A Simple Way to Estimate Current-Quarter GNP

Terry J. Fitzgerald
Research Assistant
Federal Reserve Bank of Minneapolis
and Graduate Student
University of Minnesota

Preston J. Miller
Vice President and Deputy Director of Research
Research Department
Federal Reserve Bank of Minneapolis

People interested in the economic outlook—such as financial market participants, monetary policymakers, and macroeconomic forecasters—usually spend a lot of time analyzing current economic conditions. We think the main reason they do this is that forecasts of the future course of the economy usually depend on its current state. While the economy's current state seems best summarized by real gross national product (GNP, adjusted for inflation), the U.S. Department of Commerce doesn't issue its first estimate of real GNP for a given quarter (the *advance* estimate) until late in the first month of the next quarter. So, for example, its advance estimate of first-quarter real GNP isn't available until late in April. Yet, within a given quarter a variety of daily, weekly, and monthly reports are issued on data series like interest rates, unemployment claims, employment, personal income, and housing starts. Thus, analysts who attempt to summarize the current state of the economy are led to predict the Commerce Department's advance estimate of real GNP based on the within-quarter data available before that estimate is released.

We have developed a simple method for predicting real GNP in the current quarter by using three monthly data series on the number of hours employees work. The series are available from the Labor Department's payroll employment survey. The method we've developed has generated relatively accurate forecasts. Moreover, it has interesting implications about how the GNP data are constructed and how models designed to forecast current-quarter data in *real time* (using data available at the actual time of the forecast) should be built.

Motivation for Our Method

Our simple method was developed even though another one was at hand. To forecast advance real GNP we could just have used the primary forecasting model in the Research Department of the Minneapolis Fed. That model is a Bayesian vector autoregression (BVAR) model of the U.S. economy (see the outlook paper by Runkle in this issue). It can generate forecasts for any time span, but experience suggests that its real-time forecasts of current-quarter GNP are not very accurate. This result is hardly surprising, given that the model was constructed to forecast a number of variables over many quarters—not just real GNP in the current quarter. Moreover, the model's construction essentially ignores problems concerning the real-time availability and reliability of data released within a quarter. For the most part, the model is fit based on both a complete data set that has gone through three years of revisions and an objective expressed in terms of similarly revised data. In real time, however, forecasters must base their predictions of current-quarter advance real GNP on incomplete, unrevised data.

We can be more specific about the real-time data problems that the BVAR model ignores. The BVAR model's forecast of current-quarter real GNP gives considerable weight, for instance, to monthly data series corresponding to demand components, such as retail sales, business inventories, and the merchandise trade balance. Using a complete set of data, which for the most part has gone through three years of revisions, the BVAR model's weighting seems reasonable. But in real time, until the week before the release of the advance estimate of real GNP, there is, at most, one month of within-quarter data available for either business inventories or the trade balance. Those data are not adjusted for inflation and are also very unreliable (that is, subject to large revisions). If the objective is to predict the advance estimate of current-quarter real GNP, those data should be given close to zero weight.

We have constructed a simple alternative model to deal with these problems. It has worked well in real-time forecasting of advance real GNP: the model's forecast errors seem small in an absolute sense and are smaller than those from the BVAR model. Based on 24 observations over the last four years, our model's average absolute forecast error in predicting the annual growth rate in advance real GNP is 0.86 percentage points. The corresponding error for predictions made at similar times by the BVAR model is 2.04 percentage points—or 1.18 percentage points larger than our model's error.

Our model, however, has not worked as well going back two years or longer in time to forecast revised real GNP using currently revised data: the forecast errors seem large in an absolute sense and are closer to those from the BVAR model. The standard error of estimate for the annual growth rate of current-quarter real GNP from our model is 2.39 percentage points. Although we were unable to compute the corresponding error for the BVAR model, the standard error is almost surely closer to our model's standard error than was the case with the real-time errors.[1]

From these observations we draw two implications:

- The Commerce Department seems to weigh the hours-worked data most heavily in its early estimates of real GNP but less and less so in its revised estimates.
- Analysts attempting to predict current-quarter outcomes in real time need to consider the availability and reliability of data at the time the forecasts are made.

We believe the first implication can be explained in the following way. For its advance real GNP estimate, the Commerce Department has complete labor input data and only partial demand data. Although the demand components must add up to real GNP, the Commerce Department initially can use considerable latitude in estimating the quarterly values of some of the components. What primarily constrains its advance estimate of real GNP, therefore, are the labor input data. But as a year progresses, the department acquires more complete demand data, giving it less discretion to estimate quarterly values and causing it to revise the total. Then, as more time goes by, the department acquires additional data from benchmark surveys, various types of tax receipt and customs records, and other sources. It again revises its real GNP estimate, making the labor input data even less instrumental.

The second implication stresses that if real-time predictions are made at particular dates in every quarter, analysts must consider how much within-quarter data is available at those dates for each series they plan to include in their model. They should also consider the average revision in each series relative to its preliminary reported change. Data that are sparse or subject to large revisions are unlikely to be very useful in real-time forecasting.

Our Method in Detail

The method we've devised is simple and relies on just a few variables. The only within-quarter data series used are three measures of hours worked: the index of aggregate weekly hours of production or nonsupervisory workers on private nonfarm payrolls; the component of that series for goods-producing industries; and the component for service-producing industries. We chose those series for two reasons. First, the values for all three months within a quarter are reported before the advance real GNP estimate is released. Second, the series tend to be subject to relatively small revisions. We account for the availability of the data by estimating three versions of the model, depending on whether one month, two months, or three months of hours-worked data are available at the time of the forecast.

[1] The closeness of the standard errors is implied by the following reasoning. The BVAR model's one-step-ahead forecast error for real GNP is 3.6 percentage points. That error assumes the BVAR model has no within-quarter data, whereas our model's standard error of 2.4 percentage points assumes three months of hours-worked data. Even with this difference in information, the difference in standard errors nearly matches the real-time difference in average absolute forecast errors (of about 1.2 percentage points). So, if having within-quarter data improves the accuracy of the BVAR model's forecast at all, its standard error of estimate would be closer to that of our model than are the real-time errors.

The basic structure for our method is provided by an identity:

$$(1) \qquad RGNP \equiv H \cdot (O/H) \cdot (RGNP/O)$$

where

$RGNP$ = real GNP

H = hours worked in the private nonfarm sector

O = output produced by H.

The identity states that real GNP is the product of hours worked (H) and productivity (O/H) in the private nonfarm sector multiplied by a scaling factor ($RGNP/O$) that accounts for output produced in other sectors, such as agriculture or government. We take all variables to be quarterly averages and rewrite the identity in terms of quarterly growth at annual rates, denoted by $g(\cdot)$:

$$(2) \qquad g(RGNP) \equiv g(H) + g(O/H) + g(RGNP/O).$$

Before the advance estimate of real GNP is released, we have no within-quarter information on either productivity or the scaling factor. So we make two assumptions that allow us to estimate a self-contained system based only on current hours-worked data and past real GNP. First, we assume that $g(O/H)$ depends on five variables: a constant, trend, hours worked, the ratio of the hours-worked components for goods-producing and service-producing industries ($HRAT$), and real GNP in the previous quarter $(RGNP)_{t-1}$. Second, we assume that $g(RGNP/O)$ depends only on a constant and trend. These assumptions let us estimate an equation for the growth in real GNP in quarter t:

$$(3) \qquad g(RGNP)_t = a + bt + c \cdot g(\bar{H})_t + d \cdot (\overline{HRAT})_t + e \cdot g(RGNP)_{t-1}.$$

Here, a bar (¯) over a variable indicates our estimate of its average value in quarter t.[2]

We use different estimates of H and $HRAT$ with one month, two months, and three months of data in a quarter. When one month of data is available, we estimate

$$(4) \qquad g(\bar{H})_t = r + st + u \cdot \tilde{g}(H)_{t:1} + v \cdot \tilde{g}(H)_{t-1:3} + w \cdot \tilde{g}(H)_{t-1:2} + x \cdot (HRAT)_{t:1} + y \cdot (HRAT)_{t-1:3} + z \cdot (HRAT)_{t-1:2}$$

where

$(X)_{t:i}$ = the ith monthly value of X in quarter t

$\tilde{g}(H)$ = the change at an annual rate in H over its value three months earlier.

We estimate an equation with the same variables on the right side to get $(\overline{HRAT})_t$. We estimate similar equations with two months of data by replacing $\tilde{g}(H)_{t-1:2}$ with $\tilde{g}(H)_{t:2}$ and $(HRAT)_{t-1:2}$ with $(HRAT)_{t:2}$. When three months of data are available, we compute $g(H)_t$ and $(HRAT)_t$ directly.

Its Performance . . .

To determine how well our method performs in real time, we imagined putting ourselves back in time and using data that was then available. We started in the first quarter of 1986. We initially estimated our system from the first quarter of 1964 through the fourth quarter of 1985, using data available at the time the January 1986 employment report was released in early February. (Employment data are reported around the first Friday of the following month.) We then used this model to forecast the growth in advance real GNP in the first quarter of 1986. We next repeated these steps, reestimating the system by using the (possibly revised) data set from the first quarter of 1964 to the time the second employment report was available in early March 1986. We continued this process until we had generated three monthly forecasts of quarterly advance real GNP growth for each quarter from 1986:1 through 1989:3. We computed forecast errors as the actual advance estimates of real GNP growth minus our forecasts.

When we examine these forecast errors, we first find that the errors in predicting the growth in advance real GNP using three months of data tend to be small (see Chart 1). Except for 1989:1, the errors are never off by much more than one percentage point and are well within the band of one standard error of estimate. The pattern of errors from 1988:2 through 1989:1 suggests the drought may have affected these errors. Since our

[2]The way we express the identity, our assumptions, and our description of the equation for estimation is intended to convey our thinking and intuition in the development of our model. The model itself does not rely on our particular development. For instance, we chose to write the identity with three terms, so that (O/H) is productivity in the private nonfarm sector. We could have written the identity as $RGNP \equiv H \cdot (RGNP/H)$, but the second term has no nice economic interpretation. Similarly, we don't estimate $g(O/H)$ and $g(RGNP/O)$ separately, because we only have to estimate their total and therefore don't need separate assumptions about their determinants. Finally, we chose to include only a few lags in our estimated equation—a choice based on experimentation and a desire to keep our system simple.

Charts 1–3

Our Model's Real-Time Forecast Errors of Advance Real GNP*

% Changes at an Annual Rate

Quarterly, 1986:1–1989:3

Chart 1 Using Three Months of Data

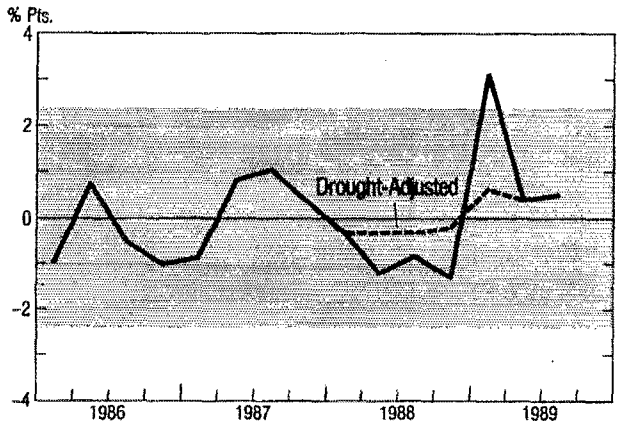
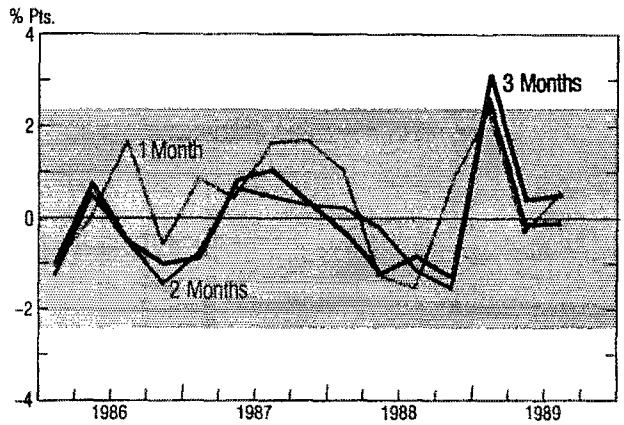


Chart 2 Using One Month, Two Months, and Three Months of Data



*Shaded bands indicate a standard error of estimate of plus or minus 2.387.

model assumes that the agricultural sector maintains a constant relationship to the rest of the economy, these errors aren't surprising. Moreover, we could argue that since the Commerce Department released its drought adjustments for the most part before its GNP reports, our forecasts could have been altered at the time to reflect the effects of the drought.[3] When we do that, the drought-adjusted forecast errors are uniformly small (see Chart 1).

We next find that having more months of hours-worked data within a quarter improves the forecasts of growth in advance real GNP by only a little. Chart 2 shows the forecast errors for the versions of the model having one month, two months, and three months of data. When two months rather than one month of data are used, some improvement in forecast accuracy is evident. But only slight improvement is seen when three months rather than two are used. A simple algebraic explanation may account for this. When quarterly average percent changes are computed, the change in the first month gets a relative weight of 1/2; the second month, a relative weight of 1/3; and the third month, a relative weight of 1/6. Thus, the quarterly average percent change is largely determined by the changes in the first two months in the quarter.[4]

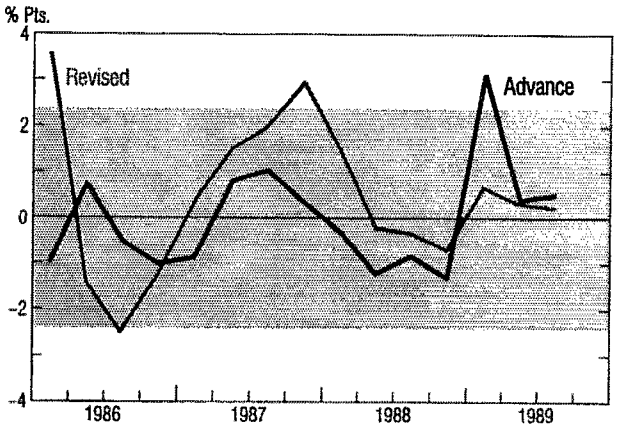
Finally, we find that the model's forecasts deteriorate as we go back in time if we use currently revised data (past data as they are reported today) instead of real-time data (past data as they were reported at the time). We computed forecasts for currently revised real GNP growth using currently revised real GNP and hours-worked data. We then computed forecast errors for revised real GNP growth and compared them to our real-time forecast errors for advance real GNP growth (see Chart 3). Between 1988:3 and 1989:3, the forecast errors using the revised data are even

[3]The Commerce Department announced its drought adjustments for the four quarters from 1988:2 through 1989:1 on September 21, 1988. The announcement came after the 1988:2 advance real GNP estimate was released but well before the estimates for the following three quarters were released.

[4]More precisely,

$$(x_t - x_{t-1})/x_{t-1} = [(\Delta x_{t-1:2} + 2\Delta x_{t-1:3}) + 3\Delta x_{t:1} + 2\Delta x_{t:2} + \Delta x_{t:3}]/3x_{t-1}.$$

Chart 3 Using Three Months of Data vs. Forecast Errors for Revised Real GNP Using Three Months of Revised Data



smaller than the real-time errors.[5] But before 1988:3, the errors grow appreciably and make the one standard error of estimate band look like a reasonable measure of forecast accuracy.

. . . And Implications

The results of this exercise show that there can be a big difference between within-sample goodness of fit (such as the standard error of estimate) and the accuracy of out-of-sample real-time forecasts. Given our model's standard error of estimate, we were continually surprised by how well it forecasted in actual use. Based on the pattern of errors indicated by the revised real GNP estimates (Chart 3), we conjecture that the model's better performance in real time is due to the Commerce Department relying more heavily on hours-worked data in its advance real GNP estimates than in its revised estimates.

In any case, our results have a more general implication for analysts doing real-time forecasting of advance or preliminarily revised data: Models that make real-time forecasts should be designed specially for real-time use. The objective for estimation should be stated with respect to the advance or preliminarily revised data that analysts are trying to predict. And the information on which the forecasts are based should include only the data available at the time of the forecast and should account for the reliability of that data.

[5]Somewhat curiously, the drought seems less important in explaining the forecast errors for the revised data than for the advance data. The 1989:1 error now seems due more to a poor advance estimate by the Commerce Department of drought-adjusted real GNP than to the drought. (The current revised estimate is 1.8 percentage points less than the advance estimate.)

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.