FORECASTING WITH REAL-TIME MACROECONOMIC DATA

Dean Croushore

University of Richmond

June 2005

Thanks to participants at the Handbook of Economic Forecasting Conference in 2004 and to two anonymous referees for their comments. Tom Stark (Philadelphia Fed) deserves credit for helping me think about many of these issues. Please send comments to the author at: Robins School of Business, 1 Gateway Road, University of Richmond, VA 23173, or e-mail: dcrousho@richmond.edu.

FORECASTING WITH REAL-TIME MACROECONOMIC DATA

Forecasts are only as good as the data behind them. But macroeconomic data are revised, often significantly, as time passes and new source data become available and conceptual changes are made. How is forecasting influenced by the fact that data are revised?

To answer this question, we begin with the example of the index of leading economic indicators to illustrate the real-time data issues. Then we look at the data that have been developed for U.S. data revisions, called the "Real-Time Data Set for Macroeconomists" and show their basic features, illustrating the magnitude of the revisions and thus motivating their potential influence on forecasts and on forecasting models. The data set consists of a set of data vintages, where a data vintage refers to a date at which someone observes a time series of data; so the data vintage September 1974 refers to all the macroeconomic time series available to someone in September 1974.

Next, we examine experiments using that data set by Stark-Croushore (2002), to illustrate how the data revisions could have affected reasonable univariate forecasts. In doing so, we tackle the issues of what variables are used as "actuals" in evaluating forecasts and we examine the techniques of repeated observation forecasting, illustrate the differences in U.S. data of forecasting with real-time data as opposed to latest-available data, and examine the sensitivity to data revisions of model selection governed by various information criteria.

Third, we look at the economic literature on the extent to which data revisions affect forecasts, including discussions of how forecasts differ when using first-available compared with latest-available data, whether these effects are bigger or smaller depending on whether a variable is being forecast in levels or growth rates, how much influence data revisions have on model selection and specification, and evidence on the predictive content of variables when subject to revision.

Given that data are subject to revision and that data revisions influence forecasts, what should forecasters do? Optimally, forecasters should account for data revisions in developing their forecasting models. We examine various techniques for doing so, including state-space methods.

The focus throughout this chapter is on papers mainly concerned with model development—trying to build a better forecasting model, especially by comparing forecasts from a new model to other models or to forecasts made in real time by private-sector or government forecasters.

I. An Illustrative Example: The Index of Leading Indicators

Figure 1 shows a chart of the index of leading indicators from November 1995, which was the last vintage generated by the U.S. Commerce Department before the index was turned over to the private-sector Conference Board, which no longer makes the index freely available. A look at the chart suggests that the index is fairly good at predicting recessions, especially those recessions that began in the 1960s and 1970s. (For more on using leading indicators to forecast, see the chapter by Marcelino on "Leading Indicators" in this volume.)

But did the index of leading indicators provide such a useful signal about the business cycle in real time? The evidence suggests skepticism, as Diebold and Rudebusch (1991a, 1991b) suggested. They put together a real-time data set on the leading indicators and concluded that the index of leading indicators does not lead and it does not indicate!

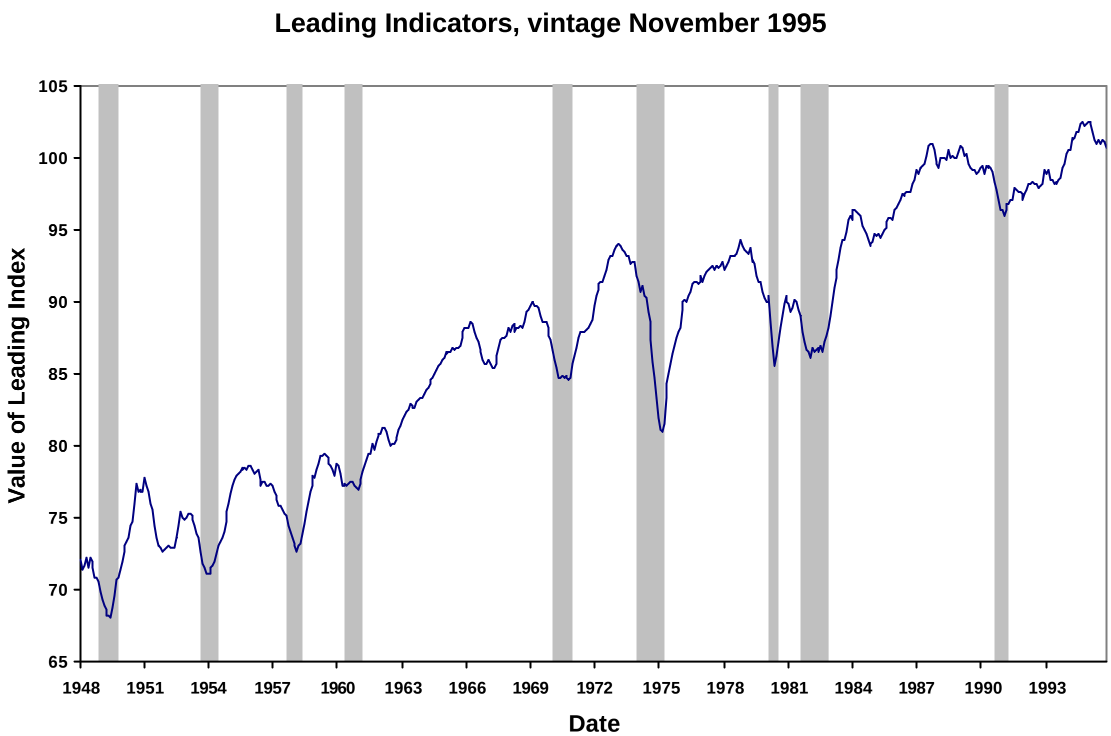

Figure 1: Leading Indicators, vintage November 1995

This chart shows the last vintage of the index of leading indicators from the Commerce Department in November 1995 before the U.S. government sold the index to the Conference Board. Note that the index declines before every recession and seems to provide a useful signal for the business cycle.
Source: Survey of Current Business (November 1995)

To see what the real-time evidence is, examine Figure 2, which shows the values of the index of leading indicators, as reported by the Department of Commerce in its publication *Business Conditions Digest* in September 1974. The index appears to be on a steady rise, with a few fits and starts. But nothing in the index suggests that a recession is likely. And the same is true if you examine any of the data vintages before September 1974. Unfortunately, a recession began in November 1973. So, even ten months after the recession began, the index of leading indicators gave no sign of a slowdown in economic activity.

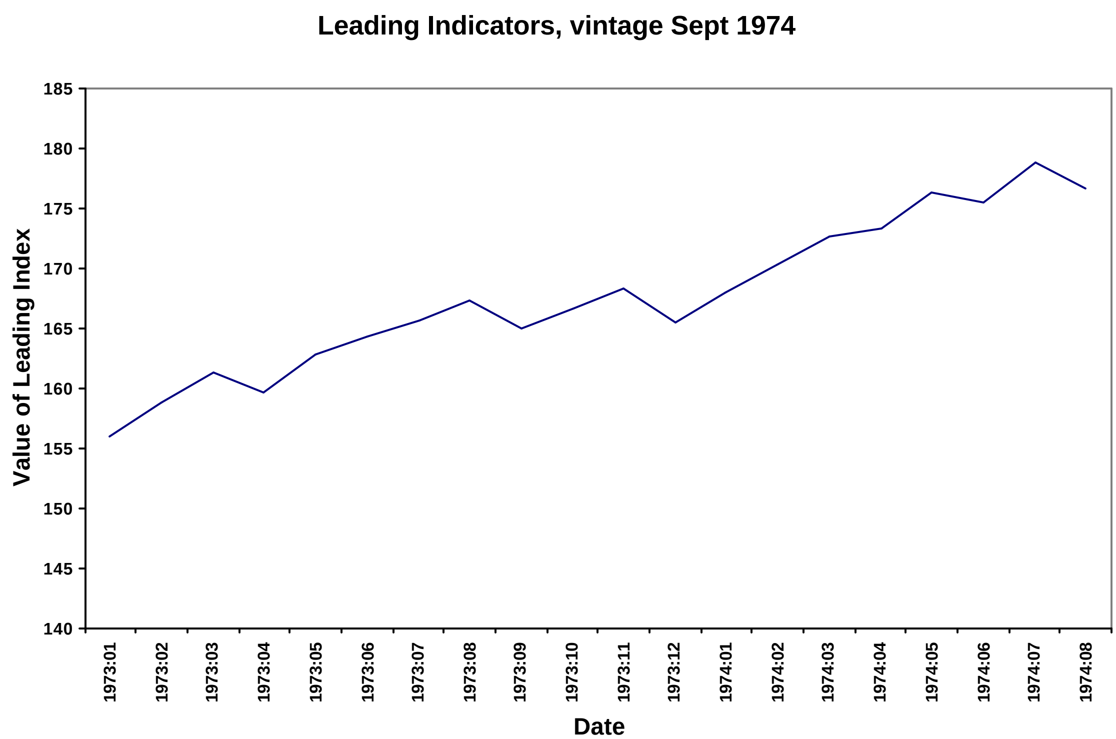

Figure 2: The Index of Leading Indicators, Vintage September 1974
This diagram shows the value of the index of leading indicators from January 1973 to August 1974, based on the data vintage of September 1974. No recession is in sight. But the NBER declared that a recession began in November 1973. *Source: Business Conditions Digest*, September 1974

Naturally, the failure to predict the recession led the Commerce Department to revise the construction of the index, which they did after the fact. The data entering the index were revised over time, but more importantly so were the methods used to construct the index. Figure 3 shows the original (September 1974 vintage) index of leading indicators and the revised index, as it stood in December 1989, over the sample period from January 1973 to August 1974. The index of leading indicators looks much better in the later vintage version; but in real time it was of no value. Thus the revised index gives a misleading picture of the forecasting ability of the leading indicators.

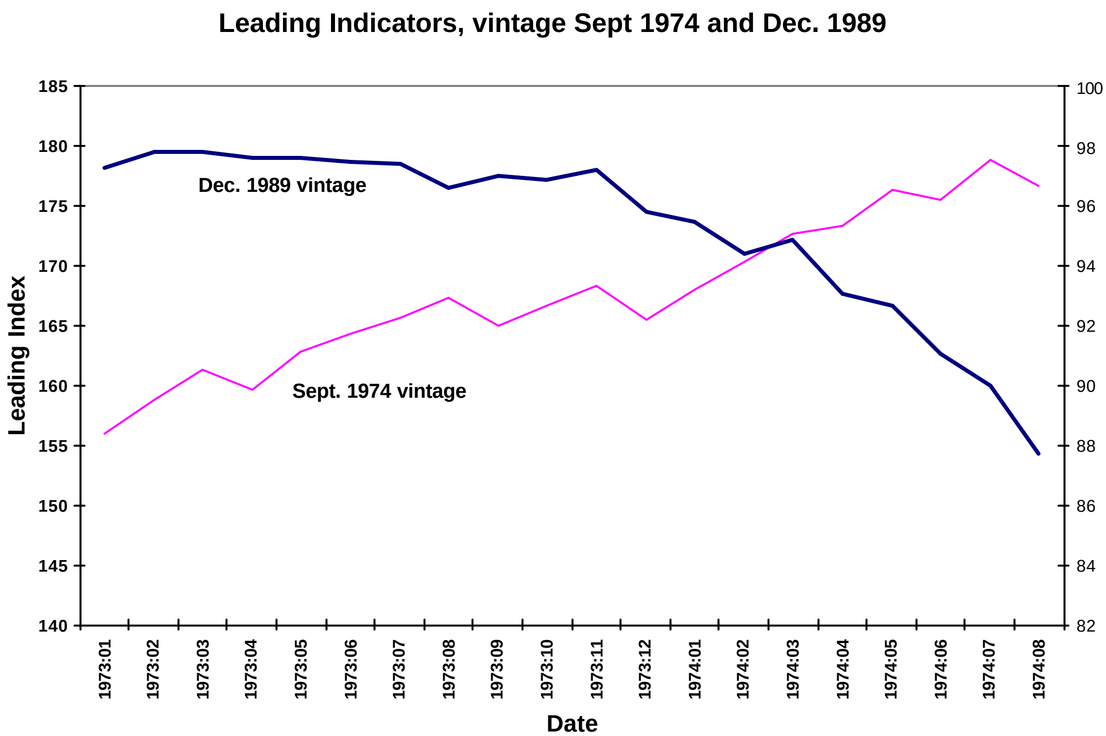

Figure 3: The Index of Leading Indicators, Vintages September 1974 and December 1989

This diagram shows the value of the index of leading indicators from January 1973 to August 1974, based on the data vintages of both September 1974 and December 1989. The revised version of the index predicts the recession nicely. But in real time, the index gave no warning at all.
Source: Business Conditions Digest, September 1974 and December 1989

II. The Real-Time Data Set for Macroeconomists

Until recently, every paper in the literature on real-time data analysis was one in which researchers pieced together their own data set to answer the particular question they wanted to address. In the early 1990s, while working on a paper using real-time data, I decided that it would be efficient to create a single, large data set containing real-time data on many different macroeconomic variables. Together with my colleague Tom Stark at the Federal Reserve Bank of Philadelphia, we created the Real-Time Data Set for Macroeconomists (RTDSM) containing real-time data for the United States.

The original motivation for the data set came from modelers who developed new forecasting models that they claimed produced better forecasts than the Survey of Professional Forecasters (a survey of forecasters around the country that the Philadelphia Fed conducted). But there was a key difference in the data sets that the researchers used (based on latest available data that had been revised many times) compared with the data set that the forecasters used in real time. Thus we hatched the idea of creating a set of data sets, one for each date in time (a vintage), consisting of data as it existed at that time. This would allow a researcher to test a new forecasting model on data that forecasters had available to them in real time, thus allowing a convincing comparison to determine if a model really was superior.

In addition to comparing forecasting models, the data set can also be used to examine the process of data revisions, test the robustness of empirical results, analyze government policy, and examine whether the vintage of the data matters in a research project. The data set is described and the process of data revisions is explored in Croushore-Stark (2001) and many tests of empirical results in macroeconomics are conducted in Croushore-Stark (2003).

The RTDSM is made available to the public at the Philadelphia Fed's web site: www.phil.frb.org/econ/forecast/reaindex.html. The data set contains vintages from November 1965 to the present, with data in each vintage going back to 1947Q1. Some vintages were collected once each quarter and others were collected monthly. The timing of the quarterly data sets is in the middle of the quarter (the 15th day of the middle month of the quarter), which matches up fairly closely with the deadline date for participants in the Survey of Professional Forecasters. The data set was made possible by numerous interns from Princeton University and the University of Pennsylvania (especially a student at Penn named Bill Wong who contributed tremendously to the data set's development), along with many research assistants from the Federal Reserve Bank of Philadelphia. In addition, some data were collected in real time, beginning in 1991. The data are fairly complete, though there are some holes in a few spots that occurred when the government did not release complete data or when we were unable to find hard copy data files to ensure that we had the correct data for the vintage in question. The data underwent numerous edit checks; errors are possible but are likely to be small.

Variables included in RTDSM to date are: Variables with Quarterly Observations and Quarterly Vintages: Nominal output, real output, real consumption (broken down into durable, nondurable, and services), real investment (broken down into business fixed investment, residential investment, and change in business inventories), real government purchases (more recently, government consumption expenditures and gross investment; broken down between federal and state-and-local governments), real exports, real imports, the chain-weighted GDP price index, the price index for imports, nominal corporate profits after taxes, nominal personal saving, nominal disposable personal income, nominal personal consumption expenditures, and nominal personal income; Variables with Monthly Observations and Quarterly Vintages: Money supply measures M1 & M2, money reserve measures (total adjusted reserves, nonborrowed reserves, and nonborrowed reserves plus extended credit; all based on Board of Governors' definitions), the adjusted monetary base (Board of Governors' definition), civilian unemployment rate, and the consumer price index; Variables with Monthly Observations and Monthly Vintages: payroll employment, industrial production, and capacity utilization. New variables are being added each year.

Studies of the revision process show that a forecaster could predict the revisions to some variables, such as industrial production. Other variables, such as payroll employment, show no signs of predictability at all. Some variables are revised dramatically, such as corporate profits, while others have very small revisions, such as the consumer price index.

The data in RTDSM are organized in two different ways. The data were initially collected in a setup in which one worksheet was created to hold the complete time series of all the variables observed at the vintage date. An alternative structure, showing all the vintage dates for one variable, is shown in Figure 4. In that structure, reading across columns shows you how the value of an observation changes across vintages. Each column represents the time series that a researcher would observe at the date shown in the column header. Dates in the first column are observation dates. For example, the upper left data point of 306.4 is the value of real output for the first quarter of 1947, as recorded in the data vintage of November 1965. The setup makes it easy to see when revisions occur. In Figure 4, note that the large changes in values in the first row are the result of changes in the base year, which is the main reason that real output jumps from 306.4 in vintages November 1965, February 1966, and May 1966, to 1481.7 in vintage November 2003, to 1570.5 in vintage February 2004.

DATA STRUCTURE

REAL OUTPUT

Vintage:	Nov65	Feb66	May66	. . .	Nov03	Feb04
Date						
47Q1	306.4	306.4	306.4	. . .	1481.7	1570.5
47Q2	309.0	309.0	309.0	. . .	1489.4	1568.7
47Q3	309.6	309.6	309.6	. . .	1493.1	1568.0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
65Q3	609.1	613.0	613.0	. . .	3050.7	3214.1
65Q4	NA	621.7	624.4	. . .	3123.6	3291.8
66Q1	NA	NA	633.8	. . .	3201.1	3372.3
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
03Q2	NA	NA	NA	. . .	9629.4	10288.3
03Q3	NA	NA	NA	. . .	9797.2	10493.1
03Q4	NA	NA	NA	. . .	NA	10597.1

Figure 4. The Data Structure of the Real-Time Data Set for Macroeconomists
Each column of data represents a vintage, so reading the column shows you what a researcher observing the data at the date shown in the column header would observe. Reading across any row of data shows how the data value for the observation date shown in the first column was revised over time.

How Big Are Data Revisions?

If data revisions were small and random, we would not worry about how they affect forecasts. But work with the RTDSM shows that data revisions are large and systematic, and thus have the potential to affect forecasts dramatically.

For example, suppose we consider the revisions to real output in the short run by looking at the data for a particular quarter. Because of changes in the base year, we generally examine revisions based on growth rates. To see what the revisions look like in the short run, consider Figure 5, which shows the growth rate (seasonally adjusted at an annual rate) of real output in 1977Q1, as recorded in every quarterly vintage of data in RTDSM from May 1977 to February 2004.

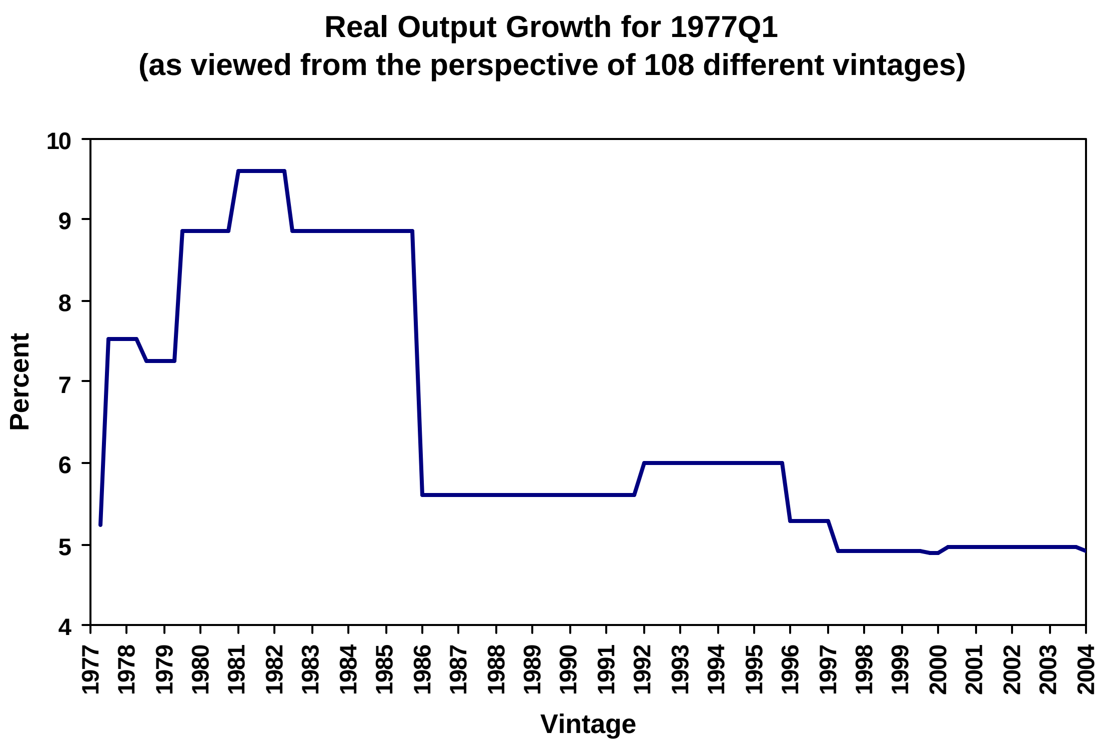

Figure 5. Real Output Growth for 1977Q1

This graph shows how the growth rate (seasonally adjusted at an annual rate) of real output for the observation date 1977Q1 has changed over vintages, from the first release vintage of May 1977 to the vintage of February 2004.

Figure 5 suggests that quarterly revisions to real output can be substantial. Growth rates vary over time from 4.9% in recent vintages, to 5.2% in the first available vintage (May 1977), to as high as 9.6% in vintages in 1981 and 1982. Naturally, short-term forecasts for real output for 1977 are likely to be greatly affected by the choice of vintage.

Although Figure 5 shows that some short-run revisions may be extreme, smaller revisions associated with seasonal adjustment occur every year in the data. To some extent, those revisions are predictable because of the government procedures for implementing seasonal adjustment, as described in the chapter by Ghysels-Osborn-Rodrigues, "Forecasting Seasonal Times Series."

Though Figure 5 might be convincing for the short run, many economic issues depend not just on short-run growth rates but on longer-term growth rates. If data revisions are small and average out to zero over time, then data revisions are not important for long-run forecasting. To investigate the issue of how long-term growth rates are influenced by data revisions, Figure 6 illustrates how five-year average growth rates are affected across vintages. In the table, each row shows the average growth rate over the period shown in the first column from the vintage of data shown in the column header. Those vintage dates are the vintage dates just before a benchmark revision to the national income accounts, except for the last column which shows the data as of November 2001.

Figure 6.
Average Growth Rates Over Five Years
For Benchmark Vintages
Annualized percentage points

Vintage Year: Period	'75	'80	'85	'91	'95	'01
			Real Output			
49Q4 to 54Q4	5.2	5.1	5.1	5.5	5.5	5.3
54Q4 to 59Q4	2.9	3.0	3.0	2.7	2.7	3.2
59Q4 to 64Q4	4.1	4.0	4.0	3.9	4.0	4.2
64Q4 to 69Q4	4.3	4.0	4.1	4.0	4.0	4.4
69Q4 to 74Q4	2.1	2.2	2.5	2.1	2.3	2.6
74Q4 to 79Q4	NA	3.7	3.9	3.5	3.4	4.0
79Q4 to 84Q4	NA	NA	2.2	2.0	1.9	2.5
84Q4 to 89Q4	NA	NA	NA	3.2	3.0	3.5
89Q4 to 94Q4	NA	NA	NA	NA	2.3	2.4
94Q4 to 99Q4	NA	NA	NA	NA	NA	3.9
			Real Consumption			
49Q4 to 54Q4	3.6	3.3	3.3	3.7	3.9	3.8
54Q4 to 59Q4	3.4	3.3	3.3	3.3	3.4	3.5
59Q4 to 64Q4	4.1	3.8	3.8	3.7	3.8	4.1
64Q4 to 69Q4	4.5	4.3	4.4	4.4	4.5	4.8
69Q4 to 74Q4	2.3	2.6	2.6	2.5	2.6	2.8
74Q4 to 79Q4	NA	4.4	4.4	3.9	3.9	4.2
79Q4 to 84Q4	NA	NA	2.8	2.5	2.5	2.8
84Q4 to 89Q4	NA	NA	NA	3.2	3.1	3.7
89Q4 to 94Q4	NA	NA	NA	NA	2.3	2.4
94Q4 to 99Q4	NA	NA	NA	NA	NA	4.0
			Prices			
49Q4 to 54Q4	2.6	2.7	2.7	2.5	2.4	2.5
54Q4 to 59Q4	2.6	2.6	2.6	2.9	2.9	2.5
59Q4 to 64Q4	1.4	1.5	1.5	1.6	1.6	1.3
64Q4 to 69Q4	3.6	3.9	3.9	4.1	4.1	3.7
69Q4 to 74Q4	6.3	6.5	6.2	6.8	6.5	6.3
74Q4 to 79Q4	NA	7.1	7.0	7.5	7.7	7.1
79Q4 to 84Q4	NA	NA	6.1	6.1	6.4	6.0
84Q4 to 89Q4	NA	NA	NA	3.3	3.6	3.1
89Q4 to 94Q4	NA	NA	NA	NA	2.9	2.8
94Q4 to 99Q4	NA	NA	NA	NA	NA	1.7

Vintage Year: Period	‘75	‘80	‘85	‘91	‘95	’01
			Nominal Output			
49Q4 to 54Q4	7.9	7.9	7.9	8.1	8.0	8.0
54Q4 to 59Q4	5.6	5.6	5.7	5.7	5.7	5.7
59Q4 to 64Q4	5.6	5.5	5.6	5.6	5.7	5.6
64Q4 to 69Q4	8.0	8.1	8.2	8.3	8.2	8.3
69Q4 to 74Q4	8.6	8.8	8.9	9.1	9.0	9.1
74Q4 to 79Q4	NA	11.1	11.2	11.3	11.4	11.4
79Q4 to 84Q4	NA	NA	8.5	8.2	8.5	8.7
84Q4 to 89Q4	NA	NA	NA	6.5	6.7	6.7
89Q4 to 94Q4	NA	NA	NA	NA	5.2	5.3
94Q4 to 99Q4	NA	NA	NA	NA	NA	5.7

Figure 6. Average Growth Rates over Five Years for Benchmark Vintages
This table shows the growth rates over the five year periods shown in the first column of four different variables (real output, real consumption, the price level, and nominal output) for each benchmark vintage shown in the column header.

Figure 6 shows that even average growth rates over five years can be affected significantly by data revisions. For example, for real output, note the large differences in the last two columns of the table. Real output growth over five-year periods was revised by as much as 0.6 percentage points from the 1995 vintage (just before chain weighting) to the newer vintage. Real consumption spending is also revised significantly, similar to the changes in output. Those differences arise in part because of revisions to the price index, as shown in the third section of the table. Changes in the base year, especially under the fixed-weight structure used before 1996, caused significant changes in price inflation and thus growth rates of real variables. In addition, redefinitions and changes in weights caused even nominal output growth to be revised, though the revisions to nominal output growth are of a smaller magnitude than the changes in the real variables.

In summary, in both the short run and the long run, data revisions may affect the values of data significantly. Given that data revisions are large enough to matter, we next examine how those revisions affect forecasts.

III. Why Are Forecasts Affected By Data Revisions?

Forecasts may be affected by data revisions for three reasons: (1) revisions change the data input into the forecasting model; (2) revisions change the estimated coefficients; and (3) revisions lead to a change in the model itself (such as the number of lags).

To see how data revisions might affect forecasts, consider a forecasting model that is an $AR(p)$. The model is:

$$Y_t = m + \sum_{i=1}^{p} f_i Y_{t-i} + e_t . \tag{1}$$

Suppose that the forecasting problem is such that a forecaster estimates this model each period, and generates forecasts of Y_t for several periods ahead. Because the forecasts must be made in real time, the data for the one variable in this univariate forecast are represented by a matrix of data, not just a vector, with a different column of the matrix representing a different vintage of the data. As in Stark-Croushore (2002), denote the data point (reported by a government statistical agency) for observation date t and vintage v as $Y_{t,v}$. The revision to the data for observation date t between vintages $v-1$ and v is $Y_{t,v} - Y_{t,v-1}$.

Now consider a forecast for date t one-period ahead (so that the forecaster's information set includes $Y_{t-1,v}$) when the data vintage is v. Then the forecast is:

$$Y_{t|t-1,v} = \hat{m}_v + \sum_{i=1}^{p} \hat{f}_{i,v} Y_{t-i,v} . \quad (2)$$

where the circumflex denotes an estimated parameter, which also needs a vintage subscript because the estimated parameter may change with each vintage.

Next consider estimating the same model with a later vintage of the data, *w*. The forecast is:

$$Y_{t|t-1,w} = \hat{m}_w + \sum_{i=1}^{p} \hat{f}_{i,w} Y_{t-i,w} . \quad (3)$$

The change to the forecast is:

$$Y_{t|t-1,w} - Y_{t|t-1,v} = (\hat{m}_w - \hat{m}_v) + \sum_{i=1}^{p} (\hat{f}_{i,w} Y_{t-i,w} - \hat{f}_{i,v} Y_{t-i,v}) \quad (4)$$

The three ways that forecasts may be revised can be seen in equation (4). First, revisions change the data input into the forecasting model. In this case, the data change from $\{Y_{t-1,v}, Y_{t-2,v,}, ..., Y_{t-p,v}\}$ to $\{Y_{t-1,v}, Y_{t-2,v,}, ..., Y_{t-p,v}\}$. Second, the revisions lead to changes in the estimated values of the coefficients from $\{\hat{m}_v, \hat{f}_{1,v}, \hat{f}_{2,v}, ..., \hat{f}_{p,v}\}$ to $\{\hat{m}_w, \hat{f}_{1,w}, \hat{f}_{2,w}, ..., \hat{f}_{p,w}\}$. Third, the revisions could lead to a change in the model itself. For example, if the forecaster were using an information criterion at each date to choose *p*, then the number of lags in the autoregression could change as the data are revised.

How large an effect on the forecasts are data revisions likely to cause? Clearly, the answer to this question depends on the data in question and the size of the revisions to the data. For some series, revisions may be close to white noise, in which case we would not expect forecasts to change very much. But for other series, the revisions will be very large and idiosyncratic, causing huge changes in the forecasts, as we will see in the literature discussed in section IV.

Experiments to illustrate how forecasts are affected in these ways by data revisions were conducted by Stark-Croushore (2002), whose results are reported here via a set of three experiments: (1) repeated observation forecasting; (2) forecasting with real-time versus latest-available data; and (3) experiments to test information criteria and forecasts.

Before getting to those experiments, we need to first discuss a key issue in forecasting: what do we use as actuals? Because data may be revised forever, it is not obvious what data vintage a researcher should use as the "actual" value to compare with the forecast. Certainly, the choice of data vintage to use as "actual" depends on the purpose. For example, if Wall Street forecasters are attempting to project the first-release value of GDP, then we would certainly want to use the first-released value as "actual". But if a forecaster is after the true level of GDP, the choice is not so obvious. If we want the best measure of a variable, we probably should consider the latest-available data as the "truth" (though perhaps not in the fixed-weighting era prior to 1996 in the United States because chain-weighted data available beginning in 1996 are superior because growth rates are not distorted by the choice of base year, as was the case with fixed-weighted data). The problem with this choice of latest-available data is that forecasters would not anticipate redefinitions and would generally forecast to be consistent with government data methods. For example, just before the U.S. government's official statistics were changed to chain weighting in late 1996, forecasters were still forecasting the fixed-weight data, because no one in the markets knew how to evaluate chain-weighted data and official chain-weighted data for past years had not yet been released. So forecasters continued to project fixed-weight values, even though there would never be a fixed-weight actual for the period being forecast.

One advantage of the Real-Time Data Set for Macroeconomists is that it gives a researcher many choices about what to use as actual. You can choose the first release (or second, or third), the value four quarters later (or eight or twelve), the last benchmark vintage (the last vintage before a benchmark revision), or the latest-available vintage. And it is relatively easy to choose alternative vintages as actuals and compare the results.

Experiment 1: Repeated Observation Forecasting

The technique of repeated observation forecasting was developed by Stark-Croushore (2002). They showed how forecasts for a particular date change as vintage changes, using every vintage available. For example: Forecast real output growth one step ahead using an $AR(p)$ model on the first difference of the log level of real output, for each date from 1965Q4 to 1999Q3, using every vintage possible from November 1965 to August 1999 (136 vintages), using the AIC to choose p. Then plot all the different forecasts to see how they differ across vintages.

Figure 7 shows many different repeated-observation forecasts from the first half of the 1970s. For example, the first column of dots for 1970Q1 is made by forecasting with data from vintages February 1970 to August 1999, all using the same sample period of 1947Q1 to 1969Q4. The second column of dots for 1970Q2 is made by forecasting with data from vintages May 1970 to August 1999, all using the same sample period of 1947Q1 to 1970Q1. The last column of dots shows forecasts for 1974Q4 made by forecasting with data from vintages November 1974 to August 1999, all using the same sample period of 1947Q1 to 1974Q3.

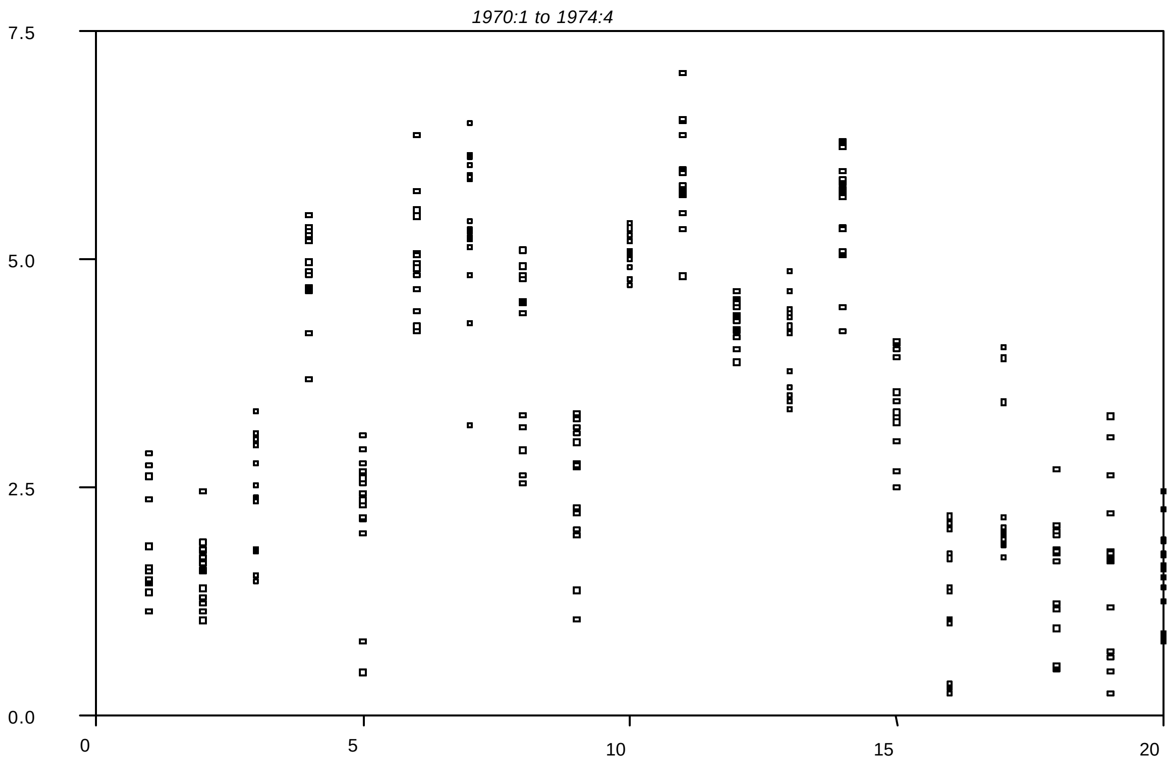

Figure 7. One-Step Ahead Forecasts for Real Output Growth, 1970Q1 to 1974Q4
Each column of points is one set of forecasts across vintages for a particular date. On the horizontal axis, each number corresponds to an observation date, with 1 = 1970Q1, 2 = 1970Q2, . . . 20 = 1974Q4. Each column of dots shows forecasts for the corresponding date. For example, the first column of dots for 1970Q1 is made by forecasting with data from vintages February 1970 to August 1999, all using the same sample period of 1947Q1 to 1969Q4. The vertical axis shows the forecasted growth rate of real output for that date.

The range of the forecasts in Figure 7 across vintages is relatively modest. But in other periods, with larger data revisions, the range of the forecasts in a column may be substantially larger. For example, Figure 8 shows the same type of graph as Figure 7, but for the second half of the 1970s. Note the increased range of forecasts in many of the columns. The increased range occurs because changes in base years affected the influence of changes in oil prices in those years, far more than was true earlier.

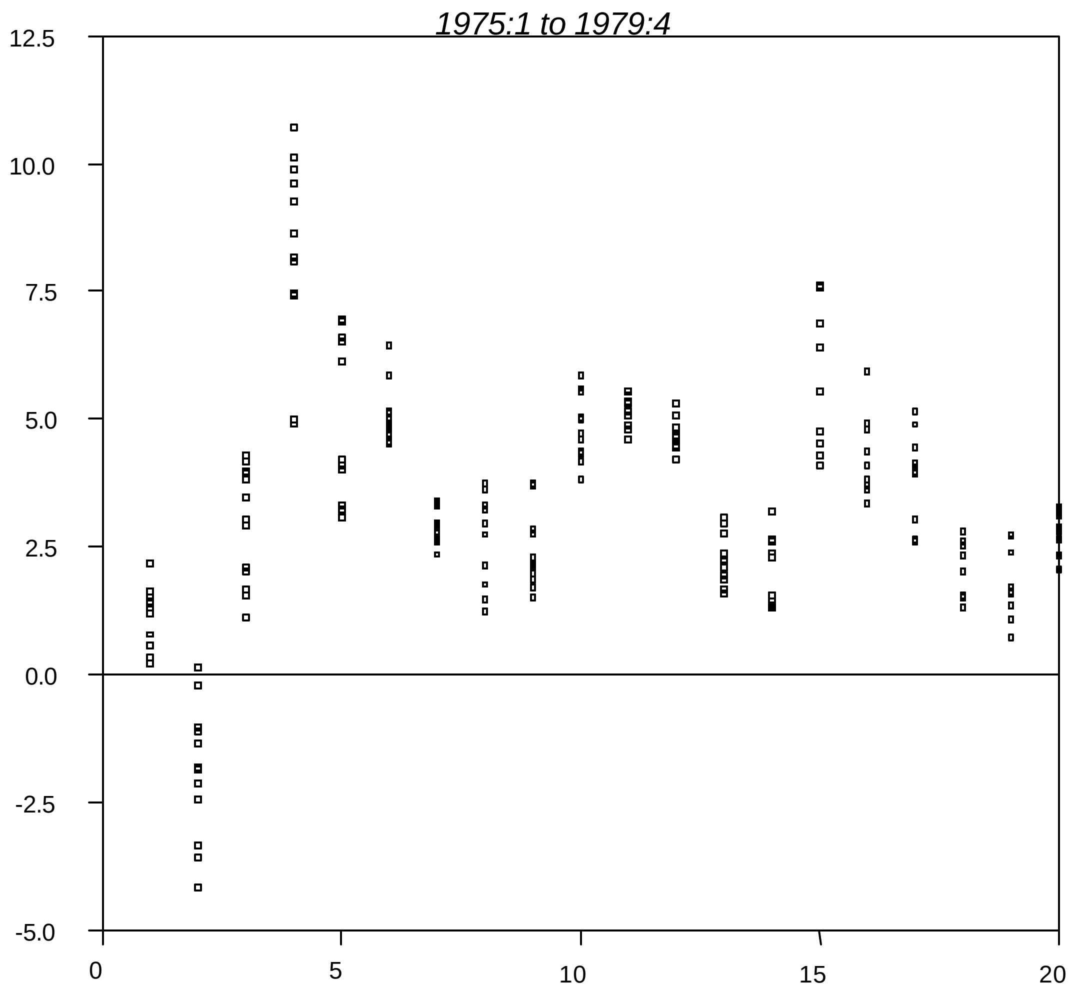

Figure 8. One-Step-Ahead Forecasts for Real Output Growth, 1975Q1 to 1979Q4
This graph is set up as in Figure 7, but covers the second half of the 1970s. The range of forecasts in the columns is much larger in many cases than in Figure 7.

In Figure 8, we can see that oil price shocks led to big data revisions, which in turn led to a large range of forecasts. In the fourth column, for example, the forecasts for 1975Q4 range from 4.89 percent to 10.68 percent.

Based on repeated-observation forecasts, Stark-Croushore suggested that inflation forecasts were more sensitive to data revisions than output forecasts. They found that the average ratio of the range of forecasts for output relative to the range of realizations was about 0.62, whereas the average ratio of the range of forecasts for inflation relative to the range of realizations was about 0.88. Possibly, inflation forecasts are more sensitive than output to data revisions because the inflation process is more persistent.

Another experiment by Stark-Croushore was to compare their results using the AIC to those of the SIC. Use of AIC rather than SIC leads to more variation in the model chosen and thus more variability in forecasts across vintages. The AIC chooses longer lags, which increases the sensitivity of forecasts to data revisions.

To summarize this section, it is clear that forecasts using simply univariate models depend strongly on the data vintage.

Experiment 2: Forecasting with Real-Time Versus Latest-Available Data Samples

Stark-Croushore's second major experiment was to use the RTDSM to compare forecasts made with real-time data to those made with latest-available data. They performed a set of recursive forecasts. The real-time forecasts were made by forecasting across vintages using the full sample available at each date, while the latest-available forecasts were made by performing recursive forecasts across sample periods with just the latest data vintage.

A key issue in this exercise is the decision about what to use as "actual," as we discussed earlier. Stark-Croushore use three alternative actuals: (1) latest available; (2)

the last before a benchmark revision (called benchmark vintages); and (3) the vintage one year after the observation date.

A priori, using the latest-available data in forecasting should yield better results, as the data reflect more complete information. So, we might think that forecasts based on such data would be more accurate. This is true for inflation data, but perhaps not for output data, as the Stark-Croushore results show.

One result of these experiments was that forecasts for output growth were not significantly better when based on latest-available data, even when latest-available data were used as actuals. This is a surprise, since such data include redefinitions and rebenchmarks, so you might think that forecasts based on such data would lead to more accurate forecasts.

However, Stark-Croushore showed that in smaller samples, there may be significant differences between forecasts. For example, in the first half of the 1970s, forecasts of output growth based on real-time data were significantly better than forecasts of output growth based on latest-available data, which is very surprising. However, in other short samples, the real-time forecasts are significantly worse than those using latest-available data. So, we can not draw any broad conclusions about forecasting output growth using real-time versus latest-available data.

Forecasts of inflation are a different matter. Clearly, according to the Stark-Croushore results, forecasts based on latest-available data are superior to those using real-time data, as we might expect. This is true in the full sample as well as sub-samples.

Stark-Croushore suggests then that forecasts can be quite sensitive to data vintage and that the vintage chosen and the choice of actuals matters significantly for forecasting results. When model developers using latest-available data find lower forecast errors than real-time forecasters did, it may not mean that their forecasting model is superior; it might only mean that their data are superior because of the passage of time.

Experiment 3: Information Criteria and Forecasts

In one final set of experiments, Stark-Croushore look at the choice of lag length in an ARIMA(p,1,0), comparing the use of AIC with the use of SIC. They examine whether the use of real-time versus latest-available data matters for the choice of lag length and hence the forecasts made by each model. Their results suggest that the choice of real-time versus latest-available data matters much more for AIC than for SIC.

Elliott (2002) illustrated and explained some of the Stark-Croushore results. He showed that the lag structures for real-time and revised data are likely to be different, that greater persistence in the latest-available series increases those differences, and that RMSEs for forecasts using revised data may be substantially less than for real-time forecasts. Monte Carlo results showed that the choices of models made using AIC or BIC is much wider using real-time data than using revised data. Finally, Elliott suggested constructing forecasting models with both real-time and revised data at hand, an idea we will revisit in section V.

IV. The Literature on How Data Revisions Affect Forecasts

In this section, we examine how data revisions affect forecasts, by reviewing the most important papers in the literature. We being by discussing how forecasts differ when using first-available compared with latest-available data. We examine whether these effects are bigger or smaller depending on whether a variable is being forecast in levels or growth rates. Then we investigate the influence data revisions have on model selection and specification. Finally, we examine the evidence on the predictive content of variables when subject to revision. The key question in this literature is: do data revisions affect forecasts significantly enough to make one worry about the quality of the forecasts?

How Forecasts Differ When Using First-Available Data Compared with Latest-Available Data

One way to illustrate how data revisions matter for forecasts is to examine a set of forecasts made in real-time, using data as it first became available, then compare those forecasts to those made using the same forecasting method but using latest-available data.

The first paper to compare forecasts using this method was Denton-Kuiper (1965). They used Canadian national income account data to estimate a six-equation macroeconomic model with two-stage-least-squares methods. They used three different data sets: (1) preliminary data (1^{st} release); (2) mixed data (real time); and (3) latest-available data. Denton-Kuiper suggests eliminating the use of variables that are revised extensively, as they pollute parameter estimates. But they were dealing with a very small data sample, from 1949 to 1958.

The next paper to examine real-time data issues is Cole (1969). She examined the extent to which data errors contribute to forecast errors, focusing on data errors in variables that are part of an extrapolative component of a forecast (e.g., extrapolating future values of an exogenous variable in a large system). Cole finds that: (1) data errors reduce forecast efficiency (variance of forecast error is higher), (2) lead to higher mean squared forecast errors because of changes in coefficient estimates, and (3) lead to biased estimates if the expected data revision is non-zero.

Cole's results were based on U.S. data from 1953 to 1963. She examined three types of models: (1) naïve projections, for which the relative root-mean-squared-error averages 1.55, and is over 2 for some variables, for preliminary data compared with latest-available data; (2) real-time forecasts made by professional forecasters, in which she regressed forecast errors on data revisions, finding significant effects for some variables and finding that data revisions were the primary cause of bias in about half of the forecasts, as well as finding a bigger effect for forecasts in levels than growth rates; and (3) a forecasting model of consumption (quarterly data, 1947–1960), in which coefficient estimates were polluted by data errors by 7 to 25 percent, depending on the estimation method, in which she found that forecasts were biased because of the data errors and that "the use of preliminary rather than revised data resulted in a *doubling* of the forecast error."

Cole introduced a useful technique, following these three steps: (1) forecast using preliminary data on model estimated with preliminary data; (2) forecast using revised data on a model estimated with preliminary data; and (3) forecast using revised data on a model estimated with revised data. Then comparing forecasts (1) and (3) shows the total effect of data errors; comparing forecasts (1) and (2) shows the direct effect of data errors for given parameter estimates; and comparing forecasts (2) and (3) shows the indirect effect of data errors through their effect on parameter estimates.

Given that data revisions affect forecasts in single-equation systems, we might wonder if the situation is better or worse in simultaneous-equation systems. To answer that question, Trivellato-Rettore (1986) showed how data errors contribute to forecast errors in a linear dynamic simultaneous-equations model. They found that data errors affect everything: estimated coefficients, lagged variables, and projections of exogenous variables. They examined a small (4 equation) model of the Italian economy for the sample period 1960 to 1980. However, the forecast errors induced by data revisions were not large. They found that for one-year forecasts, data errors led to biased coefficient estimates by less than 1% and contributed at most 4% to the standard error of forecasts. Thus, data errors were not much of a problem in the model.

Another technique used by researchers is that of Granger causality tests. Swanson (1996) investigate the sensitivity of such tests, using the first release of data compared with latest-available data and found that bivariate Granger causality tests are sensitive to the choice of data vintage.

A common method for generating inflation forecasts is to use equations based on a Phillips curve in which a variable such as the output gap is the key measure of economic slack. But a study of historical measures of the output gap by Orphanides (2001) found that such measures vary greatly over vintages—long after the fact, economists are much more confident about the size of the output gap than they are in real time. To see how uncertainty about the output gap affects forecasts of inflation, Orphanides-van Norden (2005) used real-time compared with latest-available data to show that ex-post output gap measures are useful in forecasting inflation. But in real time, out-of-sample forecasts of inflation based on measures of the output gap are not very useful. In fact, although the evidence that supports the use of the output-gap concept for forecasting inflation is very strong when output gaps are constructed on latest-available data, using the output gap is inferior to other methods in real-time, out-of-sample tests. Edge-Laubach-Williams (2004) found similar results for forecasting long-run productivity growth.

One of the most difficult variables to forecast is the exchange rate. Some recent research, however, showed that the yen-dollar and Deutschemark-dollar exchange rates were forecastable, using latest-available data. However, a real-time investigation by Faust-Rogers-Wright (2003) compared the forecastability of exchange rates based on real-time data compared with latest-available data. They found that exchange-rate forecastability was very sensitive to the vintage of data used. Their results cast doubt on research that claims that exchange rates are forecastable.

Overall, the papers in the literature comparing forecasts made in real time to those made with latest-available data imply that using latest-available data sometimes gives quite different forecasts than would have been made in real time.

Levels versus Growth Rates

A number of papers have examined whether forecasts of variables in levels are more sensitive or less sensitive to data revisions than forecasts of those variables in growth rates. The importance of this issue can be seen by considering what happens to levels and growth rates of a variable when data revisions occur. Using the log of the ratio between two successive observation dates to represent the growth rate for vintage v, it is:

$$g_{t,v} = \ln \frac{Y_{t,v}}{Y_{t-1,v}}.$$

The growth rate for the same observation dates but with a different vintage of data w is:

$$g_{t,w} = \ln \frac{Y_{t,w}}{Y_{t-1,w}}.$$

How would these growth rates be affected by a revision to a previous observation in the data series? Clearly, the answer depends on how the revision occurs. If the revision is a one-time level shift, then the growth rate will be revised, as will the level of the variable. However, suppose the revision occurs such that $Y_{t,w} = (1+a)Y_{t,v}$ and $Y_{t-1,w} = (1+a)Y_{t-1,v}$. Then the level is clearly affected but the growth rate is not. So, how forecasts of levels and growth rates are affected by data revisions is an empirical question concerning the types of data revisions that occurs. (Most papers that study data revisions themselves have not been clear about the relationship between revisions in levels compared with growth rates.)

Howrey (1996) showed that forecasts of levels of real GNP are very sensitive to data revisions while forecasts of growth rates are almost unaffected. He examined the forecasting period 1986 to 1991, looking at quarterly data and using univariate models. He found that the variance of the forecasting error in levels was four times greater using real-time data than if the last vintage prior to a benchmark revision had been used. But he showed that there is little (5%) difference in variance when forecasting growth rates. He used as "actual" values in determining the forecast error the last data vintage prior to a benchmark revision. The policy implications of Howrey's research are clear: policy should feed back on growth rates (output growth) rather than levels (output gap). This is consistent with the research of Orphanides-van Norden described above.

Kozicki (2002) showed that the choice of using latest-available or real-time data is most important for variables subject to large level revisions. She showed that the choice of data vintage is particularly important in performing real out-of-sample forecasting for the purpose of comparing to real-time forecasts from surveys. She ran tests of in-sample forecasts compared with out-of-sample forecasts using latest-available data compared with out-of-sample forecasts using real-time data and found that for some variables over short sample periods, the differences in forecast errors can be huge. Surprisingly, in-sample forecasts were not too much better than out-of-sample forecasts. In proxying expectations (using a model to try to estimate survey expectations), there is no clear advantage to using real-time or latest-available data; results vary by variable. Also, the choice of vintage to use as "actuals" matters, especially for real-time forecasts, where using latest-available data makes them look worse.

In summary, the literature on levels versus growth rates suggests that forecasts of level variables are more subject to data revisions than forecasts of variables in growth rates.

Model Selection and Specification

We often select models based on in-sample considerations, or simulated out-of-sample experiments using latest-available data. But it is more valid to use real-time out-of-sample experiments, to see what a forecaster would have projected in real time. A number of papers in the literature have discussed this issue. Experiments conducted in this area include those by Swanson-White (1997), who were the first to use real-time data to explore model selection, Harrison-Kapetanios-Yates (2002) who showed that forecasts may be improved by estimating the model on older data that has been revised, ignoring the most recent data (more on this idea later in this chapter), and Robertson-Tallman (1998), who showed how real-time data matter for the choice of model in forecasting industrial production using the leading indicators, but the choice of model for forecasting GDP is not affected much.

Overall, this literature suggests that model choice is sometimes affected significantly by data revisions.

Evidence on the Predictive Content of Variables

Few papers in the forecasting literature have examined the evidence of the predictive content of variables and how that evidence is affected by data revisions. The question is, does the predictability of one variable for another hold up in real time? Are forecasts based on models that show predictability based on latest available data useful for forecasting in real time?

To address the first question, Amato-Swanson (2001) used the latest-available data to show that M1 and M2 have predictive power for output. But using real-time data, that predictability mostly disappears; many models are improved by *not* including measures of money.

To address the second question, Croushore (2005) investigated whether indexes of consumer sentiment or confidence based on surveys matter for forecasting consumption spending in real time; previous research found them of marginal value for forecasting using latest-available data. His results showed that consumer confidence measures are not useful in forecasting consumption; in fact, in some specifications, forecasting performance is worse when the measures are included.

In summary, the predictive content of variables may change because of data revisions, according to the small amount of research that has been completed in this area.

V. Optimal Forecasting when Data Are Subject to Revision

Having established that data revisions affect forecasts, in this section we examine the literature that discusses how to account for data revisions when forecasting. The idea is that a forecaster should deal with data revisions in creating a forecasting model. The natural venue for doing so is a model based on the Kalman filter or a state-space model. (This chapter will not discuss the details of this modeling technique, which are covered thoroughly in the chapter by Harvey on "Unobserved Components Models" in this volume.)

The first paper to examine optimal forecasting under data revisions is Howrey (1978). He showed that a forecaster could adjust for different degrees of revision using the Kalman filter. He ran a set of experiments to illustrate.

In experiment 1, Howrey forecasted disposable income using the optimal predictor plus three methods that ignored the existence of data revisions, over a sample from 1954 to 1974. He found that forecast errors were much larger for non-optimal methods (those that ignored the revision process). He suggested that new unrevised data should be used (not ignored) in estimating the model, however, but the new data should be adjusted for bias and serial correlation. In experiment 2, Howrey forecasted disposable income and consumption jointly, finding the same results as in experiment 1.

Harvey-McKenzie-Blake-Desai (1983) considered how to optimally account for irregular data revisions. Their solution was to use state-space methods to estimate a multivariate ARMA model with missing observations. They used U.K. data on industrial production and wholesale prices from 1965 to 1978. Their main finding was that there was a large gain in relative efficiency (MSE) in using the optimal predictor rather than assuming no data revisions, with univariate forecasts. With multivariate forecasts, the efficiency gain was even greater. The method used in this paper assumes that there are no revisions after *M* periods, where *M* is not large, so it may not be valid for all variables.

Other papers have found mixed results. Howrey (1984) examined forecasts (using state-space methods) of inventory investment, and found that data errors are not responsible for much forecast error at all, so that using state-space methods to improve the forecasts yields little improvement. Similarly, Dwyer-Hirano (2000) found that state-space methods perform worse than a simple VAR that ignores revisions, for forecasting levels of M1 and nominal output.

One key question in this literature is that of which data set should a forecaster use, given so many vintages and different degrees of revision? Koenig-Dolmas-Piger (2003) attempted to find the optimal method for real-time forecasting of current-quarter output growth. They found that it was best to use first-release data rather than real-time data, which differs from other papers in the literature. This is similar to the result found earlier by Mariano-Tanizacki (1995) that combining preliminary and revised data is sometimes very helpful in forecasting. Patterson (2003) illustrated how combining the data measurement process and the data generation process improved forecasts, using data on U.S. income and consumption.

These papers suggest that there sometimes seems to be gains from accounting for data revisions, though not always. However, some of the results are based on data samples from further in the past, when the data may not have been of as high quality as data today. For example, past revisions to industrial production were clearly predictable in advance, but that predictability has fallen considerably as the Federal Reserve

Board has improved its methods. If the predictability of revisions is low relative to the forecast error, then the methods described here may not be very helpful. For example, if the forecastable part of data revisions arises only because seasonal factors are revised just once per year, then the gains from forecasting revisions are quite small. Further, research by Croushore-Stark (2001) and (2003) suggests that the process followed by revisions is not easily modeled as any type of AR or MA process, which many models of optimal forecasting with data revisions require. Revisions appear to be non-stationary and not well approximated by any simple time-series process, especially across benchmark vintages. Thus it may be problematic to improve forecasts, as some of the literature suggests. In addition, improvements in the data collection process because of computerized methods may make revisions smaller now than they were in the past, so using methods such as the Kalman filter may not work well.

One possible remedy to avoid issues about revisions altogether is to follow the factor model approach of Stock-Watson (1999), explained in more detail in the Stock-Watson chapter on "Forecasting with Many Predictors" in this volume. In this method, many data series, whose revisions may be orthogonal, and combined and one or several common factors are extracted. The hope is that the revisions to all the data series are independent or at least not highly correlated, so the estimated factor is independent of data revisions, though Stock-Watson did not test this because they would have needed real-time data on for more variables than are included in the Real-Time Data Set for Macroeconomists. The only test extant of this idea (comparing forecasts from a factor model based on real-time data compared with latest available data) is provided by Bernanke-Boivin (2003). They found that for the subsample of data for which they had both real-time and latest available data, the forecasts made were not significantly different, suggesting that the factor model approach is indeed promising for eliminating the effects of data revisions. However, their results could be special to the situation they examined; additional research will be needed to see how robust their results are.

Another related possibility is for forecasters to recognize the importance of revisions and to develop models that contain both data subject to revision and data that are not subject to revision, such as financial market variables. This idea has not yet been tested in a real-time context to see how well it would perform in practice.[1]

In summary, there are sometimes gains to accounting for data revisions; but predictability of revisions (today for US data) is small relative to forecast error (mainly seasonal adjustment). This is a promising area for future research.

V. Summary and Suggestions for Further Research

This review of the literature on forecasting and data revisions suggests that data revisions may matter for forecasting, though how much they matter depends on the case at hand. We now have better data sets on data vintages than ever before, and researchers in many other countries are attempting to put together real-time data sets for macroeconomists like that in the United States. What is needed now are attempts to systematically categorize and evaluate the underlying determinants of whether data revisions matter for forecasting, and to develop techniques for optimal forecasting that are consistent with the data process of revisions. This latter task may be most difficult, as characterizing the process followed by data revisions is not trivial. A key unresolved

[1] Thanks to an anonymous referee for making this suggestion.

issue in this literature is: What are the costs and benefits of dealing with real-time data issues versus other forecasting issues?

REFERENCES

Amato, Jeffery D. and Norman R. Swanson, "The Real Time Predictive Content of Money for Output," *Journal of Monetary Economics* 48, (2001) pp. 3–24.

Bernanke, Ben, and Jean Boivin. "Monetary Policy in a Data-Rich Environment." *Journal of Monetary Economics* 50 (2003), pp. 525–546.

Cole, Rosanne, "Data Errors and Forecasting Accuracy," in Jacob Mincer, ed., *Economic Forecasts and Expectations: Analyses of Forecasting Behavior and Performance*. New York: National Bureau of Economic Research, 1969, pp. 47-82.

Croushore, Dean, and Tom Stark. "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105 (November 2001), pp. 111-130.

Croushore, Dean, and Tom Stark, "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?" *Review of Economics and Statistics* 85 (August 2003), pp. 605–617.

Croushore, Dean. "Do Consumer Confidence Indexes Help Forecast Consumer Spending in Real Time?" Conference volume for conference on "Real-Time Data and Monetary Policy," Eltville, Germany; published in *North American Journal of Economics and Finance* (2005, forthcoming).

Denton, Frank T., and John Kuiper. "The Effect of Measurement Errors on Parameter Estimates and Forecasts: A Case Study Based on the Canadian Preliminary National Accounts," *Review of Economics and Statistics* 47 (May 1965), pp. 198-206.

Diebold, Francis X., and Glenn D. Rudebusch. "Turning Point Prediction With the Composite Leading Index: An Ex-Ante Analysis," in *Leading Economic*

Indicators: New Approaches and Forecasting Records, eds. K. Lahiri and G.H. Moore, Cambridge, U.K.: Cambridge University Press, 1991a, pp. 231-56.

Diebold, Francis X., and Glenn D. Rudebusch. "Forecasting Output With the Composite Leading Index: A Real-Time Analysis," *Journal of the American Statistical Association* 86 (September 1991b), pp. 603-10.

Dwyer, Mark, and Keisuke Hirano. "Optimal Forecasting Under Data Revisions," working paper 2000.

Edge, Rochelle M., Thomas Laubach, and John C. Williams. "Learning and Shifts in Long-Run Productivity Growth," manuscript, March 2004.

Elliott, Graham. "Comments on 'Forecasting with a Real-Time Data Set for Macroeconomists,'" *Journal of Macroeconomics* 24 (December 2002), pp. 533-539.

Faust, Jon, John H. Rogers, and Jonathan H. Wright. "Exchange Rate Forecasting: the Errors We've Really Made." *Journal of International Economics* 60 (2003), pp. 35-59.

Harrison, Richard, George Kapetanios, and Tony Yates. "Forecasting With Measurement Errors in Dynamic Models," working paper 2002.

Harvey, A.C., C.R. McKenzie, D.P.C. Blake, and M.J. Desai. "Irregular Data Revisions," in Arnold Zellner, ed., *Applied Time Series Analysis of Economic Data*. Washington, D.C.: U.S. Department of Commerce, Economic Research Report ER-5, 1983, pp. 329–347.

Howrey, E. Philip. "The Use of Preliminary Data in Econometric Forecasting," *Review of Economics and Statistics* 60 (May 1978), pp. 193-200.

Howrey, E. Philip. "Data Revision, Reconstruction, and Prediction: An Application to Inventory Investment." *Review of Economics and Statistics* 66 (1984), pp. 386–93.

Howrey, E. Philip. "Forecasting GNP With Noisy Data: A Case Study," *Journal of Economic and Social Measurement* 22 (1996), pp. 181-200.

Koenig, Evan, Sheila Dolmas, and Jeremy Piger. "The Use and Abuse of 'Real-Time' Data in Economic Forecasting," *Review of Economics and Statistics* 85 (2003).

Kozicki, Sharon. "Comments on: 'Forecasting with a Real-Time Data Set for Macroeconomists.'" *Journal of Macroeconomics* 24 (December 2002), pp. 541–558.

Mariano, Roberto S., and Hisashi Tanizaki. "Prediction of Final Data with Use of Preliminary and/or Revised Data." *Journal of Forecasting* 14 (1995), pp. 351–380.

Orphanides, Athanasios. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91 (September 2001), pp. 964–985.

Orphanides, Athanasios, and Simon van Norden. "The Reliability of Inflation Forecasts Based on Output Gaps in Real Time." *Journal of Money, Credit, and Banking* 37 (June 2005), pp. 583–601.

Patterson, K.D. "Exploiting Information in Vintages of Time-Series Data," *International Journal of Forecasting* 19 (2003), pp. 177-197.

Robertson, John C., and Ellis W. Tallman. "Data Vintages and Measuring Forecast Model Performance." Federal Reserve Bank of Atlanta *Economic Review* (Fourth Quarter 1998), pp. 4-20.

Stark, Tom, and Dean Croushore. "Forecasting with a Real-Time Data Set for Macroeconomists," *Journal of Macroeconomics* 24 (December 2002), pp. 507–31. Also, a "Reply" to formal comments, pp. 563–7.

Stock, James M., and Mark W. Watson. "Forecasting Inflation." *Journal of Monetary Economics* 44 (1999), pp. 293–335.

Swanson, Norman. "Forecasting Using First-Available Versus Fully Revised Economic Time-Series Data," *Studies in Nonlinear Dynamics and Econometrics* 1 (April 1996), pp. 47-64.

Swanson, Norman R. and Halbert White. "A Model Selection Approach To Real-Time Macroeconomic Forecasting Using Linear Models And Artificial Neural Networks," *Review of Economics and Statistics* (November 1997), pp. 540-550.

Trivellato, Ugo, and Enrice Rettore. "Preliminary Data Errors and Their Impact on the Forecast Error of Simultaneous-Equations Models," *Journal of Business and Economic Statistics* 4 (October 1986), pp. 445-53.