NBER WORKING PAPER SERIES

COMMUNICATING UNCERTAINTY IN OFFICIAL ECONOMIC STATISTICS

Charles F. Manski

Working Paper 20098
http://www.nber.org/papers/w20098

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2014

Communicating Uncertainty in Official Economic Statistics
Charles F. Manski
NBER Working Paper No. 20098
May 2014
JEL No. C82,E01,I32

ABSTRACT

Federal statistical agencies in the United States and analogous agencies elsewhere commonly report official economic statistics as point estimates, without accompanying measures of error. Users of the statistics may incorrectly view them as error-free or may incorrectly conjecture error magnitudes. This paper discusses strategies to mitigate misinterpretation of official statistics by communicating uncertainty to the public. Sampling error can be measured using established statistical principles. The challenge is to satisfactorily measure the various forms of non-sampling error. I find it useful to distinguish transitory statistical uncertainty, permanent statistical uncertainty, and conceptual uncertainty. I illustrate how each arises as the Bureau of Economic Analysis periodically revises GDP estimates, the Census Bureau generates household income statistics from surveys with non-response, and the Bureau of Labor Statistics seasonally adjusts employment statistics.

Charles F. Manski
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
cfmanski@northwestern.edu

1. Introduction

Government statistical agencies commonly report official economic statistics as point estimates, without accompanying measures of error. Agency publications documenting the data and methods used to produce official statistics may acknowledge verbally that the estimates are subject to sampling and nonsampling error, but the publications typically do not quantify error magnitudes. News releases that communicate official statistics to the public present the estimates with little if any mention of the possibility of error. Some prominent American examples include the reporting of GDP, household income, and employment statistics by the Bureau of Economic Analysis (BEA), the Census Bureau, and the Bureau of Labor Statistics (BLS).

Reporting official statistics as point estimates manifests a common tendency of policy analysts to project *incredible certitude*, encouraging policy makers and the public to believe that errors are small and inconsequential (Manski, 2011, 2013a). In the absence of agency guidance, persons who understand that official statistics are subject to error must fend for themselves and conjecture the error magnitudes. Thus, users of official statistics may misinterpret the information that the statistics provide.

Statistical agencies could mitigate misinterpretation of official statistics if they were to measure uncertainty and report it in their news releases and technical publications. Why is it important to communicate uncertainty in official statistics? A broad reason is that governments, firms, and individuals use the statistics when making numerous important decisions. The quality of decisions may suffer if decision makers incorrectly take reported statistics at face value or incorrectly conjecture error magnitudes. For example, a central bank monitoring official statistics on GDP growth, inflation, and unemployment may mis-evaluate the status of 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.

Using established statistical principles, agencies could report sampling error for levels and temporal

changes in important statistics based on survey data. For example, the BLS could report confidence intervals for key employment statistics, such as the level and month-to-month changes in the unemployment rate. The Census Bureau could do likewise for key income statistics, such as the poverty rate and median household income.

It is more challenging for agencies to report nonsampling errors for official statistics. There are many sources of such errors and there has been no consensus about how to measure them. Yet these facts do not justify ignoring nonsampling error. Having agency analytical staffs make good-faith efforts to measure nonsampling error would be more informative than having agencies report official statistics as if they are truths.

Beyond the traditional distinction between sampling and nonsampling error, I find it useful to distinguish transitory statistical uncertainty, permanent statistical uncertainty, and conceptual uncertainty. Transitory statistical uncertainty arises because data collection takes time. Agencies sometimes release a preliminary estimate of an official statistic in an early stage of data collection and revise the estimate as new data arrives. Hence, uncertainty may be substantial early on but diminish as data accumulates.

Permanent statistical uncertainty arises from incompleteness or inadequacy of data collection that does not diminish with time. Sampling uncertainty stemming from the finite size of survey samples is permanent. So is uncertainty stemming from nonresponse and from the possibility that some respondents may provide inaccurate data.

Conceptual uncertainty arises from incomplete understanding of the information that official statistics provide about well-defined economic concepts or from lack of clarity in the concepts themselves. Thus, conceptual uncertainty concerns the interpretation of statistics rather than their magnitudes.

This paper discusses some prominent examples of these forms of uncertainty and considers how agencies might constructively communicate the uncertainty. To illustrate transitory statistical uncertainty, Section 2 describes BEA initial measurement of GDP and the ensuing process of revising the estimate as new

data arrives. Section 3 discusses permanent statistical uncertainty due to nonresponse in sample surveys, using nonresponse to employment and income questions in the Current Population Survey to illustrate. Section 4 discusses the conceptual uncertainty inherent in seasonal adjustment of economic statistics. Section 5 concludes.

2. Transitory Uncertainty: Revisions in National Income Accounts

2.1. Bureau of Economic Analysis Reporting of GDP

The Bureau of Economic Analysis of the U.S. Department of Commerce reports quarterly estimates of GDP. The BEA initially reports an *advance* estimate based on incomplete data available one month after the end of a quarter. It reports *second* and *third* estimates after two and three months, when additional data become available. In the summer of each year, when more extensive data collected on an annual rather than quarterly basis become available, BEA reports a *first annual* estimate and then revises it further in subsequent years. Every five years, the BEA re-evaluates its operational definition of GDP and accordingly makes yet further revisions to the historical GDP record. I will not discuss the five-year revisions here because they generate conceptual rather than statistical uncertainty.

An article describing the measurement of GDP explains the reasons for revisions as follows (Landefeld, Seskin, and Fraumeni, 2008, p. 194):

> "For the initial monthly estimates of quarterly GDP, data on about 25 percent of GDP—especially
> in the service sector—are not available, and so these sectors of the economy are estimated based on
> past trends and whatever related data are available. . . . . The initial monthly estimates of quarterly
> GDP based on these extrapolations are revised as more complete data become available. . . . . The

successive revisions can be significant, but the initial estimates provide a snapshot of economic activity much like the first few seconds of a Polaroid photograph in which an image is fuzzy, but as the developing process continues, the details become clearer."

Although this passage recognizes that initial quarterly estimates of GDP growth are subject to error, BEA practice has been to report these estimates without providing quantitative measures of uncertainty.

For example, a November 29, 2012 news release compared advance and second estimates, stating (Bureau of Economic Analysis, 2012):

"Real gross domestic product . . . . increased at an annual rate of 2.7 percent in the third quarter of 2012. . . . . The GDP estimate released today is based on more complete source data than were available for the 'advance' estimate issued last month. In the advance estimate, the increase in real GDP was 2.0 percent."

Thus, the second quarterly estimate of annual GDP growth differed from the advance estimate by 0.7 percentage points. While the news release observed that the advance estimate was based on incomplete data, it did not acknowledge that the second estimate was likewise based on incomplete data and would be further revised a month later and then annually for several years.

How large do the revisions to the BEA estimates tend to be? The 0.7 percentage point revision quoted above is close to the mean absolute revision (MAR) from month to month. Fixler, Greenaway-McGrevy, and Grimm (2011) report that the MAR from the advance estimates to the second estimates of real GDP is 0.5 percentage points, that from the advance estimates to the third estimates is 0.6 percentage points, and that from the second to the third estimates is 0.3 percentage points. Considering the period 1983–2009, they report that the overall MARs to the advance, second, and third quarterly estimates (comparing these estimates with the latest available for the relevant quarter) were 1.31, 1.29, and 1.32 percentage points. Observing that the magnitude of the revisions tends not to diminish with time, despite the availability of more data when forming the second and third estimates, the authors state (p. 12): "The lack of declines in the

MARs of GDP in successive vintages of current quarterly estimates is a phenomenon that has been noted in nearly all of BEA's analyses of revisions."

2.2. The Substantive Significance of Revisions

While the magnitudes of revisions to BEA estimates of GDP are straightforward to compute, the substantive significance of the revisions is a matter of interpretation. Fixler, Greenaway-McGrevy, and Grimm (2011) remark at their beginning of their article that (p. 9): "Economic policy decisions should not need to be reconsidered in the light of revisions to GDP estimates, and policymakers should be able to rely on the early estimates as accurate indicators of the state of the economy." They provide an upbeat absolute perspective in the conclusion to their article, stating (p. 30): "The estimates of GDP and GDI are accurate; the MARs for both measures are modestly above 1.0 percentage point." Landefeld, Seskin, and Fraumeni (2008) provide an upbeat comparative perspective, stating (p. 213) : 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 considerably more cautionary perspective, stating (p. 73): "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." To illustrate, he gives a notable example (p. 73):

"In January 2009, in the middle of the financial crisis that began in September 2008, the initial release of the national income accounts showed a decline in real gross domestic product (GDP) of 3.8 percent (at an annual rate) for the fourth quarter—a bad number for sure but not as bad as might be expected considering the damage caused by the financial meltdown. But one month later, the

GDP growth rate was revised down by 2.4 percentage points, showing a decline in real GDP of 6.2

percent and confirming that the U.S. economy was in the middle of the worst recession in over

twenty-five years. The 2.4 percentage point downward revision from the initial release to the first

revised number was the largest revision ever recorded for quarterly real GDP. Real-time data

analysis of the history of revisions of real GDP shows us that the largest revision came at a very

inopportune moment."

This example is an extreme case, but it provides a stark warning that BEA revisions to GDP may be

quite large and occur at times when vital policy decisions must be made. Leaving aside the singular event

of the financial crisis, I view the MARs reported by Fixler, Greenaway-McGrevy, and Grimm (2011) for the

period 1983–2009 as too large to warrant their upbeat conclusion that BEA quarterly estimates of GDP are

accurate. A statement made by Croushore (2011) seems more on the mark (p. 77): "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."

2.3. Measurement of Transitory Uncertainty

Informative communication of the transitory uncertainty of GDP estimates should be relatively easy

to accomplish. The historical record of BEA revisions has been made accessible for study in two "real-time"

data sets maintained by the Philadelphia and St. Louis Federal Reserve Banks. See Croushore (2011) for

details regarding these data sets and similar ones for other nations.

Forward-looking 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

particularly simple extrapolation would be to suppose that the historical overall MAR of 1.3 percentage

points reported by Fixler, Greenaway-McGrevy, and Grimm (2011) will persist going forward. More broadly, it may be credible to suppose that the empirical distribution of revisions will persist going forward.

More refined measures of uncertainty can be developed by studying how the magnitude and direction of historical revisions have varied with the state of the economy. Such refinements may be important to the extent that the nature of revisions tends to vary over the business cycle (see Croushore, 2011, Sec. 2.2). If there is reason to think that the historical pattern of variation of revisions with the state of the economy will persist into the future, then it would be appropriate to measure the transitory uncertainty of future GDP estimates in a manner that conditions on the state of the economy.

A notable precedent for probabilistic communication of the transitory uncertainty in GDP estimates is the periodic release of fan charts by the Bank of England, the Norges Bank in Norway, and some other central banks. I describe the British practice below. In the absence of BEA reporting of uncertainty in its GDP estimates, the Federal Reserve could usefully emulate this practice.

2.4. Reporting GDP Growth by the Bank of England

In the United Kingdom, the Office for National Statistics (ONS) reports quarterly estimates of GDP in a manner similar to the BEA, with no quantitative measurement of uncertainty despite the fact that the estimates are revised regularly. For example, the April 2014 edition of the monthly *Economic Review* of the ONS describes the most recent GDP revisions as follows (U. K. Office for National Statistics, 2014, p. 1): "The Quarterly National Accounts left Gross Domestic Product (GDP) growth in the final quarter of 2013 unrevised at 0.7%, but reduced annual growth in 2013 to 1.7%, largely as a consequence of lower than previously estimated household expenditure."

Unlike those in the United States, users of GDP estimates in the UK have ready access to a measure of uncertainty in the fan charts reported by the Bank of England in its monthly *Inflation Report*. Figure 1

reproduces a fan chart for annual GDP growth in the *February 2014 Inflation Report* (Bank of England, 2014). The part of the plot showing growth from late 2013 on is a probabilistic forecast that expresses the uncertainty of the Bank's Monetary Policy Committee regarding future GDP growth. The part of the plot showing growth in the period 2009 through mid 2013 is a probabilistic forecast that expresses uncertainty regarding the revisions that ONS will henceforth make to its estimates of past GDP. The Bank explains as follows (p. 7): "In the GDP fan chart, the distribution to the left of the vertical dashed line reflects the likelihood of revisions to the data over the past."

Observe that Figure 1 expresses considerable uncertainty about GDP growth in the period 2010–2013. Moreover, it expresses comparable uncertainty about growth in the recent past and the near future. Thus, the Bank judges that future ONS revisions to estimates of past GDP may be large in magnitude.

Figure 1: February 2014 UK GDP Fan Chart
(Source: Bank of England)
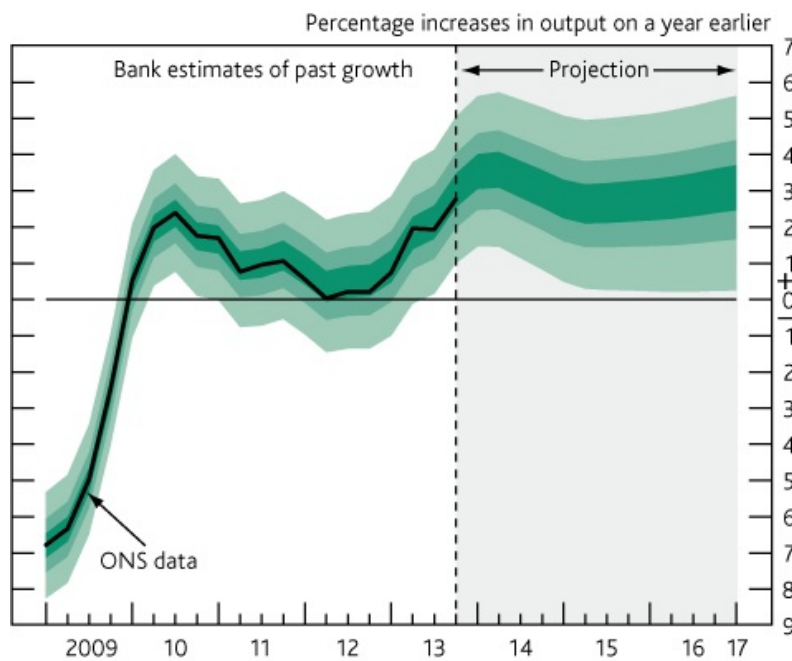
3. Permanent Uncertainty: Nonresponse in Sample Surveys

3.1. Nonresponse to Income and Employment Questions in the Current Population Survey

Nonresponse is common in the surveys used to compute important official statistics. Unit and item nonresponse may make key data missing for substantial fractions of the persons sampled. Nevertheless, the reporting of official statistics obscures the potential implications of nonresponse. Census reporting of income statistics and BLS reporting of the unemployment rate provide apt examples.

Census Reporting of Income Statistics

Each year the Census Bureau reports statistics on the household income distribution based on data collected in the Annual Social and Economic (ASEC) Supplement to the Current Population Survey (CPS). There is considerable nonresponse to the CPS income questions. During the period 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. Nevertheless, Census Bureau publications give the impression that national statistics on the income distribution are exact.

For example, in a news release issued September 12, 2012, the Census Bureau declared (U. S. Census Bureau, 2012A): "The nation's official poverty rate in 2011 was 15.0 percent, with 46.2 million people in poverty. After three consecutive years of increases, neither the poverty rate nor the number of people in poverty were statistically different from the 2010 estimates." Thus, the Census release provided point estimates, acknowledged but did not quantify sampling error, and did not mention the nonsampling error that may arise from nonresponse or misreporting.

I find it intriguing that the Census Bureau has not even provided quantitative measures of sampling error. A Census Bureau publication gives this explanation for the decision of the Bureau not to report

standard errors for income statistics (U. S. Census Bureau, 2012B, p. 7):

> "While it is possible to compute and present an estimate of the standard error based on the survey data for each estimate in a report, there are a number of reasons why this is not done. A presentation of the individual standard errors would be of limited use, since one could not possibly predict all of the combinations of results that may be of interest to data users. Additionally, data users have access to CPS microdata files, and it is impossible to compute in advance the standard error for every estimate one might obtain from those data sets."

This reasoning explains why the Bureau cannot measure sampling error for every logically possible application of the CPS data. It does not explain why the Bureau chooses not to report sampling error for the income statistics that it highlights in news releases.

Interestingly, a positive role model for quantitative measurement of sampling error in news releases of official statistics can be found elsewhere in the Census Bureau. Each month the Census Bureau and the Department of Housing and Urban Development jointly release statistics on new residential home sales in the previous month. Expression of uncertainty is prominent in these releases. For example, the one for March 2014 begins this way (U. S. Census Bureau, 2014):

> "Sales of new single-family houses in March 2014 were at a seasonally adjusted annual rate of 384,000, according to estimates released jointly today by the U.S. Census Bureau and the Department of Housing and Urban Development. This is 14.5 percent (±12.9%) below the revised February rate of 449,000 and is 13.3 percent (±9.9%) below the March 2013 estimate of 443,000."

The Explanatory Notes section of the release states "All ranges given for percent changes are 90-percent confidence intervals and account only for sampling variability." This transparent expression of sampling uncertainty could easily be emulated when the Census Bureau reports income statistics.

BLS Reporting of Employment Statistics

On the first Friday of each month, the BLS issues *The Employment Situation*, a monthly news release reporting official employment statistics for the previous month. For example, the BLS reported this on October 5, 2012 (U. S. Bureau of Labor Statistics, 2012): "The unemployment rate decreased to 7.8 percent in September, and total nonfarm payroll employment rose by 114,000." The unemployment-rate statistic is based on data on households sampled in the CPS. The one on nonfarm employment is based on data collected from employer establishments sampled in the Current Employment Statistics survey (CES).

The BLS monthly news release reports employment statistics as point estimates, without measures of potential error. A Technical Note issued with the news release contains a section on *Reliability of the estimates* that acknowledges the possible presence of errors, beginning with the statement "Statistics based on the household and establishment surveys are subject to both sampling and nonsampling error." The section describes the conventional use of standard errors and confidence intervals to measure sampling error, providing a few numerical illustrations.

The Technical Note then turns to nonsampling errors, stating that they "can occur for many reasons, including the failure to sample a segment of the population, inability to obtain information for all respondents in the sample, inability or unwillingness of respondents to provide correct information on a timely basis, mistakes made by respondents, and errors made in the collection or processing of the data." The Note does not indicate the magnitudes of the nonsampling errors that may be present in the employment statistics.

3.2. Nonresponse Imputations and Weights

To deal with survey nonresponse, statistical agencies use traditional but untenable assumptions, namely that nonresponse is random conditional on specified observed covariates. These assumptions are implemented as weights for unit nonresponse and imputations for item nonresponse.

In particular, the Census Bureau applies *hot-deck* imputations to the CPS and other surveys, describing the hot deck this way (U. S. Census Bureau, 2006, p. 9-2):

"This method assigns a missing value from a record with similar characteristics, which is the hot deck. Hot decks are defined by variables such as age, race, and sex. Other characteristics used in hot decks vary depending on the nature of the unanswered question. For instance, most labor force questions use age, race, sex, and occasionally another correlated labor force item such as full- or part-time status."

Thus the agency staff select a vector of covariates for which response is complete and compute the empirical distribution of the outcome of interest among sample members who have this covariate value and who report their outcomes. An outcome is imputed to a sample member with missing data by drawing a realization at random from the available empirical distribution.

The CPS documentation of hot-deck imputation offers no evidence that the method yields an outcome distribution for missing data that is close to the actual distribution of such outcomes. However, another Census Bureau document describing the American Housing Survey is revealing. The document states (U. S. Census Bureau, 2011):

"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.

A potentially fruitful way to assess the closeness of imputed values to actual ones is to enrich the available data by matching CPS respondents to administrative records that provide information on their income and employment. For example, Hokayem, Bollinger, and Ziliak (2014) obtain Social Security earnings records for CPS respondents. The authors caution that Social Security records exclude various

forms of income. Hence, they do not consider the records to provide the true values of missing data. Nevertheless, their joint analysis of the available CPS and Social Security data yields a cautionary conclusion (p. 5):

> "Our results suggest that assumption of missing at random, even conditional on known characteristics, is not valid in modern data. Hence any correction which assumes missing completely at random or missing at random, such as the hot deck procedure, is likely to be biased. We show that the ASEC underestimates the number of persons in poverty by an average of about 1.0 percentage point."

3.3. Measurement of Uncertainty due to Nonresponse

Modern econometric research on partial identification has shown how to measure uncertainty due to nonresponse without making any assumptions about the nature of the missing data. See Manski (1989, 1994, 2003) and Horowitz and Manski (1998, 2000) inter alia. A notable early precedent in the statistics literature was the Cochran, Mosteller, and Tukey (1954) consideration of the potential implications of nonresponse in the Kinsey survey of male sexual behavior. However, subsequent statistical research on analysis of surveys with nonresponse did not follow up and instead recommended generation of point estimates under assumptions of random nonresponse. See, for example, Little and Rubin (1987).

The basic idea in partial identification analysis, simply enough, is to contemplate all the values that the missing data might take. Doing so, the available data yield interval rather than point estimates of official statistics. The literature derives the forms of these intervals for population means, quantiles, and other parameters of possible interest. Econometricians have also shown how to form confidence intervals that jointly measure sampling error and potential nonresponse error (e.g., Horowitz and Manski, 2000; Imbens and Manski, 2004). Thus, we know how to measure uncertainty for official statistics with survey

nonresponse.

To illustrate, Manski (2013b) used ASEC data collected in 2002-2012 to form interval estimates of median household income and the fraction of families with income below the official poverty threshold in the years 2001-2011. I used monthly CPS data to form interval estimates of the unemployment rate in March of 2002-2012. I provided one set of estimates that take into account item nonresponse alone and another that recognizes unit response as well. The estimates show vividly that item nonresponse poses a huge potential problem for inference on the American income distribution, and that unit nonresponse exacerbates the problem. While item nonresponse is a relatively minor source of error for the unemployment rate, unit nonresponse is highly consequential.

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. I am, however, unaware of any official statistics reported this way.

In personal communications, colleagues have sometimes remarked that interval estimates obtained without assumptions on nonresponse are not reported because they are "too wide to be informative." Indeed, the illustrative empirical analysis of Manski (2013b) shows that, with current nonresponse rates, the intervals are quite wide. For example, the interval estimate for the family poverty rate in 2011 is [0.139, 0.339] if only item nonresponse is acknowledged and is [0.128, 0.390] if both item and unit nonreponse are recognized. The analogous interval estimates for the unemployment rate in March 2012 are [0.078, 0.090] and [0.071, 0.162].

Nevertheless, I would argue that the Census Bureau and the BLS should report such intervals. Even when wide, interval estimates obtained without assumptions on nonresponse are valuable for two reasons. First, they are maximally credible in the sense that they express all possible values of the statistic of interest, whatever values the missing data may take. Second, they make explicit the fundamental role that

assumptions play in inferential methods that yield tighter findings.   Wide bounds reflect real data uncertainties that cannot be washed away by assumptions lacking credibility.

The above argument does not imply, of course, that statistical agencies should entirely refrain from making assumptions about nonresponse. Interval estimates obtained with 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 the standard practice of point estimation assuming that nonresponse is conditionally random. The middle ground obtains interval estimates based on assumptions that may include random nonresponse as one among various possibilities. It is unlikely that any one middle-ground assumption will be appropriate in all settings. Manski (2013b) poses some alternatives that statistical agencies may want to consider.

4. Conceptual Uncertainty: Seasonal Adjustment of Official Statistics

I wrote in the Introduction that conceptual uncertainty arises from incomplete understanding of the information that official statistics provide about well-defined economic concepts or from lack of clarity in the concepts themselves. Section 2 mentioned one example, this being that the BEA re-evaluates its operational definition of GDP every five years and revises the historical GDP record accordingly.

Another example arises in BLS measurement of unemployment.   The BLS classifies CPS respondents as unemployed if they "do not have a job, have actively looked for work in the prior 4 weeks, and are currently available for work" (www.bls.gov/cps/cps_htgm.htm#unemployed).   Responses to a sequence of CPS questions are used to determine whether a person has "actively looked for work in the prior 4 weeks" and is "currently available for work." The notions of actively looking for work and being currently available for work are inherently vague to some degree, so there is resulting uncertainty about how

unemployment statistics should be interpreted.

4.1. Seasonal Adjustment in Principle and Practice

A particularly troublesome conceptual uncertainty arises in the conventional practice of seasonally adjusting official statistics, including quarterly GDP estimates and monthly unemployment rates. Viewed from a sufficiently high altitude, the purpose of seasonal adjustment appears straightforward to explain. However, it is much less clear from ground level how one should actually perform seasonal adjustment.

The BLS explains seasonal adjustment of employment statistics this way (U. S. Bureau of Labor Statistics (2001):

"What is seasonal adjustment? Seasonal adjustment is a statistical technique that attempts to measure and remove the influences of predictable seasonal patterns to reveal how employment and unemployment change from month to month. Over the course of a year, the size of the labor force, the levels of employment and unemployment, and other measures of labor market activity undergo fluctuations due to seasonal events including changes in weather, harvests, major holidays, and school schedules. Because these seasonal events follow a more or less regular pattern each year, their influence on statistical trends can be eliminated by seasonally adjusting the statistics from month to month. These seasonal adjustments make it easier to observe the cyclical, underlying trend, and other nonseasonal movements in the series."

The explanation is heuristically appealing but it does not specify how, in practice, one may "remove the influences of predictable seasonal patterns." Views on appropriate ways to perform seasonal adjustment have long varied among econometricians. See, for example, the exchange between Granger (1979) and Sims (1979) as well as the recent contribution of Wright (2013).

In practice, the BLS uses the X-12-ARIMA method, developed by the Census Bureau and described

in Findley *et al.* (1998). The X-12 method, along with its predecessor X-11 and successor X-13, may be a sophisticated and successful algorithm for seasonal adjustment. Or it may be an unfathomable black box containing a complex set of statistical operations that lack economic foundation. Wright (2013) eloquently expresses the difficulty of understanding X-12, writing (p. 2):

> "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. Statistical agencies emphasize SA data (and in some cases don't even publish NSA data), and while they generally document their seasonal adjustment process thoroughly, it is not always done in a way that facilitates replication, or encourages entry into this research area."

Wright's remark that statistical agencies sometimes do not publish non-seasonally-adjusted (NSA) data refers particularly to a decision of the BEA to stop publication of NSA GDP data. He comments on this as follows (p. 10): "It is very unfortunate that for the most basic measure of economic activity in the largest country in the world, researchers are effectively prevented from evaluating any difficulties associated with seasonal adjustment."

Understanding the practice of seasonal adjustment matters because, as Wright states (p. 1): "Seasonal adjustment is extraordinarily consequential." He gives this example concerning BLS reporting of estimates of non-farm payrolls (p. 1): "In terms of monthly changes, the average absolute difference between the seasonally adjusted (SA) and not-seasonally-adjusted (NSA) number is 660,000, which dwarfs the normal month-over-month variation in the SA data. All this implies that we should think very carefully about how seasonal adjustment is done."

4.2. Measurement of Uncertainty Associated with Seasonal Adjustment

There presently exists no clearly appropriate way to measure the conceptual uncertainty associated with seasonal adjustment. The Census Bureau's X-12 is an algorithm, not a method based on a well-specified dynamic theory of the economy. Hence, it is not obvious how to evaluate the extent to which it accomplishes the objective of removing the influences of predictable seasonal patterns. One might perhaps juxtapose X-12 with other seemingly reasonable algorithms, perform seasonal adjustment with each one, and view the range of resulting estimates as a measure of conceptual uncertainty.

More principled ways to evaluate uncertainty may open up if statistical agencies were to use a seasonal adjustment method derived from a well-specified model of the economy. One could then assess the sensitivity of seasonally adjusted estimates to variation in the parameters and the basic structure of the model.

A more radical departure from present practice would be to abandon seasonal adjustment and leave it to the users of official statistics to interpret unadjusted statistics as they choose. Publication of unadjusted statistics should be particularly valuable to users who want to make year-to-year rather than month-to-month comparison of statistics. Suppose, for example, that one wants to compare unemployment in March 2013 and March 2014. It is arguably more reasonable to compare the unadjusted estimates for these months than to compare the seasonally adjusted estimates. Comparison of unadjusted estimates for the same month each year (March in this case) sensibly removes the "influences of predictable seasonal patterns" that the BLS noted in its 2001 document. Moreover, it compares data actually collected in the two months of interest. In contrast, the seasonally adjusted estimates for March 2013 and March 2014 are comprised of data collected not only in these months but over a lengthy prior period.

5. Conclusion

The National Research Council (NRC) publication *Principles and Practices for a Federal Statistical Agency* recommends that agencies adhere to various good practices. Practice 4, titled "Openness About Sources and Limitations of the Data Provided," states this (National Research Council, 2013, p. 18):

"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."

As documented in this paper, federal statistical agencies typically pay only lip service to this practice in their reporting of official economic statistics. The norm has been to acknowledge potential errors verbally but not quantitatively. The news releases and technical documentation published by statistical agencies may caution readers that point estimates of official statistics are subject to sampling and nonsampling errors but agency publications typically do not measure the errors. They do not, in the words of the NRC, provide "appropriate methods for analysis that take account of variability and other sources of error."

Nor do agencies justify the ways that they use incomplete and imperfect data to produce the point estimates they report. I have called attention to several prevalent agency practices that lack justification: use of trend extrapolations to form advance GDP estimates, imputation of missing data in sample surveys, and use of the X-12 algorithm to seasonally adjust statistics. In these and other respects, agencies commonly do not make clear "what is known and not known about the quality and relevance of the data."

This paper has suggested ways to measure the transitory statistical uncertainty in estimates of

official statistics based on incomplete data and the permanent statistical uncertainty stemming from survey nonresponse. I have also 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 agency communication of uncertainty would affect policymaking and private decision making. We now have little understanding of the ways that users of official statistics interpret them. Some may mistakenly take the statistics at face value. Others may conjecture that the statistics are prone to errors of varying directions and magnitudes. 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.

References

Bank of England (2014), *Inflation Report Fan Charts February 2014*,
www.bankofengland.co.uk/publications/Documents/inflationreport/2014/ir14febfc.pdf, accessed April 26,
2014.

Bureau of Economic Analysis, U. S. Department of Commerce (2012), *News Release: Gross Domestic Product*, November 29,
www.bea.gov/newsreleases/national/gdp/gdpnewsrelease.htm, accessed April 26, 2014.

Cochran, W., F. Mosteller, and J. Tukey (1954), *Statistical Problems of the Kinsey Report on Sexual Behavior in the Human Male,* Washington, DC: American Statistical Association.

Croushore, D. (2011), "Frontiers of Real-Time Data Analysis," *Journal of Economic Literature*, 49, 72-100.

Findley, D., B. Monsell, W. Bell, M. Otto and B. Chen (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program," *Journal of Business & Economic Statistics*, 16, 127-152.

Fixler, D., R. Greenaway-McGrevy, and B. Grimm (2011), "Revisions to GDP, GDI, and Their Major Components," *Survey of Current Business*, 91 (7), 9-31.

Granger, C. (1979), "Seasonality: Causation, Interpretation, and Implications," in A. Zellner (editor), *Seasonal Analysis of Economic Time Series*, National Bureau of Economic Research, 33-46.

Hokayem, C., C. Bollinger, and J. Ziliak, "The Role of CPS Nonresponse on the Level and Trend in Poverty," University of Kentucky Center for Poverty Research Discussion Paper Series, DP2014-05.

Horowitz, J. and C. Manski (1998), "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37–58.

Horowitz, J. and C. Manski (2000), "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77–84.

Imbens, G. and C. Manski (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Landefeld, J., E. Seskin, and B. Fraumeni (2008), "Taking the Pulse of the Economy: Measuring GDP," *Journal of Economic Perspectives*, 22, 193-216.

Little, R. and D. Rubin (1987), *Statistical Analysis with Missing Data*, New York: Wiley.

Manski, C. (1989), "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343–360.

Manski, C. (1994), "The Selection Problem," in C. Sims (editor) *Advances in Econometrics, Sixth World Congress*, Cambridge: Cambridge University Press.

Manski, C. (2003), *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

Manski, C. (2011), "Policy Analysis with Incredible Certitude," *The Economic Journal*, 121, F261-F289.

Manski, C. (2013a), *Public Policy in an Uncertain World*, Cambridge, MA: Harvard University Press.

Manski, C. (2013b), "Credible Interval Estimates for Official Statistics with Survey Nonresponse," Department of Economics, Northwestern University.

National Research Council (2013), *Principles and Practices for a Federal Statistical Agency: Fifth Edition,* Washington, DC: The National Academies Press.

Sims, C. (1979), "Comments on 'Seasonality: Causation, Interpretation, and Implications,' by Clive W. J. Granger," in A. Zellner (editor), *Seasonal Analysis of Economic Time Series*, National Bureau of Economic Research, 47-49.

U. K. Office for National Statistics (2013), *Economic Review, April 2014*, www.ons.gov.uk/ons/dcp171766_358477.pdf, accessed April 26, 2014.

U. S. Bureau of Labor Statistics (2012), *Employment Situation News Release*, October 5, www.bls.gov/news.release/archives/empsit_10052012.htm, accessed April 26, 2014.

U. S. Bureau of Labor Statistics (2001), *Labor Force Statistics from the Current Population Survey*, www.bls.gov/cps/seasfaq.htm, accessed April 26, 2014.

U. S. Census Bureau (2006), *Current Population Survey Design and Methodology*, Technical Paper 66, Washington, DC: U. S. Census Bureau.

U.S. Census Bureau (2011), *Current Housing Reports, Series H150/09, American Housing Survey for the United States: 2009*, Washington, DC: U.S. Government Printing Office.

U. S. Census Bureau (2012A), *Income, Poverty and Health Insurance Coverage in the United States: 2011*, September 12, www.census.gov/newsroom/releases/archives/income_wealth/cb12-172.html, accessed April 26, 2014.

U. S. Census Bureau (2012B), *Source and Accuracy of Estimates for Income, Poverty, and Health Insurance Coverage in the United States: 2011*, Publication P60_243sa, www.census.gov/hhes/www/p60_243sa.pdf, accessed April 26, 2014.

U. S. Census Bureau (2014), *New Residential Sales in March 2014*, www.census.gov/construction/nrs/pdf/newressales.pdf, accessed April 26, 2014.

Wright, J. (2013), "Unseasonal Seasonals?" Department of Economics, Johns Hopkins University.