A Comment on the Role of Professional Journals in Facilitating Data Access

Douglas L. Young

Author's note *The scientific community is increasingly concerned about the availability and accuracy of data and procedures underlying published scientific research The Board and editors of* The Journal of Agricultural Economics Research *(JAER) recently decided against requiring its authors to furnish access to ready-to-use data in published articles, but they encourage such access In May 1990, the* American Journal of Agricultural Economics *urged authors to make their procedures and data available to other researchers for purposes of replication but stopped short of requiring such sharing*

I favor open access to data and methodology used in published scientific research The *Western Journal of Agricultural Economics* (WJAE), on which I have served as co-editor for the past 3 years enforces a data access policy The WJAE experience may be of particular interest to the JAER audience because USDA employees who submit articles to WJAE and who have used confidential government data have been, to date, the only authors encountering difficulty with the WJAE policy

Only a handful of economics professional journals, including the *Journal of Applied Econometrics* (JAE), the *Journal of Econometrics*, the *Journal of Money, Credit, and Banking* (JMCB), and WJAE, require access to data and methodology prior to reviewing submitted articles Given the integral role of careful documentation and openness in the scientific method, it is surprising that these requirements have been occasionally controversial for the few journals that have enacted them After all, scientific verification or refutation of research results is impossible if the methodology and data are unavailable or too vaguely documented to reconstruct

Careful documentation and access to data are important because science is a cumulative process The process is most efficient when researchers can confirm that the earlier results upon which they are basing their inquiries are sound A sound base of verifiable research should reduce the number of false starts and dead-end hypotheses Insistence on replication is an important safeguard for the foundations of any science (*1*, *5*) [1]

Concern about possible errors in published research results due to errors in the data and/or methodology is well founded In a landmark study published in the *American Economic Review*, Dewald, Thursby and Anderson found that errors in published research articles due to data transcription errors, computer programming mistakes, and related "data problems" were relatively common (*3*) This conclusion was based upon an ambitious project to replicate empirical results of several articles published in JMCB

Openness in the sharing of data and methodology also fosters efficiency in the use of scarce and costly data At the time the research is conducted, researchers often can store the data and methodology for future use at a low marginal cost

Insistence by professional journals that authors provide access to their data can help encourage flexibility and openness in government agencies that collect data Data access policies provide motivation for working out compromise solutions to the conflicting but mutually important requirements of confidentiality and access with respect to government data

WJAE Policy

WJAE is the only agricultural economics journal in North America that makes consent to data access and documentation a formal condition for entering the review process Since January 1, 1988, authors submitting manuscripts to WJAE must consent to document fully their procedures and "to make data used available at cost upon request for five years "

On the whole, the policy has worked smoothly Most authors have agreed routinely to the data and estimation policy Since July 1988, authors of only three manuscripts were unable to comply at first All three cases involved USDA authors who had used confidential government data One of these cases, which involved Department of Commerce agricultural census data as well, was resolved by a compromise The other two, involving USDA survey data, were eventually withdrawn due to failure to reach a timely compromise, but ongoing negotiations may yet achieve a long-term solution

The government is justified in legally restricting access to some observations from private citizens The key component of the successful compromise with WJAE asks an authorized government employee acting as a third party to obtain requested aggregated statistical results from the confidential data set These results are provided to the requester through the editors of WJAE This compromise meets the spirit of

Young is a professor in the Department of Agricultural Economics Washington State University Pullman and co-editor of the *Western Journal of Agricultural Economics*

[1] Italicized numbers in parentheses cite sources listed in the References section at the end of this article

WJAE's data access policy It complies with census and USDA confidentiality requirements that forbid divulging information on individual respondents or even on small groups of respondents (*2*) The WJAE compromise essentially substitutes "reasonable access to aggregated results' for "access to the raw data "

USDA's initial proposal for providing access to the data in the two as yet unsuccessful cases was burdensome USDA proposed an extensive application and internal review process for anyone requesting access to confidential USDA data used in the potential WJAE articles Once the application cleared that hurdle, USDA would have required the applicant to obtain the data personally in Washington, DC

This proposal was rejected by the Western Agricultural Economics Association because both the uncertainty of access and the formidable practical barriers were judged to violate the spirit of the access policy Subsequent discussions between USDA officials and WJAE editors have focused on resolving these two concerns and a third concern of USDA that a reasonable limit be placed on its staff workload

A data access policy like WJAE's forces editors to consider what they would do if an author who has consented to make data available upon request fails to do so Our first recourse would be to use editorial persuasion Our second recourse would be simply to include a brief statement in the Journal indicating that the author had not made data available in compliance with an earlier agreement

So, how has the WJAE data/methodology access policy affected the workload of editors, authors, and government agencies that collect data? In private discussions, this question has been a major concern of agency officials who administer confidential government data sets To date this concern has been moot WJAE has not received a single formal request for data from published articles in the past 3 years

In Conclusion

A policy of guaranteed access to data and methodology for articles published in WJAE has not spawned a rush of requests for data Why? Possibly, because science does not reward or encourage replication (*3*, p 587)

> Thomas Kuhn (1970) emphasized that replication—however valuable in the search for knowledge—does not fit within the "puzzle-solving" paradigm which defines the reward structure in scientific research Scientific and professional laurels are not awarded for replicating another scientist's findings Further, a researcher undertaking a replication may be viewed as lacking imagination and creativity or of being unable to allocate his time wisely among competing research projects In addition, replications may be interpreted as reflecting a lack of trust in another scientist's integrity and ability, as a critique of the scientist's findings, or as a personal dispute between researchers Finally ambiguities and/or errors in the documentation in the original research may leave the researcher unable to distinguish between errors in the replication and in the original study Months of effort may yield the replicater only inconclusive results regarding the validity of the original study, and thus no foundation for his future research in the area These circumstances nurture a natural reluctance to undertake replication studies

Some observers might argue that the lack of incentives for replication points up the irrelevance of data/methodology access policies in professional journals I argue that it indicates the opposite By encouraging preservation of the utilized data and estimation procedures, these policies help dismantle one important practical barrier to replication Such policies can also improve the quality of research at the outset Dewald Thursby, and Anderson (*3*, p 589) provide evidence " that the existence of a requirement that authors submit to the journal their [computer] programs and data along with each manuscript would significantly reduce the frequency and magnitude of errors [T]he very process of authors compiling their programs and data for submission revealed to them ambiguities, errors, and oversights which otherwise would be undetected " They convincingly argue that authors can more efficiently and accurately assemble their data and methodology when they submit an article than they can several months or years later Indeed, approximately a third of the authors in their sample had "lost or destroyed" their data for articles submitted as recently as 6 months in the past JMCB and JAE ask authors to submit data sets and computer programs to the editorial office This is more stringent than the WJAE policy which requires authors to keep data to satisfy potential requests for up to 5 years

Dewald, Thursby, and Anderson also reported that a journal can provide a cost-effective central storage location for data sets and programs JMCB reports no unreasonable storage or logistics problem with its policy The JMCB data bank, furthermore, has been used in graduate classes where students have been assigned to replicate the results of published studies Using a journal as a third party for data access mitigates possible personal fears that might preclude a researcher from asking a published author for original data

Professional journals have a role in fostering access to estimation methodology and data used in published articles Given the current incentive structure, neither authors nor readers have much motivation to facilitate

replications or to undertake them If professional journals do not foster an environment conducive to scientific replication, who will?

References

1 Broad, W , and N Wade *Betrayers of Truth* New York Simon and Schuster, 1983

2 Caudill, C E "A Federal Agency Looks for Answers to Data Sharing/Confidentiality Issues " Unpublished paper U S Dept Agr , Nat Agr Stat Serv , Aug 1990

3 Dewald, G D , J G Thursby, and R G Anderson "Replication in Empirical Economics The Journal of Money, Credit, and Banking Project," *American Economic Review* Vol 76, 1986, pp 587-606

4 Kuhn, T S *The Structure of Scientific Resolutions* 2nd ed Chicago University of Chicago Press, 1970

5 Ladd, G "Thoughts on Building an Academic Career," *Western Journal of Agricultural Economics* Vol 16, 1991, pp 1-10