IMPLICATIONS OF REAL-TIME DATA FOR FORECASTING AND MODELING EXPECTATIONS

Sharon Kozicki

NOVEMBER 2001

RWP 01-12

Research Division
Federal Reserve Bank of Kansas City

Sharon Kozicki is an assistant vice president and economist at the Federal Reserve Bank of Kansas City. This paper was prepared in response to a request by the *Journal of Macroeconomics* for comments on "Forecasting with a Real-time Data Set for Macroeconomics" by Tom Stark and Dean Croushore. The views expressed herein are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

sharon.kozicki@kc.frb.org

Abstract

This note extends the analysis in Stark and Croushore (2001) with an emphasis on the importance of data vintage for survey forecasts and modeling expectations. For both of these types of empirical exercises, results suggest that the choice of latest available or real-time data is critical for variables subject to large level revisions, but almost irrelevant for variables subject to only small revisions. Other forecasting practices were examined, with some surprising results.

JEL Classification: C5

Keywords: Forecasting, real-time data, modeling expectations, survey data, data vintage

IMPLICATIONS OF REAL-TIME DATA FOR FORECASTING AND MODELING EXPECTATIONS

1 Introduction

How should forecasting models be evaluated? While the general procedure has become somewhat standardized, specifics are gradually coming under greater scrutiny. The contribution of Stark and Croushore is to point out that the choice of data vintage matters in important ways when making and evaluating forecasts. For instance, forecasts of a variable for a particular date and rankings of models based on properties of forecast errors can both be quite different depending on the vintage of data used.

Until recently, forecasting models were evaluated by comparing forecasts of an economic variable to the actual outcome as reported in the most recent vintage of historical data. However, Stark and Croushore point out that this approach gives recent model developers a considerable advantage: The data they are using have been revised and may differ significantly from the data used by forecasters in real time. For example, by using the most recent vintage of historical data, recent model developers have knowledge about historical changes in methodology used to construct data—information that forecasters in the past could not reasonably have been expected to anticipate.

Stark and Croushore argue that "for purposes such as modeling expectations or evaluating forecast errors of survey data, the use of latest-available data is questionable; comparisons between forecasts generated from new models and benchmark forecasts, generated in real time, should be based on real-time data." The arguments they put forward are compelling. However, while Stark and Croushore examine the differences between real-time and latest-available data in a study of forecast accuracy, they don't actually investigate the importance of data vintage for evaluating survey forecasts or modeling expectations. This note extends the analysis in Stark and Croushore and provides such an investigation.

Data vintage issues are particularly important when evaluating survey forecasts. Forecasts collected by surveys are made conditional on the version of the historical data available at the time of the survey. However, the relationship between the forecasts and the historical data is not recorded by the surveys. For variables subject to considerable revision, it is unreasonable to expect that forecasts would remain unchanged if historical data were revised. Yet, this is implicitly what analysts are doing when they compare survey forecasts to a measure of the actual outcome as reported in the latest available vintage of data. Since forecasts in surveys cannot be revised to reflect revisions in historical data, survey forecasts should be compared to an early release version of the actual outcome. Furthermore, many forecasters, and commercial forecasters in particular, are more interested in forecasting an early release of the data—after all, this is most likely to be the metric their customers will use to evaluate their forecast performance. Thus, it would be more appropriate to evaluate forecasts collected by surveys to a version of actual that is published soon after the date being forecast has passed.

Data vintage issues are likely to be relevant when modeling expectations, because they too are conditioned on the version of the historical data available at the time the expectations are formed. Expectations variables are important in macroeconomic analysis. For instance, investment decisions are based on real funding costs, or the difference between market nominal rates and expected inflation; monetary policy actions have been described as responding to forecasts of economic activity and inflation (Clarida, Gali, and Gertler (1998 and 2000), Kozicki (1999)); and, theories of the term structure of interest rates relate multi-period yields to expectations of future one-period yields. Despite the prevalence of expectations variables in macroeconomic models, few studies have evaluated how well standard empirical proxies approximate expectations.[1]

The next section discusses how data vintage influences forecast evaluation by affecting forecast errors–through conditioning variables, estimates of model coefficients, and measures of actual. An empirical exercise in section 3 explores the implications of data vintage for the evaluation of survey and other forecasts. In addition, a comparison of proxies for expectations that take into account real-time data issues is provided. Concluding comments are offered in section 4.

2 Background

This section reviews the basic steps followed when constructing and evaluating forecasts, emphasizing where data vintage considerations may enter into the process. Forecast evaluation is based on statistical properties of forecast errors, which may be influenced by data vintage in

[1] Kozicki and Tinsley (1998) show how time series models with shifting (or moving) endpoints provide long-horizon forecasts of inflation that are more consistent with survey expectations than more commonly used models with fixed or moving-average endpoints.

several ways. In constructing forecast errors, data vintage is relevant for the measurement of "actuals" against which forecasts are compared and data vintage may affect forecasts. Forecasts are influenced by data vintage because they depend on the vintage of conditioning variables and on the vintage of data used to estimate model coefficients.

The relevance of data vintage can be seen by considering a simple model of the form:

$$y_{t+h}(v_y) = x_t(v_x)\alpha_h(\tau, v_\alpha) + \epsilon_t(v_y, v_x, v_\alpha) \tag{1}$$

where $y_{t+h}(v_y)$ is the value of the variable to be forecast in quarter $t+h$ as recorded in data vintage v_y, $x_t(v_x)$ contains the values of the conditioning variables in quarter t as recorded in data vintage v_x, α_h are estimated model coefficients, and ϵ_t is the regression residual in quarter t.[2] In this notation, h is the forecast horizon, τ refers to the sample period used to estimate the model coefficients, and v_α is a double that contains the vintages of regressors and regressand used during estimation. T is used to reflect the quarter corresponding to the latest available vintage of data. To account for lags in the release of data, the last observation of data is for quarter $T-1$.

This notation is useful to contrast three approaches to constructing forecasts. Although other approaches have been suggested in the literature, this note focuses on the three most common approaches used to construct forecasts: in-sample, out-of-sample, and real-time out-of-sample.[3]

The *in-sample* approach to forecasting uses the full sample of latest available data to both estimate and evaluate a forecasting model. The model is estimated only once, using the full sample of data. Consequently, evaluation of "forecasts" from the model is based on observations that were also in the sample used to estimate the model. In-sample forecasts of y in quarter $t+h$ conditional on data through quarter t are constructed as:

$$f^I_{t+h|t} = x_t(T)\alpha_h(T-1, (T,T)) \qquad t \leq T-1 \tag{2}$$

where dependence of α_h on $T-1$ reflects that the full sample of data including observations through

[2]The notation in this note differs from that used by Stark and Croushore. To reconcile the two representations, note that in general $y_t(v_y)$ in this note corresponds to $Y_{t,v}$ in Stark and Croushore. The notation used in this note allows for the possibility that regressands and regressors may be from different vintages of data.

[3]For all three approaches examined here, $v_x = v_y$. Hoever, such a restriction is not necessary. Koenig, Dolmas, and Piger (2001) consider alternative approaches that use real-time data. In addition to considering cases for which $v_x \neq v_y$, they argue that analysts should generally use data of as many different vintages as there are in their samples each time a model is estimated. In particular, they recommend that at every date within a sample, data on conditioning variables out to be measured as they would have been at that time.

quarter $T-1$ of y and $T-h-1$ of x are used during estimation, and dependence on the double (T,T) reflects that observations on both x and y are vintage T data.

The typical *out-of-sample* approach to forecasting also uses latest available data for estimation and forecasting. The difference between in-sample and out-of-sample approaches is that for out-of-sample approaches, model coefficients are estimated using only observations from before the forecast period so that forecast evaluation is based on forecast errors calculated using observations from outside the estimation sample.[4] To forecast y in quarter $t+h$ conditional on data through quarter t, estimates of α_h use data on y through quarter t and data on x through quarter $t-h$. Out-of-sample forecasts are constructed using vintage T data as:

$$f^{O}_{t+h|h} = x_t(T)\alpha_h(t,(T,T)) \qquad t \leq T-1. \tag{3}$$

Forecast errors are typically calculated using an iterative procedure. An out-of-sample evaluation interval is chosen. A forecast for the first period of the out-of-sample evaluation interval is computed using model coefficients estimated using only data prior to the evaluation interval. Then, the estimation sample is extended by one observation, the model is reestimated, and a forecast for the next period of the out-of-sample evaluation interval is computed. This procedure repeats until forecasts have been constructed for each period of the interval over which forecast performance is to be evaluated.

Real-time out-of-sample forecasts use real-time data for estimation and forecasts. As in the typical out-of-sample approach, forecasts are constructed by iterating over estimation, forecast construction, and extension of the estimation sample. The only difference between the typical out-of-sample approach and the real-time out-of-sample approach is that each time the estimation sample is extended a quarter, a new vintage of data is used to estimate the forecasting model and to compute the out-of-sample forecast. Thus, real-time out-of-sample forecasts of y in quarter $t+h$ conditional on data through quarter t are constructed as:

$$f^{R}_{t+h|t} = x_t(t+1)\alpha_h(t,(t+1,t+1)), \qquad t \leq T-1 \tag{4}$$

where both conditioning variables and estimates of model coefficients are drawn from data vintage

[4]Although coefficient estimates only use observations from before the forecast period, the use of latest available data implies that this approach does not replicate a procedure that could have been followed in real time. To emphasize this feature of the approach, Stock and Watson (2001) refer to out-of-sample forecasting exercises that use latest available data as *simulated* out-of-sample forecasting.

$t + 1$. This approach comes closest to approximating the procedure followed by forecasters in real time.

Data vintage influences forecasts because the observations of conditioning variables may differ across data vintages *and* estimates of model coefficients may be affected by the data vintages of the regressand and regressors. Stark and Croushore show that data vintage may matter for model specification issues—such as how many lags of y to include in x. They also show that out-of-sample forecasts of y in a given quarter can vary widely for different vintages of data. However, they did not examine how much the variation in x across vintages and the variation in estimates of α_h across vintages each contributed to variation in forecasts across vintages. Such a decomposition is also beyond the scope of this note and is left for future research.

Data vintage is also relevant for the choice of actual to which forecasts are compared. Forecast evaluation is based on an analysis of the properties of forecast errors, constructed as the difference between "actual" y in $t + h$ and forecasts of y in $t + h$. To examine whether the choice of data vintage used as actual is important when computing forecast errors and evaluating the relative performance of alternative forecasts, Stark and Croushore use three different measures of actual.[5] An alternative choice of actual is a measure of market expectations, such as the median or mean of a survey of forecasts. Such a choice is relevant for assessing the ability of forecasts from a given model to approximate market expectations.

Forecasts are evaluated by comparing summary statistics of forecast errors for different forecasts. Examples of summary statistics that measure forecast performance include the mean error, the mean absolute error (MAE), and the root mean square error (RMSE). Better forecasts have mean errors closer to zero and smaller MAEs and RMSEs. Usually, most studies emphasize performance as measured by RMSE. Using these metrics, the forecast performance of a proposed model is often compared to that of alternative models. One common alternative is a naive forecasting model that asserts that the variable being forecast will be constant over the forecast horizon at its last observed value. A second alternative is the median or mean of forecasts from a survey. Sometimes, the benchmark for comparison is a forecasting model proposed by an earlier study. Since Stark and Croushore were more interested in examining the implications of real-time data

[5] Stark and Croushore use latest-available data (vintage T), the last vintage before a benchmark release, and the vintage of data four quarters after their four-step-ahead forecasts.

issues for forecasting, they did not provide such benchmarks for comparison. In this note, the median forecast from the Survey of Professional forecasters is provided as an alternative forecast.

3 Empirical Results

This section compares the relative performance of four forecasts. Three are based on an AR(4) forecasting model and differ according to the choice of data vintage and the sample over which the model is estimated. The three AR(4) forecasts correspond to the three approaches to constructing forecasts that were discussed in section 2: in-sample ($f^I_{t+h|t}$), out-of-sample ($f^O_{t+h|t}$), and real-time out-of-sample ($f^R_{t+h|t}$). The fourth forecast is the median forecast from the Survey of Professional Forecasters. These four forecasts are evaluated against three measures of actual: first release data ($y_{t+h}(t+h+1)$), latest available data ($y_{t+h}(T)$), and the median forecast from the Survey of Professional Forecasters.

Four variables are examined: real output growth, the real consumption share, output price inflation, and the unemployment rate.[6] The real-time data were obtained from the the Real-Time Data Set for Macroeconomists, available from the Federal Reserve Bank of Philadelphia. These data are described in Stark and Croushore and, in greater detail, in Croushore and Stark (forthcoming). The median of forecasts from the Survey of Professional Forecasters published by the Federal Reserve Bank of Philadelphia provides a measure of the expectations of economic agents. Two forecast horizons are examined: one step-ahead, and four-steps ahead. Data is quarterly, so one-step ahead and four-step ahead forecasts correspond to one-quarter ahead and four-quarter ahead forecasts respectively.

Stark and Croushore examine real output growth and output price inflation. The real consumption share and the unemployment rate are considered here as well because they are subject to different degrees of revision. Stark and Croushore illustrate that both real output growth and

[6] Real output growth is defined as $400 * log(real\ output(t)/real\ output(t-1))$. Real output growth is measured using real GNP early in the sample and real GDP later in the sample. The real consumption share is defined as $100{*}log(real\ consumption/real\ output)$. Output price inflation is defined as $400{*}log(output\ price(t)/output\ price(t-1))$. Beginning with the February 1996 vintage data set, the output price is defined as the GDP chain-weighted price index. For earlier vintages, the output price is defined as the implicit price deflator, i.e., the ratio of the measure of nominal output to real output. The Survey of Professional Forecasters provided median forecasts of nominal output, real output, real consumption and the unemployment rate. Thus median forecasts of real output growth, the real consumption share, and output price inflation were constructed by applying the transformations described above to the median forecasts. Thus, for example, the "median survey forecast" of output growth is constructed as $400 * log(median\ output\ forecast(t)/median\ output\ forecast(t-1))$.

output price inflation are subject to sizable revisions. However, the mean levels of these series do not change very much with revisions. Means of the four macroeconomic variables being analyzed were calculated over 1960:Q1 through 1981:Q2 for each data vintage dated 1981:Q3 through 2001:Q3. The standard error over vintages of the inflation means was 0.16 percentage point and the standard error of the output growth means was 0.14 percentage point. By contrast, as shown in Figure 1, the level of the real consumption share differs considerably for different data vintages.[7] The standard error over vintages of the real consumption share means was 2.02 percentage points. At the other extreme, revisions to the unemployment rate are very small. The standard error over vintages of the unemployment rate means was less than 0.002 percentage point.

By expanding the list of variables and also including information on median survey forecasts, this note extends the analysis of Stark and Croushore and examines the robustness of some of their results. In addition, the paper reviews whether some well-accepted views and practices are justified.

Tables 1 through 4 contain results for real output growth, the real consumption share, output price inflation, and the unemployment rate, respectively. Tables contain mean errors and root mean square errors (RMSE) for each of the three different measures of actual. Results are summarized in three pairs of columns, with each pair of columns providing results for a different measure of actual. The first and second columns contain results for actual measured using latest available data (2001:Q3 vintage data) and first release data ($y_{t+h}(t+h+1)$), respectively. The third pair of columns compares forecasts to the median forecast from the Survey of Professional Forecasters. Each table contains two panels of results, one with results for 1-step ahead forecasts ($h = 1$) and the second with results for 4-step ahead forecasts ($h = 4$).[8] In each panel results are included for three forecast evaluation intervals, 1981:Q3 - 2000:Q3 and two subsamples, 1981:Q3 - 1991:Q2 and 1991:Q3 - 2000:Q3.[9]

The discussion of the results is organized using a selection of questions. The first question revisits some issues addressed by Stark and Croushore. The next two questions check whether the

[7] The real consumption share is negative because it is defined as $100 * log(real\ consumption/real\ output)$. Although the ratio ($real\ consumption/real\ output$) is a positive fraction, the natural logarithm transformation results in a negative real consumption share. Nevertheless, differences between different vintages of data and changes in levels of a given vintage of data are measured in percentage points.

[8] Four-step ahead forecasts were constructed recursively from an estimated AR(4) model.

[9] In all cases, models were estimated using data starting in 1960:Q2.

results in this note are consistent with some commonly held views about forecasting. The third and fourth questions examine the performance of median survey forecasts and the fifth through seventh questions address the modeling of expectations. A summary of the questions and answers is provided in Table 5, with details provided in the following discussion.

(1) Are real-time out-of-sample forecasts or out-of-sample forecasts based on latest available data better?

Stark and Croushore expected that using latest available data to forecast and latest available data as actuals would lead to smaller forecast errors than using real-time data to forecast and latest available data as actuals. However, they found little difference between out-of-sample forecasts based on latest available data and real-time out-of-sample forecasts when forecasts were evaluated over long intervals. Over short intervals, they found larger differences. For three of the variables examined here, differences were small as well. However, for the real consumption share, the variable subject to the largest revisions, the differences between RMSEs were huge. Measuring actual with latest available data, for one-step ahead forecasts over 1981:Q3 through 2000:Q3, the out-of-sample forecast had an RMSE of 0.62 percent while the real-time out-of-sample forecast had an RMSE of 2.29 percent. When actuals are measured using first release data instead, the out-of-sample RMSE jumps to 2.18 percent and the real-time out-of-sample RMSE plummets to 0.85 percent. For the real consumption share, latest available data is relatively effective at forecasting latest available data and real-time data is relatively effective at forecasting early release data. These results suggest that if historical revisions are larger then the potential differences between out-of-sample forecasts and real-time out-of-sample forecasts are also larger and the choice of data vintage to measure actuals becomes more important.

(2) Do in-sample forecasts perform better than out-of-sample forecasts?

It is generally believed that in-sample forecasts will perform better than out-of-sample forecasts (at least when actuals are measured using latest available data) because observations to be forecast are included in the estimation sample. Empirical results confirm that in-sample forecasts had smaller RMSEs than out-of-sample forecasts in all cases when actuals were measured using latest available data. In-sample forecasts of all variables other than the real consumption share also had smaller RMSEs when actuals were measured using first release data. However, while the in-sample forecasts were generally better, the improvement over out-of-sample forecasts tended to be very small. The largest improvement (over the full sample) for 1-step ahead forecasts was for output growth and the improvement was 0.11 percentage point. For 4-step ahead forecasts, the largest improvement was for inflation and the improvement was 0.14 percentage point.

(3) Do median survey forecasts outperform real-time out-of-sample forecasts?
It is commonly believed that a combination of forecasts tends to forecast better than a single forecast based on a more limited source of information. In this case, the real-time out-of-sample forecast is very constrained in terms of information used to generate the forecast–only historical data on the series being forecast is used. The median survey forecast is implicitly using a much richer dataset that includes all historical data. The empirical results provide some support for the preference of the median survey forecast. One-step ahead median survey forecasts have smaller RMSEs than real-time out-of-sample AR(4) forecasts for all variables. However, for four-step ahead forecasts, median survey forecasts only have smaller RMSEs than the real-time out-of-sample forecasts for the unemployment rate and the real consumption share (in the early sample for the latter).

(4) Do median survey forecasts perform relatively better when compared to first release data rather than latest available data?
Because forecasters don't try to forecast changes in methodology (particularly those that haven't been announced) it seems reasonable to argue that forecasters try to forecast early releases of data. This argument suggests that survey forecasts should perform better relative to other forecasts when actuals are measured using first release data.

For 1-step ahead forecasts of real output growth in the early sample, the median survey strongly outperforms other forecasts regardless of the measure of actual. However, the improvement over the other forecasts is larger for first release data (even compared to the real-time out-of-sample forecast). For 4-step ahead forecasts, results differ across subsamples, but, in general, there is no clear difference in performance of the three AR specifications. This result may reflect the tendency of forecasts to revert toward estimates of trend growth at longer horizons. Since changes in average growth rates across vintages are small, it is not surprising that no one model strongly dominates at the 4-step horizon.

The choice of vintage to measure actuals is critical for evaluation of forecasts of the real consumption share. When latest available data are used to measure actuals, real-time out-of-sample forecasts and median survey forecasts have large magnitude mean errors and large RMSEs, while in-sample and out-of-sample forecasts have small mean errors and small RMSEs. However, these results flip when first release data are used as actuals. Similar results are obtained for one-step-ahead and four-step-ahead forecasts. Since the real consumption share is the series with the largest revisions to the mean level of the series, it should not be surprising that the vintage used to measure actual is very important.

The relative performances of the median survey forecasts of output price inflation and the unemployment rate do not clearly depend on how actuals are measured. This result shouldn't be surprising for the unemployment rate since revisions to the unemployment rate are small.

(5) Do AR models that generate better forecasts also do a better job at matching median survey forecasts?

Better forecasts (in the sense of smaller RMSEs and smaller mean errors) might be expected to provide better proxies for survey forecasts constructed in real time. However, when real-time data differs from latest available data, this view seems more likely to hold only when forecast performance is evaluated relative to actuals measured using first release data.

Empirical results support this theory for real output growth forecasts, although the evidence isn't strong. Real-time out-of-sample forecasts are the closest to matching median survey forecasts and are at least as good as the other AR forecasts when compared to first release data. The evidence isn't strong, however, because the difference in performance of the three AR forecasts when compared to first release data are very small (the largest difference between RMSEs is 0.13 percentage points). Also, the importance of evaluating forecast performance relative to first release data is supported by the observation that real-time out-of-sample forecasts do not provide the best forecast of latest available data.

Real-time out-of-sample forecasts of the real consumption share are much better than the other AR forecasts when compared to first release data and much better at matching median survey forecasts. The huge differences between the performance of the real-time out-of-sample forecasts and the other two forecasts is due to the large revisions in the level of the real consumption share that show up in the different vintages of data. Another consequence of the large revisions is that real-time out-of-sample forecasts are much worse than the other AR forecasts at forecasting latest available data.

For output price inflation the hypothesis that better forecasts are better expectations proxies does not hold. Interestingly, for four-step ahead forecasts in the early sample, in-sample forecasts provided the best forecast of first release data but the worst proxy for median survey expectations, while real-time out-of-sample forecasts provided the worst forecast of first release data but the best proxy for median survey expectations.

For the unemployment rate, better forecasts of first release data turn out to provide better proxies for survey forecasts. However, the differences between the performance of all forecasts are quite small.

(6) Are out-of-sample forecasts or in-sample forecasts closer to survey expectations?

A common approach to proxy for expectations is to use out-of-sample forecasts, typically with latest available data. This approach addresses the standard criticism that in-sample predictions are using information that wouldn't have been available to economic agents in real time and that out-of-sample forecasts correct for this. Of course, this isn't necessarily true. If data is revised and out-of-sample forecasts are based on latest available data, then the version of data being used to construct the forecast would not have been available to economic agents in real time. Ignoring for now the real-time data issue, it is interesting to know whether in fact out-of-sample forecasts do provide a better proxy for survey expectations.

A review of the results suggests that, while out-of-sample forecasts have smaller RMSEs than in-sample forecasts in more than half of the cases, the differences are very small. The largest difference is for 4-step ahead forecasts of inflation in the early subsample, where the RMSE for the in-sample AR forecast is 0.92 and the out-of-sample RMSE is 0.77—a difference of 0.15 percentage point. In most cases, the difference is less than 0.1 percentage point.

(7) Are real-time out-of-sample forecasts or forecasts based on latest available data closer to median survey forecasts?

The expectation is that forecasts based on real-time data would provide better proxies for survey forecasts than forecasts based on latest available data because the latter include revisions that were not known at the time of the surveys. In the tables, both in-sample and out-of-sample forecasts use latest available data, so they might be expected to provide poorer proxies for expectations.

The answer to this question also depends on the variables being analyzed. For both real output growth and the real consumption share, real-time out-of-sample forecasts appear to provide clearly superior proxies for 1-step ahead expectations. Once again, the results are particularly strong for the real consumption share, where the RMSE is about 75 percent smaller (in some cases on the order of 2 percentage points) for real-time out-of-sample forecasts than for forecasts based on latest available data.

For the unemployment rate, forecasts based on latest available data provide better proxies for median survey forecasts than real-time out-of-sample forecasts. However, the penalty to using real-time forecasts is on the order of 0.05 percentage point—less than the precision of the reported data and, thus, not economically significant.

4 Conclusions

This note extended the analysis in Stark and Croushore with an emphasis on the importance of data vintage for evaluating survey forecasts and for modeling expectations. In addition, the justification for some widely accepted practices was reviewed.

The main result reinforces that of Stark and Croushore. The choice between latest available and real-time data matters in important ways. When evaluating forecasts or modeling expectations of variables subject to large level revisions (such as the real consumption share), the choice of latest available or real-time data is critical. Of course, for variables that are subject to only small revisions (such as the unemployment rate) or to no revisions, data vintage issues are almost irrelevant.

The paper also showed that median survey forecasts tend to perform better than simple AR forecasts, particularly at a 1-step horizon. At a 4-step horizon, the relative performance of median survey forecasts deteriorated for more variable series, such as real output growth and output price inflation. For evaluating the relative performance of median survey forecasts, the choice of whether to use first release or latest available to measure actuals was only important for the real consumption share. Once again, data vintage appears to matter more for variables subject to large level revisions.

Several results on modeling expectations were obtained. When compared against first release data, AR models that generated better forecasts also did a better job at matching median survey forecasts for 3 out of 4 variables. For the same 3 variables, real-time out-of-sample forecasts came the closest to the median survey forecasts. Finally, if limited to latest available data, there appeared to be little advantage to following the standard practice of using out-of-sample forecasts rather than in-sample forecasts to proxy for expectations. The performance of in-sample forecasts and out-of-sample forecasts was similar.

References

Clarida, Richard, Jordi Gali, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics,* 115, 147 - 180.

Clarida, Richard, Jordi Gali, and Mark Gertler, "Monetar Policy Rules in Practice: Some International Evidence," *European Economic Review,* 42, 1033-1067.

Croushore, Dean and Tom Stark, forthcoming, "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics.*

Koenig, Evan F., Sheila Dolmas, and Jeremy Piger, 2001, "The Use and Abuse of 'Real-time' data in Economic Forecasting," mimeo, Federal Reserve Bank of Dallas.

Kozicki, Sharon and P.A. Tinsley, 1998, "Moving Endpoints and the Internal Consistency of Agents' Ex Ante Forecasts," *Computational Economics,* 11, 21-40.

Stark, Tom and Dean Croushore, 2001, "Forecasting with a Real-Time Data Set for Macroeconomists," Federal Reserve Bank of Philadelphia Working Paper 01-10.

Stock, James H. and Mark W. Watson, 2001, "Forecasting Output and Inflation: The Role of Asset Prices," Mimeo.

Table 1: Performance of real output growth forecasts

for alternative measures of "actual"

	Measure of "actual"					
	Latest Available		First release		Median survey	
	Mean	RMSE	Mean	RMSE	Mean	RMSE
Panel a: One-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	-.07	2.47	-.60	2.12	-.94	1.56
Out-of-sample AR	-.05	2.58	-.58	2.19	-.92	1.62
Real-time out-of-sample AR	.30	2.51	-.23	2.11	-.57	1.38
Median survey	.87	2.13	.34	1.51	.00	.00
1981:3 - 1991:2						
In-sample AR	-.26	2.89	-.98	2.57	-1.08	1.90
Out-of-sample AR	-.28	3.05	-1.00	2.68	-1.10	2.01
Real-time out-of-sample AR	.13	2.92	-.59	2.56	-.68	1.75
Median survey	.81	2.20	.09	1.54	.00	.00
1991:3 - 2000:3						
In-sample AR	.14	1.92	-.18	1.50	-.79	1.07
Out-of-sample AR	.20	1.95	-.13	1.51	-.74	1.04
Real-time out-of-sample AR	.49	1.95	.16	1.47	-.45	.82
Median survey	.93	2.06	.61	1.49	.00	.00
Panel b: Four-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	.01	2.37	-.48	2.11	-.53	.97
Out-of-sample AR	.06	2.42	-.43	2.14	-.48	.98
Real-time out-of-sample AR	.43	2.41	-.15	2.06	-.20	.87
Median survey	.54	2.46	.05	2.11	.00	.00
1981:3 - 1991:2						
In-sample AR	-.09	2.74	-.79	2.54	-.28	1.01
Out-of-sample AR	-.08	2.80	-.77	2.58	-.27	1.09
Real-time out-of-sample AR	.26	2.75	-.47	2.45	.03	1.00
Median survey	.20	2.70	-.50	2.43	.00	.00
1991:3 - 2000:3						
In-sample AR	.12	1.90	-.14	1.51	-.79	.91
Out-of-sample AR	.20	1.92	-.06	1.51	-.71	.85
Real-time out-of-sample AR	.46	1.98	.20	1.51	-.45	.71
Median survey	.91	2.16	.65	1.70	.00	.00

Table 2: Performance of real consumption share forecasts

for alternative measures of "actual"

	Measure of "actual"					
	Latest Available		First release		Median survey	
	Mean	RMSE	Mean	RMSE	Mean	RMSE
Panel a: One-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	.09	.59	-.91	2.21	-.95	2.19
Out-of-sample AR	.12	.62	-.88	2.18	-.92	2.17
Real-time out-of-sample AR	1.11	2.29	.11	.85	.07	.34
Median survey	1.04	2.24	.04	.71	.00	.00
1981:3 - 1991:2						
In-sample AR	.11	.68	-2.49	2.74	-2.50	2.69
Out-of-sample AR	.16	.73	-2.44	2.70	-2.45	2.65
Real-time out-of-sample AR	2.70	2.89	.10	.95	.09	.46
Median survey	2.61	2.78	.01	.70	.00	.00
1991:3 - 2000:3						
In-sample AR	.07	.47	.80	1.42	.73	1.48
Out-of-sample AR	.08	.48	.81	1.43	.74	1.49
Real-time out-of-sample AR	-.62	1.37	.12	.74	.05	.13
Median survey	-.66	1.43	.07	.73	.00	.00
Panel b: Four-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	.33	1.14	-.66	2.26	-1.04	2.35
Out-of-sample AR	.44	1.23	-.55	2.22	-.93	2.27
Real-time out-of-sample AR	1.36	2.53	.37	1.52	-.01	.52
Median survey	1.37	2.50	.37	1.38	.00	.00
1981:3 - 1991:2						
In-sample AR	.35	1.36	-2.21	2.78	-2.69	2.86
Out-of-sample AR	.51	1.50	-2.05	2.70	-2.52	2.73
Real-time out-of-sample AR	3.00	3.27	.44	1.71	-.04	.63
Median survey	3.04	3.17	.47	1.43	.00	.00
1991:3 - 2000:3						
In-sample AR	.31	.84	1.01	1.50	.75	1.62
Out-of-sample AR	.35	.87	1.06	1.54	.80	1.64
Real-time out-of-sample AR	-.42	1.33	.29	1.27	.02	.35
Median survey	-.44	1.48	.27	1.32	.00	.00

Table 3: Performance of output price inflation forecasts for alternative measures of "actual"

	Measure of "actual"					
	Latest Available		First release		Median survey	
	Mean	RMSE	Mean	RMSE	Mean	RMSE
Panel a: One-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	-.19	.77	-.16	.94	.17	.49
Out-of-sample AR	-.25	.82	-.22	.97	.11	.48
Real-time out-of-sample AR	-.35	.92	-.32	1.06	.01	.56
Median survey	-.36	.74	-.33	.92	.00	.00
1981:3 - 1991:2						
In-sample AR	-.20	.84	-.00	1.10	.22	.56
Out-of-sample AR	-.29	.89	-.09	1.14	.13	.55
Real-time out-of-sample AR	-.48	1.08	-.29	1.24	-.07	.67
Median survey	-.41	.80	-.22	1.03	.00	.00
1991:3 - 2000:3						
In-sample AR	-.18	.70	-.33	.72	.12	.40
Out-of-sample AR	-.21	.72	-.35	.75	.10	.40
Real-time out-of-sample AR	-.21	.71	-.36	.82	.10	.40
Median survey	-.31	.67	-.45	.78	.00	.00
Panel b: Four-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	-.49	1.03	-.44	1.05	.40	.71
Out-of-sample AR	-.68	1.17	-.63	1.17	.22	.61
Real-time out-of-sample AR	-.73	1.22	-.69	1.21	.16	.59
Median survey	-.90	1.31	-.85	1.28	.00	.00
1981:3 - 1991:2						
In-sample AR	-.54	1.22	-.38	1.20	.61	.92
Out-of-sample AR	-.82	1.41	-.66	1.35	.33	.77
Real-time out-of-sample AR	-.94	1.43	-.79	1.38	.20	.71
Median survey	-1.15	1.57	-.99	1.47	.00	.00
1991:3 - 2000:3						
In-sample AR	-.45	.77	-.52	.85	.18	.39
Out-of-sample AR	-.53	.84	-.60	.92	.10	.35
Real-time out-of-sample AR	-.51	.95	-.58	1.00	.12	.41
Median survey	-.63	.95	-.70	1.04	.00	.00

Table 4: Performance of unemployment rate forecasts

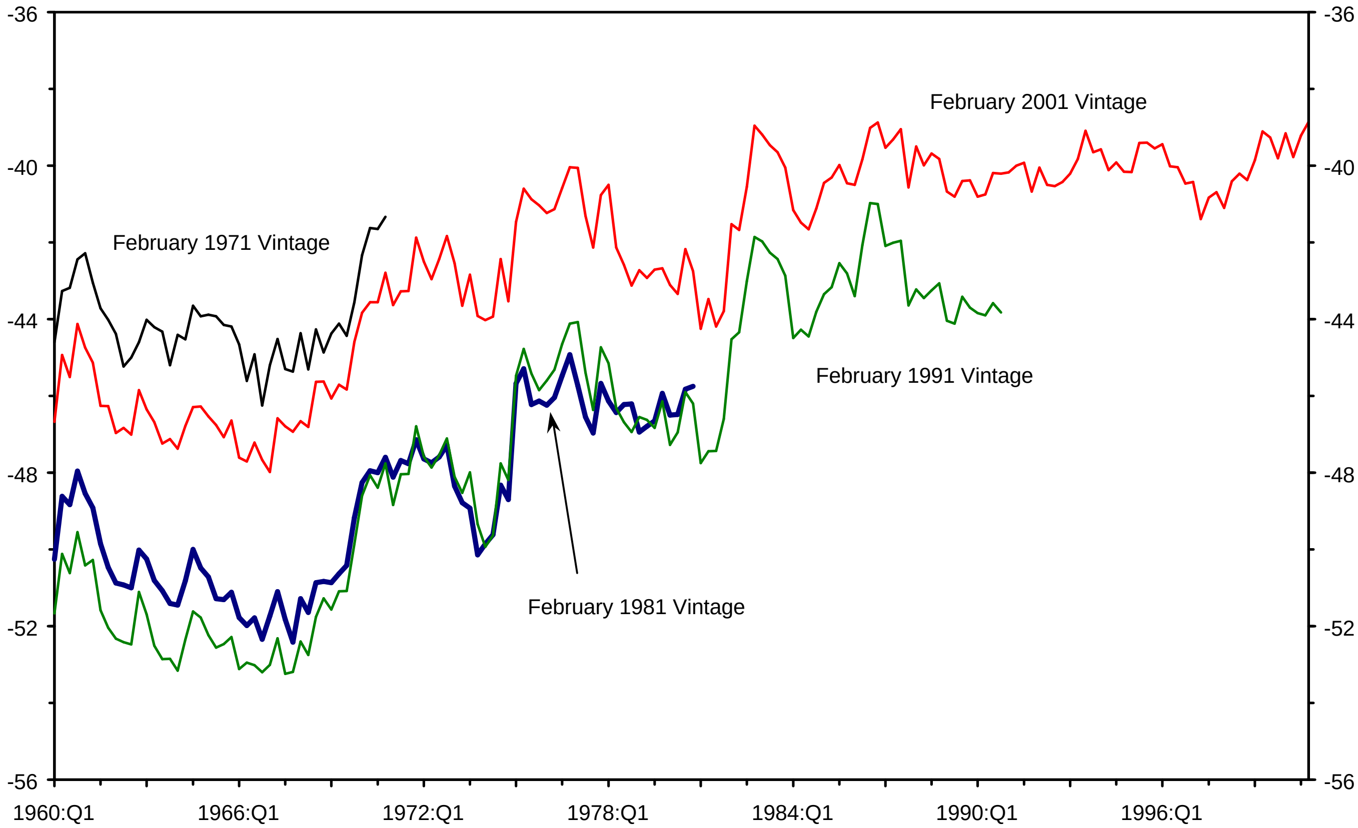
for alternative measures of "actual"

	Measure of "actual"					
	Latest Available		First release		Median survey	
	Mean	RMSE	Mean	RMSE	Mean	RMSE
Panel a: One-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	-.00	.21	-.02	.23	.01	.17
Out-of-sample AR	-.01	.23	-.04	.24	.00	.18
Real-time out-of-sample AR	.01	.27	-.01	.28	.03	.22
Median survey	-.01	.14	-.04	.14	.00	.00
1981:3 - 1991:2						
In-sample AR	.04	.26	.02	.27	.05	.21
Out-of-sample AR	.02	.27	.00	.29	.04	.22
Real-time out-of-sample AR	.04	.32	.03	.34	.06	.26
Median survey	-.01	.16	-.03	.16	.00	.00
1991:3 - 2000:3						
In-sample AR	-.04	.16	-.07	.18	-.03	.13
Out-of-sample AR	-.05	.16	-.08	.18	-.04	.13
Real-time out-of-sample AR	-.03	.20	-.06	.21	-.01	.16
Median survey	-.02	.12	-.05	.12	.00	.00
Panel b: Four-step ahead forecasts						
1981:3 - 2000:3						
In-sample AR	-.04	.76	-.06	.76	.06	.32
Out-of-sample AR	-.13	.84	-.15	.84	-.03	.34
Real-time out-of-sample AR	-.11	.86	-.13	.85	-.01	.38
Median survey	-.10	.69	-.12	.68	.00	.00
1981:3 - 1991:2						
In-sample AR	.20	.93	.18	.93	.21	.37
Out-of-sample AR	.09	1.04	.06	1.04	.09	.39
Real-time out-of-sample AR	.11	1.06	.09	1.05	.12	.45
Median survey	-.01	.85	-.03	.84	.00	.00
1991:3 - 2000:3						
In-sample AR	-.30	.50	-.33	.51	-.11	.25
Out-of-sample AR	-.37	.55	-.39	.56	-.17	.27
Real-time out-of-sample AR	-.35	.57	-.37	.57	-.15	.29
Median survey	-.20	.46	-.22	.46	.00	.00

Table 5: Summary of results

	Output Growth	Consumption Share	Inflation	Unemployment Rate
(1) Are real-time out-of-sample forecasts or out-of-sample forecasts based on latest available data better?	similar	depends on actual	latest available (slight)	latest available (slight)
(2) Do in-sample forecasts perform better than out-of-sample forecasts?	slightly	yes (4-step)	slightly	slightly
(3) Do median survey forecasts outperform real-time out-of-sample forecasts?	yes (1-step)	yes	yes (1-step)	yes
(4) Do median survey forecasts perform relatively better when compared to first release data than latest available data?	slightly	yes	no	no
(5) Do AR models that generate better forecasts also do a better job at matching median survey forecasts?	yes (first release)	yes (first release)	no	yes
(6) Are out-of-sample forecasts or in-sample forecasts closer to median survey forecasts?	similar	similar	similar	similar
(7) Are real-time out-of-sample forecasts or forecasts based on latest available data closer to median survey forecasts?	real-time	real-time	mixed	latest available (slight)

Figure 1: Different Vintages of the Real Consumption Share



Source: Real-time Dataset for Macroeconomists, Federal Reserve Bank of Philadelphia.
Note: Series are constructed as 100 * log(real consumption / real output).