COMMUNICATING UNCERTAINTY IN OFFICIAL ECONOMIC STATISTICS

(forthcoming: *Journal of Economic Literature*)

Charles F. Manski

Department of Economics & Institute for Policy Research
Northwestern University

"Perhaps the greatest step forward that can be taken, even at short notice, is to insist that economic statistics be only published together with an estimate of their error. Even if only roughly estimated, this would produce a wholesome effect. Makers and users of economic statistics must both refrain from making claims and demands that cannot be supported scientifically. The publication of error estimates would have a profound influence on the whole situation."

Oscar Morgenstern, *On the Accuracy of Economic Observations*, 1963, pp. 304-305.

Government statistical agencies report official statistics as point estimates.

Publications documenting data and methods acknowledge that estimates are subject to sampling and nonsampling error, but they do not quantify error magnitudes.

News releases present estimates with little mention of error.

Examples include the employment, household income, and GDP statistics reported by the BLS, Census Bureau, and BEA.

Reporting official statistics as point estimates manifests the tendency of policy analysts to project *incredible certitude*.

Users of statistics may misinterpret the information that the statistics provide.

Some may take them at face value.

Others may conjecture error directions and magnitudes.

Agencies could mitigate misinterpretation if they were to measure and report the uncertainty in official statistics.

Why is it important to communicate uncertainty?

Governments and private entities use the statistics as conditioning information when making important decisions.

The quality of decisions may suffer if decision makers incorrectly believe the statistics to be accurate or incorrectly conjecture error magnitudes.

For example, a central bank may mis-evaluate the economy and consequently set inappropriate monetary policy.

Agency communication of uncertainty would enable decision makers to better understand the information actually available regarding key economic variables.

Sampling and Non-Sampling Uncertainty

Agencies could use established principles to report sampling error in statistics based on survey data.

It is more challenging to measure nonsampling error.

Yet good-faith efforts would be more informative than reporting official statistics as if they are truths.

Transitory Statistical Uncertainty

arises because data collection takes time.

Agencies may release a preliminary statistic with incomplete data and revise as new data arrives.

Uncertainty diminishes as data accumulates.

<u>Example</u>: BEA initial measurement of GDP and revision of the estimate as new data arrives.

Permanent Statistical Uncertainty

arises from lasting incompleteness or inadequacy of data collection.

In surveys, sources include finite sample size, nonresponse and misreporting.

Example: nonresponse to employment and income questions in the Current Population Survey.

Conceptual Uncertainty

arises from incomplete understanding of the information that official statistics provide about economic concepts or from lack of clarity in the concepts themselves.

Conceptual uncertainty concerns the interpretation of statistics rather than their magnitudes.

Example: seasonal adjustment of statistics.

Transitory Uncertainty in National Income Accounts

The BEA reports multiple vintages of quarterly GDP estimates:

* An *advance* estimate combines data available one month after the end of a quarter with trend extrapolations.
* *Second* and *third* estimates are released after two and three months.
* A *first annual* estimate is released in the summer, based on more extensive data collected annually.
* There are subsequent annual and five-year revisions.

BEA reports GDP estimates without quantitative measures of uncertainty.

The Substantive Significance of Revisions

BEA analysts provide an upbeat perspective on the accuracy of GDP statistics.

Landefeld, Seskin, and Fraumeni (2008):

> "In terms of international comparisons, the U.S. national accounts meet or exceed internationally accepted levels of accuracy and comparability. The U.S. real GDP estimates appear to be at least as accurate—based on a comparison of GDP revisions across countries—as the corresponding estimates from other major developed countries."

Croushore (2011) offers a more cautionary perspective:

> "Until recently, macroeconomists assumed that data revisions were small and random and thus had no effect on structural modeling, policy analysis, or forecasting. But realtime research has shown that this assumption is false and that data revisions matter in many unexpected ways."

> "If monetary policy depends on short term growth rates, then clearly policy mistakes could be made if the central bank does not account for data uncertainty."

Measurement of Transitory Uncertainty

Communication of the transitory uncertainty of GDP estimates should be relatively easy to accomplish.

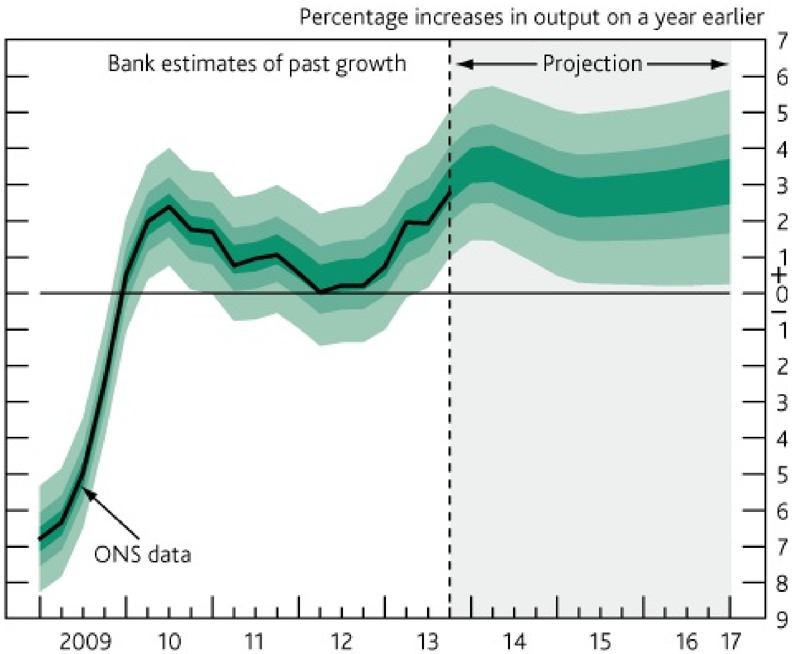
The historical record of BEA revisions has been made accessible in two *real-time* data sets maintained by the Philadelphia and St. Louis Federal Reserve Banks.

Measurement of transitory uncertainty in GDP estimates is straightforward if one finds it credible to assume that the revision process is time-stationary.

Then historical estimates of the magnitudes of revisions can credibly be extrapolated to measure the uncertainty of future revisions.

A precedent exists in the fan charts of the Bank of England.

Figure 1: February 2014 UK GDP
Fan Chart
(Source: Bank of England)



Permanent Uncertainty: Nonresponse in Sample Surveys

Census Reporting of Income Statistics

Each year the Census Bureau reports statistics on the household income distribution based on data collected in a supplement to the CPS. There is considerable nonresponse to the income questions.

During 2002-2012, 7 to 9 percent of the sampled households yielded no income data due to unit nonresponse and 41 to 47 percent of the interviewed households yielded incomplete income data due to item nonresponse.

Nonresponse Imputations and Weights

Statistical agencies routinely assume that nonresponse is random conditional on specified observed covariates.

These assumptions are implemented as weights for unit nonresponse and imputations for item nonresponse.

The Census Bureau applies *hot-deck* imputations.

CPS documentation of hot-deck imputation offers no evidence that the method yields a distribution for missing data that is close to the actual distribution.

Another Census document describing the American Housing Survey is revealing:

> "Some people refuse the interview or do not know the answers. When the entire interview is missing, other similar interviews represent the missing ones . . . . For most missing answers, an answer from a similar household is copied. The Census Bureau does not know how close the imputed values are to the actual values."

Indeed, lack of knowledge of the closeness of imputed values to actual ones is common.

Measurement of Uncertainty due to Nonresponse

Econometric research has shown how to measure uncertainty due to nonresponse without making assumptions about the nature of the missing data.

One contemplates all values that the missing data can take. Then the data yield interval estimates of official statistics.

The literature derives these intervals for population means, quantiles, and other parameters. The literature shows how to form confidence intervals that jointly measure sampling and nonresponse error.

Manski (1989, 1994, 2003), Horowitz & Manski (1998, 2000), Imbens and Manski (2004).

Interval estimates of official statistics that place no assumptions on the values of missing data are credible, easy to understand, and simple to compute.

One might therefore think that it would be standard practice for government statistical agencies to report them.

Official statistics are not reported this way.

It is sometimes said that such interval estimates are "too wide to be informative."

Nevertheless, statistical agencies should report them.

Even when wide, interval estimates making no assumptions on nonresponse are valuable for two reasons.

1. They are maximally credible.

2. They make explicit the fundamental role that assumptions play in inferential methods that yield tighter findings.

Wide bounds reflect real uncertainties that cannot be washed away by assumptions lacking credibility.

The above does not imply that statistical agencies should refrain from making assumptions about nonresponse.

Interval estimates making no assumptions may be excessively conservative if agency analysts have some understanding of the nature of nonresponse.

There is much middle ground between interval estimation with no assumptions and point estimation assuming that nonresponse is conditionally random.

The middle ground obtains interval estimates using assumptions that may include random nonresponse as one among various possibilities.

Manski (2015) poses some alternatives that agencies may want to consider.

Conceptual Uncertainty: Seasonal Adjustment of Official Statistics

Viewed from a sufficiently high altitude, the purpose of seasonal adjustment appears straightforward to explain.

It is less clear from ground level how one should perform seasonal adjustment.

The X-12-ARIMA method was developed by Census.

X-12 may be a sophisticated and successful algorithm for seasonal adjustment.

Or it may be an unfathomable black box containing a complex set of operations that lack economic foundation.

Wright (2014) notes the difficulty of understanding X-12:

> "Most academics treat seasonal adjustment as a very mundane job, rumored to be undertaken by hobbits living in holes in the ground. I believe that this is a terrible mistake, but one in which the statistical agencies share at least a little of the blame.
>
> "Seasonal adjustment is extraordinarily consequential."

Measurement of Uncertainty Associated with Seasonal Adjustment

There now exists no clearly appropriate way to measure the uncertainty associated with seasonal adjustment.

X-12 is a standalone algorithm, not a method based on a theory of the economy.

It is not obvious how to evaluate the extent to which it accomplishes the objective of removing the influences of predictable seasonal patterns.

Principled ways to evaluate uncertainty may open up if agencies were to use a seasonal adjustment method derived from a well-specified model of the economy.

.

Conclusion

The NRC-CNSTAT publication *Principles and Practices for a Federal Statistical Agency* recommends adherence to various good practices. Practice 4 states

"A statistical agency should be open about the strengths and limitations of its data, taking as much care to understand and explain how its statistics may fall short of accuracy as it does to produce accurate data. Data releases from a statistical program should be accompanied by a full description of the purpose of the program; the methods and assumptions used for data collection, processing, and reporting; what is known and not known about the quality and relevance of the data; sufficient information for estimating variability in the data; appropriate methods for analysis that take account of variability and other sources of error; and the results of research on the methods and data."

The practice in the reporting of official statistics has been to acknowledge potential errors verbally but not quantitatively.

The news releases and technical documentation published by statistical agencies caution that point estimates of official statistics are subject to sampling and nonsampling errors.

Agency publications do not measure non-sampling errors.

Nor do agencies justify the ways that they produce point estimates.

I have suggested ways to measure transitory statistical uncertainty arising from incomplete data and permanent statistical uncertainty stemming from survey nonresponse.

I have called attention to the conceptual uncertainty in seasonal adjustment.

Statistical agencies would better inform policymakers and the public if they were to measure and communicate these and other significant uncertainties in official statistics.

An open question is how communication of uncertainty would affect policymaking and private decision making.

We have little understanding of the ways that users of statistics interpret them.

Some may take the statistics at face value. Others may conjecture the directions and magnitudes of errors.

We know essentially nothing about how decision making would change if statistical agencies were to communicate uncertainty regularly and transparently.

I urge behavioral and social scientists to initiate empirical studies that would shed light on this subject.