.wustl.edu/WebEc—could begin building a body of knowledge in empirical economics by collecting stylized "important empirical facts" on which authors and replicators agree.

[24] Membership is required to access most ICPSR archives. The journal archive is open to all, no membership required.

visible to Internet web spiders and search engines so that a search in Google or Yahoo locates the material.[25]

None of this will occur unless the editors of professional journals take the lead. Ten years ago, Anderson and Dewald (1994) re-visited the issue of replication in economics as a follow-up to Dewald et al. (1986) and found that little had changed. The subsequent ten years have shown progress, however—five economics journals now have archives for data and programs, and three more have data-only archives. Prospects for the future of replicable economic research are bright, primarily due to the fact that the number of journals with data plus program code archives has doubled in the past year, after having remained stagnant for many years. This list must expand in the coming years; much remains to be done. For empirical economics to at last become a science, the profession must embrace the scientific method by archiving and sharing data and program code.

---

[25] On the Virtual Data Center and Data Description Initiative projects, see Altman et al. (2001) and Blank and Rasmussen (2004).

References

Altman, Micah, L. Andreev, M. Diggory, G. King, A. Sone, S. Verba, Daniel L. Kiskis, and M. Krot (2001). "A Digital Library for the Dissemination and Replication of Quantitative Social Science Research: The Virtual Data Center," *Social Science Computer Review*, vol. 19, pp. 458-70.

Altman, Micah, Jeff Gill and Michael P. McDonald, eds. (2003). *Numerical Issues in Statistical Computing for the Social Scientist* (Wiley).

Anderson, Richard G. and William G. Dewald (1994). "Replication and Scientific Standards in Applied Economics a Decade After the *Journal of Money, Credit and Banking* Project," Federal Reserve Bank of St. Louis *Review*, November/December 1994, pp. 79-83.

Ashenfelter, Orley, Robert H. Haveman, John G. Riley, and John T. Taylor (1986), "Editorial Statement," *American Economic Review* 76(4), p. v

Bai, J. and P. Perron (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18, 1-22

Bailar, John C. (2003). "The Role of Data Access in Scientific Replication," paper presented at the October 16-17, 2003 workshop 'Confidential Data Access for Research Purposes,' held by the Panel on Confidential Data Access for Research Purposes, Committee on National Statistics, National Research Council. <www7.nationalacademies.org/cnstat/John_Bailar.pdf>

Bernanke, Ben S. (2004). "Editorial Statement," *American Economic Review*, 94(1), March, p. 404.

Blank, Grant and Karsten Boye Rasmussen (2004). "The Data Documentation Initiative," *Social Science Computer Review*, 22 (3), Fall, pp. 307-18.

Blume, A. and T. Arnold (2004). "Learning to Communicate in cheap-talk games," *Games and Economics* Behavior, 46(2), February, pp. 240-259.

Bornstein, Robert F. (2001). "Publication Politics, Experimenter Bias and the Replication Process in Social Science Research," in Neuliep (2001), pp. 71-84.

Boruch, Robert F. and David S. Cordray (1985). "Professional Codes and Guidelines in Data Sharing," in Fienberg et.al.

Breusch, Trevor and Edith Gray (2004), "New Estimates of Mothers' Foregone Earnings Using HILDA Data," *Australian Journal of Labour Economics* 7(2), 125-150

Breusch, Trevor (2005), "The Canadian Underground Economy: An Examination of Giles and Tedds," *Canadian Tax Journal* (to appear)

Brooks, Chris, Simon P. Burke, and Gita Persand (2001). "Benchmarks and the accuracy of GARCH model estimation, " *International Journal of Forecasting* 17, pp.45-56

Bruno, Giuseppe and Riccardo De Bonis (2004), "A Comparative Study of Alternative Econometric Packages with an Application to Italian Deposit Interest Rates," *Journal of Economic and Social Measurement* 29(1-3), 271-295

Buckheit, J. B. and D. L. Donoho (1995). "WaveLab and Reproducible Research" In: *Wavelets and Statistics*, ed. A. Antoniadis. New York, Springer-Verlag: 53-81

Calzolari, Giorgio and Lorenzo Panattoni (1988). "Alternative Estimators of FIML Covariance Matrix: A Monte Carlo Study," *Econometrica* 56:3, May, pp. 701-714.

Cecil, Joe Shelby and Eugene Griffen (1985). "The Role of Legal Policies in Data Sharing," in Fienberg et al.

Chapman, Bruce, Yvonne Dunop, M. Gray, Amy Liu, and D. Mitchell (2001), "The Impact of Children on the Lifetime Earnings of Australian Women: Evidence from the 1990's," *The Australian Economic Review* 34, 373-389

Collins, Harry M. (1993). "Comment [on Feigenbaum and Levy]," *Social Epistemology*, vol. 7, no. 3, July-September, pp. 233-236.

Crawford, Vincent P. and Joel Sobel (1982). "Strategic Information Transmission," *Econometrica*, 50(6), November.

Dewald, William G. (1993). "Comment [on Feigenbaum and Levy]," *Social Epistemology*, vol. 7, no. 3, July-September, pp. 243-4.

＿＿＿ , Jerry G. Thursby, and Richard G. Anderson (1986), "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project," *American Economic Review* 76(4), pp. 587-603

Diamond, Arthur M. (1986). "What is a Citation Worth?" *Journal of Human Resources*, 21(2), Spring, pp. 200-15.

＿＿＿ (1996), "The Economics of Science," *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization*, vol. 9, pp. 6-49.

Drukker, David and Weihua Guan. (2003) "Replicating the Results in 'On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators'." *Journal of Applied Econometrics*, 18:1, 119.

Engle, Robert (2001), "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics," *Journal of Economic Perspectives*, 15(4), Fall, pp. 157-68.

Feige, Edward (1975). "The Consequences of Journal Editorial Policies and a Suggestion for Revision," *Journal of Political Economy*, 83:6. 1291-1295.

Feigenbaum, S. and D. Levy (1993). "The Market for (Ir)Reproducible Econometrics," pp. 215-232, and "Response to the Commentaries," pp. 286-292, *Social Epistemology*, 7:3.

___ (1996). "The Technological Obsolescence of Scientific Fraud," *Rationality and Society*, vol, 8, pp. 261-76.

Feldstein, Martin S. (1982). "Social Security and Private Saving: Reply," *Journal of Political Economy*, June, pp. 630-42.

Fienberg, Stephen E., Margaret E. Martin, and Miron L. Starf, eds., (1985). *Sharing Research Data*. (Washington, DC: National Academy Press, for the National Research Council).

Fiorentini, G., G. Calzolari and L. Panattoni (1996), "Analytic Derivatives and the Computation of GARCH Estimates," *Journal of Applied Econometrics*, 11, 399-417

Giles, David E. A. and Lindsay M. Tedds (2002), *Taxes and the Canadian Underground Economy*, Canadian Tax Foundation: Toronto

Gleditsch, Nils Petter, Claire Metelits, and Havard Strand. (2003) "Posting Your Data: Will You Be Scooped or Will You Be Famous?" *International Studies Perspectives*, 4:1, 89-97.

Goffe, William L. (2004). "The Internet and the American Economic Association—A Set of Proposals," *Social Science Computer Review*, 22(4), Winter, pp. 493-511.

Greene, W. H. (2004), "Reply to Bruno and De Bonis," *Journal of Economic and Social Measurement* 29(1-3), 297-305

Hamermesh, Daniel S. (1997) "Some Thoughts on Replications and Reviews." *Labour Economics*, 4, 107-109.

Hansen, Bruce E. and Byeongseon Seo (2002), "Testing for Two-Regime Threshold Cointegration in Vector Error-Correction Models," *Journal of Econometrics* 110, 293-318

Hert, Carol A., Sheila Denn and Stephanie W. Haas (2004). "The Role of Metadata in the Statistical Knowledge Network," *Social Science Computer Review*, vol. 22, no. 1, Spring, pp. 92-99.

Hoffman, Walter E. and Richard E. Just (2000), "Setting Efficient Incentives for Agricultural Research: Lessons from Principal-Agent Theory," *American Journal of Agricultural Economics*, November, v. 82, iss. 4, pp. 828-41.

Hunter, John E. and Frank L. Schmidt (1996). "Cumulative Research Knowledge and Social Policy Formulation: The Critical Role of Meta-Analysis," *Psychology, Public Policy and Law*, 2(2), pp. 324-47.

[JEP] Editors (1975). "Editorial Comment," *Journal of Political Economy,* 83:6, pp. 1295-6.

Kane, Edward J. (1984). "Why Journal Editors Should Encourage the Replication of Applied Econometric Research," *Quarterly Journal of Business and Economics*, 23(1).

Kankanhalli, Atreyi, Bernard C.Y. Tan, and Kwok-Kee Wei (2005). "Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation," *MIS Quarterly*, 29(1), March, pp. 113-143.

Kilpatrick, Henry E. (1998), "Some Useful Methods for Measuring the Benefits of Social Science Research," Impact Assessment Discussion Paper No. 5, International Food Policy Research Institute, Washington DC, October.

King, Gary (1995). "Replication, Replication," *PS: Political Science and Politics*, vol. XXVIII, No. 3, September, pp. 443-499.

___ (2003). "The Future of Replication," *International Studies Perspectives,* vol 4:1, pp. 100-105, February.

Leamer, Edward (1993). "Comment [on Feigenbaum and Levy]," *Social Epistemology*, vol. 7, no. 3, July-September, pp. 268.

Levitt, Steven (1987). "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review*, June, *87*(3), pp. 270 –90

Levy, David M. and Sandra J. Peart (2001). "Econometrics and the Truth-Seeking Assumption: Ethics and Research Independence," mimeo, George Mason University. Paper presented at the Southern Economic Association meetings, Tampa, 2001.

Lin, Lihui, Xianjum Geng, and Andrew B.Whinston (2005). "A sender-receiver framework for knowledge transfer," *MIS Quarterly*, June, 29(2), pp. 197-219.

Mann, Michael E., Raymond S. Bradley, and Malcolm K. Hughes (1998). "Global-scale temperature patterns and climate forcing over the past six centuries," *Nature*, vol. 392, April, pp. 779-787.

___ (1999). "Northern Hemisphere Temperatures During the Past Millenium: Inferences, Uncertainties, and Limitations," *Geophysical Research Letters*, 26, pp. 759-762.

___ (2004). "Corrigendum: Global-scale temperature patterns and climate forcing over the past six centuries," *Nature*, 430 (105).

Mayer, Thomas (1980). "Economics as a Hard Science: Realistic Goal or Wishful Thinking," *Economic Inquiry,* vol. 81, April, pp. 165-78.

McCrary, Justin (2002). "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment," *American Economic Review*, September, 92(4), pp. 1236–43.

McCullough, B.D., Kerry Anne McGeary and Teresa D. Harrison (2005a). "Lessons from the JMCB Archive," *Journal of Money, Credit and Banking*, forthcoming.

___ (2005b), "Lessons from the Fed. St. Louis *Review* Archive," manuscript, Drexel University, January.

McCullough, B. D. and C.G. Renfro (1999). "Benchmarks and Software Standards: A Case Study of GARCH Procedures," *Journal of Economic and Social Measurement*.

McCullough, B. D. and H. D. Vinod. (1999) "The Numerical Reliability of Econometric Software." *Journal of Economic Literature*, 37:2, 633-665.

___ . (2003a) "Verifying the Solution from a Nonlinear Solver: A Case Study." *American Economic Review*, 93:3, 873-892.

___ (2003b), "Comment: Econometrics and Software," *Journal of Economic Perspectives* 17(1), 223-224

McIntyre, Stephen and Ross McKitrick (2003). "Corrections to the Mann et al. (1998) Proxy Data Base and Northern Hemispheric Average Temperature Series," *Energy and Environment*, 14, pp. 751-771.

___ (2005a). "The M&M Critique of the MBH98 Northern Hemisphere Climate Index: Update and Implications," *Energy and Environment*, 16 (1), forthcoming.

___ (2005b). "Hockey Sticks, Principal Components and Spurious Significance," *Geophysical Research Letters*, forthcoming.

McKitrick, Ross (2004). "The Mann et al. Northern Hemisphere 'Hockey Stick' Climate Index: A Tale of Due Dilgence," mimeo, Department of Economics, University of Guelph, October.

Mirowski, Philip and Steven Sklivas. (1991) "Why Econometricians Don't Replicate (Although They Do Reproduce)." *Review of Political Economy*, 3:2, 146-163.

Mittelstaedt, Robert A. and Thomas S. Zorn (1984). "Econometric Replication: Lessons from the Experimental Sciences," *Quarterly Journal of Business and Economics*, 23:1.

Neuliep, James W., ed. (1991), *Replication Research in the Social Sciences* (Sage Publications).

Racine, Jeff (2001), "On the Nonlinear Predictability of Stock Returns Using Financial and Economic Variables," *Journal of Business and Economic Statistics* 19(3), 380-382

Ray, J.L. and B. Valeriano (2003). "Barriers to Replication in Systematic Empirical Research on World Politics," Symposium on Replication in International Studies Research," *International Studies Perspectives*, 4:1, pp. 79-85, February.

Stanley, T.D. (2001). "Wheat from Chaff: Meta-Analysis As Quantitative Literature Review," *Journal of Economic Perspectives,* 15(3), Summer, pp. 131-150.

Stokes, Houston H. (2004), "On the Advantage of Using Two or More Econometric Software Systems to Solve the Same Problem," *Journal of Economic and Social Measurement* 29(1-3), 307-320

Tauchen, George (1993). "Remarks on My Term at the *JBES*," *Journal of Business and Economic Statistics*, October, 11(4), pp. 428-31.

Vinod, H. D. (2001). "Care and Feeding of Reproducible Econometrics," *Journal of Econometrics* 100, 87-88.

Wasko, Molly McLure and Samer Faraj (2005). "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly*, 29(1), March, pp. 35–57.

Zeileis, Achim and Christian Kleiber (2004), "Validating Multiple Structural Change Models – A Case Study," manuscript, May. <http://www.sfb475.uni-dortmund.de/berichte/tr34-04.pdf>

Zellner, A. and H. Thornber (1966), "Computational Accuracy and Estimation of Simultaneous Equation Econometric Models," *Econometrica* 34(3), 727-729