University of Waterloo W. H. Cherry

ERROR: Measurement Error – *The mismeasure of statistical reckoning*

EM9634: The Globe and Mail, November 30, 1996, page D5

FAULTY FIGURING / *Economists crunch numbers to justify policy, predict the future, explain the past. What they're not doing is checking to see if their original figures are accurate.*

The mismeasure of statistical reckoning

The Economist
London

"THERE are three kinds of economists: those who can count and those who can't." That old joke gets a good laugh at economics conferences, yet it cuts dangerously close to the bone. Economists spend much time churning statistics through computer models or using them to justify policy, but few worry about the reliability of those numbers. They ought to – traditional measures of economic performance are becoming increasingly dodgy.

Number-crunching is not just an academic issue. Important questions, such as why all the billions of dollars invested in computers have failed to boost productivity growth, rest upon the accuracy of official statistics. Faulty figures distort people's vision. The economic debate in the United States, for example, has been shaped partly by official numbers showing that productivity growth has slowed – from an annual rate of 2.6 per cent in 1960-1973 to 0.9 per cent in 1980-1995 – and that real wages have stagnated. Calculate those figures correctly, however, and the true rate of U.S. productivity growth in the 1990s could be almost as high as in the 1960s, while real wages could be rising at a respectable pace.

The question of the reliability of official statistics will get a public airing in the U.S. in the next few weeks, when a commission headed by economist Michael Boskin issues a report to the Finance Committee of the U.S. Senate on the consumer-price index. It is expected to conclude that the CPI has overstated annual inflation in recent years by one to two percentage points. This, in turn, suggests that real growth in GDP and productivity has been understated, since a variant of the CPI is used to create inflation-adjusted figures.

This is no small matter. If the CPI overstates inflation, the budget deficit swells because of the over-indexation of welfare benefits, government pensions and income-tax brackets. If the U.S. inflation rate is overstated by just one percentage point a year, then after 10 years this would add an estimated $140 billion to its federal budget deficit.

In recent years, governments have made big efforts to improve the accuracy, timeliness and integrity of the statistics they collect. But are the numbers they collect relevant? Far too many statisticians are still trying to measure the output of the 19th century rather than the expected wizardry of the 21st. Three powerful forces pose special challenges to traditional statistics:

If the measurement bias were constant over time, it might not matter. But the margin of error is widening. Rapid technological change and the expansion of the so-called 'intangible economy' has exacerbated the problem of adjusting for quality changes and for new goods and services. Many goods did not exist two years ago, let alone 10 years ago. This makes it hard to compare products over time, and thus declines in price (and gains in output) are missed.

- **Globalization.** A growing proportion of trade and investment reflects internal decisions made by multinational corporations, which makes it harder to define national economic performance. Is Germany's economy measured best by including Volkswagen's production abroad but excluding Ford's output in Germany? Some economists would argue that it is. But this would change the economic picture dramatically – and would instantly give the United States a trade surplus instead of a chronic deficit.

- **Invisibility.** Conventional statistics were originally devised for tracking the production of physical goods. But a growing slice of output consists not of material things but the production and manipulation of ideas. Output has become less visible and hence less measurable. The number-crunchers have failed to keep pace. They still churn out masses of figures on the lamb population, steel output or sales of "rubber and miscellaneous plastic products." Yet fast-growing sectors such as software, telecommunications, entertainment, health care and financial services are barely tracked. In fields such as education and finance, government statistics often assume that output simply rises in line with the number of hours worked. Thus, by definition, productivity never rises.

- **Technology.** New goods, shorter product cycles and rapid quality improvements make it harder to measure changes in output and prices over time. Isn't it great that faster recovery times from operations mean patients spend less time in hospital? Not from a statistician's point of view: if measured by occupancy of hospital beds, output would show a decline. A trucking firm might improve its service by using computer navigation to run its trucks more efficiently. But productivity as measured by tonne-miles would drop if trucks reach their destinations more directly.

The problem here is insoluble. To measure the increase in real output over time, it is necessary to define a unit of production. This is easy for basic goods, such as steel, but for a growing slice of the economy the concept of a unit of output is becoming increasingly fuzzy. As long as that is true, statistics will fail to capture many of the advances in the modern economy.

The share of the economy that can be measured accurately is dwindling. Zvi Griliches, an economist at Harvard University, points out that in 1947 half of the U.S. output was accounted for by farming, mining and manufacturing, and so was relatively easy to measure. Today these sectors account for less than 30 per cent of output. And even manufacturing is becoming trickier to track as shorter product cycles and more rapid improvements in quality have made traditional measures less reliable.

Yet governments have been slow to examine the problem. There have been pathetically few studies that try to assess the size of measurement errors, and virtually all of those have been conducted in the United States. Anecdotal evidence suggests that measurement problems in Europe and other regions of the globe are just as severe.

If the measurement bias were constant over time, it might not matter. But the margin of

(*continued overleaf*)

error is widening. Rapid technological change and the expansion of the so-called "intangible economy" has exacerbated the problem of adjusting for quality changes and for new goods and services. Many goods did not exist two years ago, let alone 10 years ago. This makes it hard to compare products over time, and thus declines in price (and gains in output) are missed. In computers, for instance, the average life of a model is now less than 12 months. Some 30 per cent of consumer-electronics sales are of products that did not exist a year earlier.

Moreover, standard economic statistics fail to capture many of the benefits of information technology, which increasingly take the form not of cost saving or greater volume but of improved quality, time saving, convenience and increased consumer choice. Banking with a machine rather than a human teller can save customers both time and anguish. Neither benefit will appear in countries' national accounts.

Leonard Nakamura, an economist at the Federal Reserve Bank of Philadelphia, argues that if full account is taken of the benefits to consumers of new goods, services and product improvements, then GDP growth may have been understated by as much as two to three percentage points a year. After 10 years, that would leave the official American GDP one-third smaller than the true GDP. The degree of mismeasurement has been far higher since 1974, Mr. Nakamura says. That could account for almost all the productivity slowdown in the U.S. since 1974. Luc Soete, an economist at the University of Limburg in the Netherlands, reckons that the margin of error in Europe's growth rate is probably just as large.

If productivity and real incomes are growing faster than the official figures say, does this mean governments can sit back and relax? Far from it. Better policies would still yield an even better performance. What it does suggest, however, is that the correct policies may not be those that are signalled by the flawed statistics.

Statisticians face a big challenge in trying to boost the quality of their own production. There is certainly room for improvement. However, it is probably an unavoidable (if ironic) fact that in the so-called information age, when super-computers crunch mountains of data and satellites can track the precise movements of every man or machine, people's knowledge about the economy may be less exact than it was back when adding machines ruled the Earth.

1 The words *statistics* and *statisticians* are used with*out* an adjective in six places (twice each in the left- and right-hand columns overleaf on page HL14.1, once in the middle column overleaf and once in the last paragraph above) in the article EM9634 reprinted in this Highlight #14. Indicate briefly how these terms could be qualified to clarify their usage in the article.

2 The calculation of the consumer price index (CPI) is mentioned overleaf on page HL14.1 in the third paragraph of the article EM9634. Outline, in point form, the main *statistical* issues involved in this matter and their relevance to the discussion.
- Summarize the repercussion(s) of the mis-statement of the CPI as they are described in the article.

3 At two places in the third paragraph of the right-hand column overleaf on page HL14.1, the word *unit* is used. Compare and contrast these use(s) with the elements that make up a population in the terminology of the Formulation and Design stages of the FDEAC cycle.

4 In the second-last paragraph overleaf on page HL14.1, the phrase *measurement errors* is used. Outline what the phrase encompasses as it is used in the article EM9634.
- Compare and contrast the meaning(s) you have identified with the use of *error* in the statement:

  overall error = study error + non-response error + sample error + sample attribute measurement error.

  [This expression is equation (HL18.1) on page HL18.4 in Statistical Highlight #18.]

5 In the last line overleaf on page HL14.1 and the first line above, and at the bottom of the middle column above, the phrase *margin of error* is used. Outline what the phrase means as it is used in the article.
- Compare and contrast the meaning you have identified with that of the same phrase in media reports of poll results.
  - Suggest an alternate to *margin of error* for use in the article EM9634 that might help avoid confusion with the more usual meaning of the phrase in statistics.

6 Explain briefly the issues raised in the context of the article by the sentence in the last two lines overleaf on page HL14.1: *If the measurement bias were constant over time, it might not matter.*
- Outline why measuring inaccuracy is a matter of concern in statistics.

7 Discuss critically in the context of the article EM9634 the statement at the bottom of the left-hand column above and the top of the middle column: *Banking with a machine rather than a human teller can save customers both time and anguish.*

8 Present a calculation that provides a basis for the statement in the second paragraph of the middle column above: *After 10 years, that (GDP growth understated by as much as two to three percentage points a year) would leave the official American GDP one-third smaller than the true GDP.*

The article EM9634 reprinted overleaf and above is also used in Figure 2.21 in the STAT 220 Course Materials.