# **Fellow's Opinion: Econometrics, Data, and the World Wide Web**

by William A. Barnett[1]

Washington University in St. Louis

February 7, 1996

Forthcoming as a J. E. Fellow's article in the *Journal of Econometrics*

---

[1]This research was partially supported by NSF grant SES 9223557.

**Econometrics, Data, and the World Wide Web**

As is increasingly evident, the World Wide Web is producing an information revolution. The long run implications of this revolution for the field of econometrics may exceed our current expectations and hence are difficult to anticipate. Nevertheless, I would like to use this Fellow's Opinion article as a means of speculating about one such likely effect that I believe to be very positive and topical.

Lately the subject of governmental data quality has become visible and fashionable, even to the public and to Congress. Recent Congressional testimony has raised serious questions about the merits of the consumer price index (CPI) published by the Bureau of Labor Statistics, and technical issues (e.g., the cumulative upwards bias) regarding the index's infamous unchained, fixed market basket are now fashionable discourse for talk-show hosts. In addition, the press recently has published unusually harsh attacks on governmental data quality, as in the recent *Business Week* article (Mandel (1994, pp. 110-118)) containing the overstated but nevertheless provocative conclusion:

> "The economic statistics that the government issues every week should come with a warning sticker: User beware. In the midst of the greatest information explosion in history, the government is pumping out a stream of statistics that are nothing but myths and misinformation."

Congress is considering proposals to alter the manner in which data are produced and supplied, with some proposals involving major governmental reorganizations, such as the possible creation of a central data production agency, as exists in Canada and in many other countries. The purpose of that controversial proposal appears to be to eliminate duplication of effort across agencies and to minimize conflict of interests by removing the data creation responsibility from those agencies which also use that data in policy.[2]

To the degree that governmental data quality is questionable, applied econometricians should be very concerned, since bad or poor quality data can contaminate an empirical literature, as I believe has been the case in

---

[2]A conspicuous example is the creation of monetary aggregate data by the Federal Reserve, which is held responsible, at least sometimes, to Congress in terms of target ranges for those aggregates. A game theoretic model of this phenomenon would produce serious issues regarding monitoring and accountability. A stickier example of possible conflict of interests could be suggested by the fact that the pensions of federal government employees are indexed by the CPI, which is produced by a governmental agency.

the field of applied monetary economics for at least two decades.[3] Saving tax dollars may motivate Congressional and public concern, but econometricians have a different reason for concern, and that reason is the integrity of our science. While all of this recent notoriety, and the resulting proposals for governmental reform, may have an alarming appearance, I wish to suggest that the problem can and may be solved privately, whether or not the current noisy public actions result in governmental changes or reforms.[4]

Historically, economic data were gathered, aggregated, and supplied privately. In the nineteenth century, price indexes were produced and published by newspapers in the United States and England. Prior to the Second World War, much of the best economic data were provided by private sources, and until relatively recently, the National Bureau of Economic Research was the source of much of the best economic data.[5] Yet in recent decades, production and distribution of economic data have become heavily centralized in the federal government. I suggest that there are three reasons: (1) a comparative advantage of government in collection of disaggregated national economic data, (2) a comparative advantage of government in the distribution of aggregated national economic data, and (3) conflict of interests. The third reason should motivate competition from private data sources, and the second reason I argue is completely eliminated by the Web. Only the first reason remains.

---

[3]See, e.g., Belongia (1993), Swofford and Whitney (1987), and Barnett, Fisher, and Serletis (1992) for some relevant empirical evidence regarding the seriousness of the matter. Simple sum monetary aggregation has made no sense since monetary aggregates began to include assets providing a positive rate of return, such as NOW accounts. Simple sum aggregation, in aggregation theory, is correct only over perfect substitutes (one can add apples to apples, but not apples to oranges). If goods are perfect substitutes and have different prices, a corner solution will result. If in fact the simple sum monetary aggregates are correct (i.e. the component assets are perfect substitutes) and if the components yield different rates of return (currency versus certificates of deposit, etc.), then everyone must be holding only the highest yielding asset. Advocates of simple sum monetary aggregation therefore must believe that the appearance of the existence of currency and demand deposits in the economy is an illusion. Those assets must have been wiped out by a corner solution decades ago.

In addition to the contaminating effects of the simple sum monetary aggregates on empirical research, there also is the contaminating effect of the opportunity cost variables appearing in that published research. If the true monetary aggregate were a simple sum, then the dual user cost price would be the Leontief index, which would be the maximum rate of return on any component asset, since only that component asset would be held. If multiple component assets are held, then not only is the simple sum the wrong monetary aggregate, but no single interest rate is the dual user cost price and hence no one interest rate is the opportunity cost of monetary services. The correct measure is the user cost index that is dual to the quantity index aggregating over the component quantities. See Barnett (1980).

[4]In fact I am less than convinced that Canadian data are of better quality than US data, and hence I do not view the proposals for change in Washington to be the ultimate solution for econometricians needing scientifically valid data that are coherent with economic theory. Regarding that need for coherence (the so called "Barnett critique"), see Chrystal and MacDonald (1994).

[5]See, e.g., Kuznets (1961) and Friedman and Schwartz(1970) regarding some of the widely used NBER data, and Dewhurst and Associates (1955) and Kravis, Heston, and Summers (1982) regarding many of the other private sources.

**Reason 1:** The first reason is likely always to remain. Consider, for example, commercial bank data acquired by the Federal Reserve. Some of that data ("the reserves file") are acquired from commercial banks under a confidentiality agreement intended to protect the competitive position of individual banks. Only after aggregation is the Federal Reserve permitted to supply any of that data to the public. Other examples would include the decennial census and other data sources that could not be aquired without potential penalties of law. It seems unlikely that we shall see comparable private contractual agreements evolving such that collection of disaggregated data by private sources would grow dramatically. Hence we are likely to continue depending upon governmental agencies for disaggregated data collection.

**Reason 2:** Distribution of data through the Web is increasingly becoming the preferred method, and even governmental data sources are becoming available by that means. But private data producers can distribute their data in the same manner. For example, experimental laboratories can supply to the public the data that they generate, after they have published their research results. The same is true for Monte Carlo data. Since much data are available from the government in relatively disaggregated form, private researchers wishing to aggregate in manners well suited for research can do so and supply the data on the Web in direct "competition" with the disreputable fixed base Paasche and Laspeyres indexes commonly supplied by governmental agencies.

**Reason 3:** As suggested above, there are instances in which conflicts of interest exist between governmental production of data and governmental use of data. There also are conflicting interests among non-governmental users of governmental data. For example, the CPI is used for many different purposes in the private sector of the economy. Only one of those purposes is as data in econometric research. To the degree that econometricians have reason to want price indexes and quantity indexes constructed in specific ways, such as chained Fisher ideal, Törnqvist-Divisia, or Frisch indexes, why should we be dependent upon the fixed base Laspeyres and Paasche indexes---or worse yet the simple sum aggregates---often supplied by government and frequently used by the private and public sectors for very politically sensitive purposes? Such purposes include indexing wages in labor contracts, indexing Social Security benefits, determining cost of living adjustments for federal government employees and retirees, targeting monetary policy, and indexing federal tax rates? The implications of some of those forms of indexing for the deficit are widely known to the public at present.

There is a significant principal-agent problem in econometricians' relying on governments for aggregate data to be used in models. Expecting or requiring government to aggregate and supply data in all of the forms that would be useful to all of the potential users of that data is unrealistic. Pressures on government from the conflicting needs of competing possible users can result in aggregated data that may be best for no one. The truth is that there is no method of data construction that is best for all possible users, and the econometrics profession is only one among the many potential users.[6]

It is my observation that the Web, by eliminating Reason 2 above, is already in the process of solving the problem. Cases in point are the Web sites recently put on-line at the University of Mississippi and at the St. Louis Federal Reserve Bank and the online version of the Penn World Tables.[7] The Mississippi site supplies an international database of Divisia monetary aggregates, component asset quantities, and dual user costs. Much of those data come from economists on the staffs of those central banks; and for purposes of econometric research, those data are much to be preferred to the simple-sum (shudder) data supplied officially by those central banks.[8] The St. Louis Federal Reserve database, FRED (Federal Reserve Economic Data), supplies analogous historical data on the Web for the United States "unofficially."[9] At present, the data being supplied on the Web are accessible at no cost, but it is possible in principle for subscription rates to be charged, as is being done with some of the new on-line journals. Hence research centers that might contemplate contributing to this evolving new source of economic data could cover their costs by charging for access to their data on the Web.

In my opinion, private competition with government in the production and distribution of aggregated national economic data, along with expected online availability of Monte Carlo and experimental data, is very good news for applied econometricians. As a means of encouraging that competition and hopefully the resulting growth in high quality data availability, I have put on-line a resource Web page, which has as its sole purpose providing

[6]An illustration of the sometimes poor connection between governmental data production methods and the needs of econometricians is the lack of any link between the data supplied by the Commerce Department and the earlier historical data supplied by Kuznets for years going back to the late nineteenth century. If the Commerce Department supplied their data in a form aggregated into categories that would link with Kuznets' published data, this country would have over a century of data on consumption allocation patterns. This capability, which easily could have been provided to researchers by the Commerce Department, soon will become available on the Web from a private source (i.e. from two of my students), with support from the Federal Reserve Bank of St. Louis.
[7]For the origins of the important international database, the Penn World Tables, see Kravis, Heston, and Summers (1982).
[8]An exception is the Bank of England, which is itself publishing Divisia monetary aggregate data.
[9]The historical United States Divisia data available from the Federal Reserve Bank of St. Louis are the subject of current research intended to permit regular published updates by that bank in the future.

links for econometricians to unusually valuable Web pages, especially those supplying high quality data produced outside of official governmental channels.[10] Unlike other resource Web pages, this one, instead of being comprehensive, is selective. I encourage correspondence regarding links that should be added to those already included on the page. Access to my Web resource page is, and will remain, free, and can be located on the Web at:

**http://econwpa.wustl.edu:80/~barnett/**

---

REFERENCES


Barnett, William A. (1980), "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory," *Journal of Econometrics*, Summer 1980, 14(1), pp. 11-48.

Barnett, William A., Fisher, Douglas, and Apostolos Serletis (1992), "Consumer Theory and the Demand for Money," *Journal of Economic Literature*, vol 30, Dec. 1992, pp. 2086-2119.

Barnett, William A., Gallant, A. Ronald, Hinich, Melvin J., Jungeilges, Jochen A., Kaplan, Daniel T., and Mark J. Jensen (1995), "A Single-Blind Controlled Competition among Tests for Nonlinearity and Chaos," *Journal of Econometrics*, submitted.

Belongia, Michael (1993), "Measurement Matters: Recent Results from Monetary Economics Re-examined," *Journal of Political Economy*, forthcoming.

Chrystal, K. Alec and Ronald MacDonald (1994), "Empirical Evidence on the Recent Behavior and Usefulness of Simple-Sum and Weighted Measures of the Money Stock," Federal Reserve Bank of St. Louis *Review*, March/April, pp. 73-109.

Dewhurst and Associates (1955), *America's Needs and Resources*, Twentieth Century Fund, New York.
Mandel, Michael J., "The Real Truth about the Economy: Are Government Statistics so much Pulp Fiction? Take a Look," *Business Week*, cover story, November 7, 1994, pp. 110-118.

Friedman, Milton and Anna Schwartz (1970), *Monetary Statistics of the United States: Estimates, Sources, Methods* (Columbia University Press, NY).

Kravis, Irving B., Heston, Alan W., and Robert Summers (1982), *World Product and Income: International Comparisons of Real Gross Product*, Johns Hopkins Press, Baltimore.

Kuznets, Simon (1961), *Capital in the American Economy* (National Bureau of Economic Research, NY).

Swofford, James L. and Gerald A. Whitney (1987), "Nonparametric Tests of Utility Maximization and Weak Separability for Consumption, Leisuure, and Money," *Review of Economics and Statistics*, Aug. 1987, vol 69, pp. 458-464.


---

[10]Private competition on the Web, using innovative modern approaches to data aggregation, will also likely result in improved data from government. In fact there is growing evidence that some of the inertia on the public side is decreasing. For example, the Commerce Department's recent release of chain-weighted gross domestic product data is encouraging. One, nevertheless, might wonder why it took 50 years, and why finally now? Another interesting development has been the recent decision of the Commerce Department to turn over computation of the Index of Leading Economic Indicators to the New York based Conference Board, which is a private research group.