Evaluating BLS labor force, employment, and occupation projections for 2000

*In 1989, BLS first projected estimates for the year 2000
of the labor force, employment, and occupations;
in most cases, the accuracy of BLS projections were
comparable to estimates from naïve extrapolated models*

H.O. Stekler
and
Rupin Thomas

The purpose of any evaluation of economic forecasts is to find the sources of the errors and to improve future forecasts. The errors may result from internal procedures, assumptions, or methods, and from external inputs.[1] Moreover, because the forecasts are intended to be used for some function or purpose, the evaluation should pose questions that determine how well the predictions fulfilled this intended purpose.

Thus, for a forecast evaluation to be valuable, it must pose the right questions that need to be addressed. This is true whether the forecasts are short-term macroeconomic predictions or the long-term BLS projections of labor force, employment, and occupation trends. However, an evaluation of these BLS long-term projections poses three methodological issues that usually are not encountered in analyses of short-term macroeconomic forecasts. First, no other organization made projections of these variables. Consequently, there is no benchmark for judging the BLS forecasts. Second, these projections are long-term rather than the short-term macroeconomic forecasts that have been evaluated in the past. Thus, the questions that must be addressed in this evaluation can differ from those addressed in the macro forecasts. Finally, this is a one-time forecast—that is, the evaluation is concerned with the BLS projections for a *single* year, 2000—while most forecast evaluations have examined multiple forecasts.

This article evaluates the labor force, employment by industry, and occupation projections that BLS made in 1989 for the year 2000.[2] While these forecasts have already been evaluated individu-

ally,[3] it is possible to both ask additional questions that were not addressed in those studies and to use evaluation methodologies different from those employed previously. In addition, this article, whenever possible, uses the same methodologies to evaluate the projections of all three of these variables.

Methodological issues

Because there are no other forecasts that are comparable to the BLS projections, it is necessary to construct a benchmark for the projections of each variable. In each case, BLS projections are compared with similar data obtained from the forecasts of a benchmark. The benchmarks that were selected all use data that were available at the time when BLS projections were prepared. In actuality, the benchmarks are naïve models such as: (1) projecting the latest available information; or (2) predicting that the change over the forecast period is equal to that observed over the previous time interval, which is of the same length as the forecast period.[4]

Because the projections that are being analyzed in this article were prepared in 1988, the forecast period is 12 years in length. Consequently, the change from 1976 to 1988 was used as the basis for this benchmark.

At a minimum, the BLS projections should be more accurate than the forecasts of these naïve models.

Long-term projections vs. short-term forecasts. The questions that are appropriate for evaluating

H.O. Stekler
is a research professor
in—and Rupin Thomas,
a graduate of—the
Department
of Economics,
George Washington
University.
E-mail:
hstekler@gwu.edu

the short-term forecasts have been examined in detail,[5] but the questions that should be asked in analyzing longer run projections have not been given the same degree of attention. Because BLS projections primarily focus on long-run trends, the questions asked and the statistics used in evaluating these forecasts should be related to the primary emphasis of the forecast. Thus, the two *basic* questions to be asked in evaluating these projections are: (1) Have the trends, especially structural changes, been predicted correctly? (2) Were these forecasts better than those that could have been produced by a benchmark method? Additional questions such as what the sources of the errors were and if the forecasts improved over time can also be posed.

The statistics that can answer these questions include the following: (1) the percentage of components where the direction of change was predicted correctly; (2) dissimilarity indexes that measure the structure of the labor force, and so forth; (3) contingency tables that determine whether the actual and predicted directions of change are related; and (4) Spearman Rank Correlation Coefficients that measure the relationship between the predicted and actual changes of the components of an aggregate forecast. Whenever possible, the same statistical procedures are used to measure the accuracy of the forecasts of all three variables for which BLS made projections.

One-time projections. In most forecast evaluations, the analyst examines a set (time series) of forecasts. It is then possible to discuss the characteristics of the *average* forecast. We cannot do this here, because BLS projections do not constitute a time series of forecasts. Rather, the projections (of the labor force, employment by industry, and of occupational employment) that BLS made in 1989 for the year 2000 are examples of *predictions made for a single end point*. Consequently, there are two reasons why the procedures that have been employed to evaluate sets of forecasts cannot be utilized in this case.[6]

First, the *magnitude* of the forecast error involving predictions made for a single end point may be a function of events that were unique to that particular year. This would be especially true if the target year is a recessionary year and not one in which full employment prevails. Thus, one should not base the forecast evaluation of *this one prediction* on how close the projection was to the outcome.[7] Instead, it is necessary to develop measures and to use benchmarks or standards of comparison that are independent of the magnitude of unique events.

Second, evaluations of a set of forecasts consider the characteristics of the *average* forecast. By focusing on the average forecast, the random shocks that affect particular years are canceled out. In such evaluations, it has been customary to use

quantitative measures such as mean absolute error (MAE), mean absolute percentage error (MAPE), or mean square error (MSE) to describe the characteristics of the forecasts.[8]

Because we are evaluating a single forecast, we use measures that are appropriate and that answer the two basic questions listed above. However, the questions asked and the descriptive statistics used in past forecast evaluations were also examined. The analysis is done separately for the forecasts of the three different variables.

Labor force projections

The BLS projection of the labor force for 2000 was based on two estimates obtained from different sources. The Census Bureau provided the population estimates for 2000 for 14 classifications of age and gender. BLS then multiplied these population numbers by its own participation rate estimates for each of these 14 classifications. It then summed the 14 estimates to obtain the overall estimate of the labor force size in 2000. This projection has been evaluated by Howard N Fullerton, Jr.[9]

Exhibit 1 presents some of the questions that were asked in both the original forecast and subsequent evaluation. These include: What is the projected size of the labor force, by age and gender? What is the growth rate of the labor force? What are the participation rates of the various groups? What is the distribution of the total labor force by age and gender? The error measures that were used in evaluating these projections are also presented in exhibit 1. They include the direction of error, the absolute and percentage error, the dissimilarity index, and so forth. The limitations of these questions and statistics are also noted.

Both the questions and measures used to evaluate the projections are relevant and appropriate. The major shortcoming is that there are no benchmark standards with which the forecasts can be compared. In addition, there are several other questions that can be posed in these evaluations.

Was the labor force projection accurate relative to the benchmark? Table 1 indicates that 1989's projection of the 2000 labor force overestimated the actual data by 0.2 million persons. Fullerton, however, also indicates that this small error was the result of offsetting errors made by both the Census Bureau (in underestimating the population) and by BLS (in overestimating the participation rates). In order to evaluate this projection, we calculated three alternative estimates of the 2000 labor force. The first uses the actual 2000 population in combination with the predictions of the participation rates made in 1989. This estimate can be used to measure the magnitude of the error that is entirely attributable to the misestimates of the participation rates. This projection is 3.8 million too high. (See table 1).

The second alternative is based on the actual 2000 population and the 1988 participation rates. This measure can be used as a standard with which the participation rates forecasted in 1989 can be compared. This projection is 1.9 million too low. Thus, if the actual population had been known in 1989, the naïve procedure of using the 1988 participation rates would have yielded a more accurate forecast than one using participation-rate estimates projected for 2000.

In 1989, however, BLS would not have known the actual 2000 population. Consequently, as a benchmark or standard of comparison, a projection is presented based entirely on data available at the end of 1988—that is, Census Bureau population projections available in 1989 and the 1988 participation rates. This estimate of the labor force is 134.8 million, yielding an error of 6.1 million. Comparing the BLS projection made in 1989 with this estimate clearly shows that the BLS estimate of the 2000 labor force published in 1989 was more accurate than the standard of comparison. (See table 2.)

The same analysis was applied to the projections of the male and female components of the labor force. We again conclude that the BLS estimates were more accurate than the standard of comparison. Because these results hold for all

Exhibit 1. Questions about the labor force forecasts

Questions	Accuracy measure	Problem with questions and/or accuracy measure	New question and/or measure
What is the size of the total labor force?	Absolute error, percentage direction of error	Does not distinguish between census population errors and participation rate errors, standard of comparison	How much of total labor force error is the result of participation rate errors? Standard of comparison: 1988 participation rates
What is the size of the labor force by gender and so forth?	Mean absolute percentage error, direction of error	Same as total labor force	Same as total labor force
What is the growth rate of the total labor force?	Error in percentage points	Same as total labor force	How much of the error in the growth rate forecast is the result of participation rate errors? Standard of comparison: 1988 participation rates
What are the participation rates of total labor force? Of men? Of women? By age and sex?	Error in percentage points, or absolute error/participation rate; mean absolute percentage error.	Does not indicate whether direction of change in participation rate was predicted, no standard of comparison.	Were the directions of change in the participation rates accurately predicted? Standard of comparison: number of changes accurately predicted versus predictions by chance (binomial, p=0.5)
What was the distribution of the labor force by age and sex?	Dissimilarity Index	No standard of comparison	Dissimilarity Index: comparison with naïve model

Table 1. Alternative estimates of the 2000 labor force and its rate of growth

Labor force	Actual	1989 BLS estimate	Actual population		Census estimated population, 1988 participation rate (standard of comparison)
			BLS estimated participation rate	1988 participation rates	
Total labor force (millions)	140.9	141.1	144.7	138.2	134.8
Rate of growth	1.2	1.2	1.5	1.1	.9
Male labor force (millions)	75.2	74.3	76.5	76.8	74.6
Rate of growth	1.0	.9	1.1	1.1	.9
Female labor force (millions)	65.6	66.8	68.2	61.7	60.4
Rate of growth	1.5	1.7	1.8	1.0	.8

Table 2. Participation rates, actual 1988, BLS forecast for 2000, actual 2000 and forecast errors

Group	Participation rates			Forecast error	
	Actual 1988	BLS 2000 forecast	Actual 2000	BLS	Naïve (1988)
Total, 16 and older	65.9	69.0	67.2	−1.8	1.3
Men					
16 years and older	76.2	75.9	74.7	−1.2	−1.5
16 to 19	56.9	59.0	53.0	−6.0	−3.9
20 to 24	85.0	86.5	82.6	−3.9	−2.4
25 to 34	94.3	94.1	93.4	−.7	−.9
35 to 44	94.5	94.3	92.6	−1.7	−1.9
45 to 54	90.9	90.5	88.6	−1.9	−2.3
55 to 64	67.0	68.1	67.3	−.8	.3
65 and older	16.5	14.7	17.5	2.8	1.0
Women					
16 and older	56.6	62.6	60.2	−2.4	3.6
16 to 19	53.6	59.6	51.3	−8.3	−2.3
20 to 24	72.7	77.9	73.3	−4.6	.6
25 to 34	72.7	82.4	76.3	−6.1	3.6
35 to 44	75.2	84.9	77.3	−7.6	2.1
45 to 54	69.0	76.5	76.8	.3	7.8
55 to 64	43.5	49.0	51.8	2.8	8.3
65 and older	7.9	7.6	9.4	1.8	1.5

labor force size estimates, a *fortiori*, the same conclusions apply to the estimates of the various growth rates.[10]

The accuracy of participation rates. Although the BLS projections of the 2000 labor force were more accurate than those of the standard of comparison, they benefitted from offsetting errors. The Census Bureau population estimates were too low, while the participation rates were overestimated. In evaluating these estimates of the participation rates, the following two questions are posed: (1) Did the estimates correctly predict the direction of change between 1988 and 2000? (2) Were these projections of the level of the participation rates more accurate than those generated by a standard of comparison?

There are 14 classifications of the labor force based on age and gender. (See table 2). The direction of change between 1988 and 2000 was projected correctly for 9 of these 14 classifications.[11] Using the binomial distribution with $p = 0.5$, it is possible to test the null hypothesis that this favorable result could have occurred purely by chance. We are unable to reject this hypothesis. As an additional test, we compared the levels of the participation rates observed in 2000 with the following: (1) the ones that BLS projected for 2000; and (2) the 1988 participation rates, which are used as the benchmark. The latter had smaller absolute errors in a majority of the cases. These results indicate that there was room for improving the projections of participation rates.

Measuring structural change: dissimilarity indexes. In order to determine whether the structural changes and major trends that occurred between 1988 and 2000 were predicted accurately, a statistic is used that directly addresses this question. The forecast of the total labor force is an aggregated estimate, and it is important to also examine the disaggregated component predictions. Such an analysis enables one to determine whether the structure of the aggregate has been predicted accurately.

If the aggregate, *X*, is predicted according to some scenario (for example, full employment), one would want to determine whether the structure is accurate even if the total is wrong. R.A. Kolb and H.O. Stekler developed a procedure for decomposing the total error into two components—where the first measures the scenario discrepancy and the second, the structural error.[12] They calculated the proportion of the aggregate predicted and actual totals that were associated with each of the *i* components. While their analysis was based on an information content statistic, using dissimilarity indexes would yield the same result.

A dissimilarity index is a statistic that can be used to determine whether one distribution approximates another one. Specifically, it measures the amount by which the forecasted distribution would have to change to be identical to the actual distribution. The formula for the dissimilarity index is:

$$D = 0.5 \, 3 \, | \, (P_{fi} / P_f) - (P_{ai} / P_a) \, |$$

where P_{fi} is the forecast of the labor force that will be in the

Table 3.	Dissimilarity indexes of labor force projections			
		Standards of comparison		
Age	BLS projections	Actual population and—		Census population estimate and—
		BLS participation rate	1988 participation rate	1988 participation rate
Gender, age	1.83	2.02	2.24	2.32
Men, age	1.63	.91	.62	1.37
Women, age	1.91	2.86	2.4	1.32

*i*th group, and P_f is the forecast for the total labor force. Similarly, P_{ai} and $\hat{P}_a$ are the corresponding actual data. *D* is bounded in the interval 0 to 100 percent. The smaller the value of *D*, the smaller the difference is between the predicted and actual distributions—that is, the more accurate the forecast.

The dissimilarity index for the BLS projections was based on the 14 age/gender categories that had been used in 1989 to prepare the estimates for 2000. Similar dissimilarity indexes were constructed for the other distributions that serve as standards of comparison. The values of the various dissimilarity indexes are presented in table 3.

The results are mixed. In some cases, the dissimilarity indexes obtained from the BLS projections are smaller (and thus more accurate) than those of the standards of comparison. In other cases, the opposite results were obtained. However, the dissimilarity index for the actual BLS forecast never exceeds 2 percent for all age/gender categories or for men and women separately. The values of the dissimilarity indexes of the standards of comparison were also around 2 percent, indicating that the BLS projection was comparable to but not superior to these (naïve) benchmark forecasts.

While there is no statistical distribution for the dissimilarity index, the BLS projection substantially predicted the structural changes that occurred in the labor force between 1988 and 2000. On the other hand, similar results were obtained from the naïve models that served as the benchmarks.

Did the forecasts improve? The primary focus of this analysis is on the projections that were published in 1989 for 2000. While there had been a second set of projections for industry employment by industry and occupation, BLS actually made five forecasts of the 2000 labor force. These were published in 1987, 1989, 1991, 1993, and 1995.[13] It is thus possible to determine whether the accuracy of the forecasts improved as the forecast horizon declined.

The results are mixed. (See table 4.) The forecasts of the labor force that were made in 1988 (and published in 1989) were more accurate than those made in any other year. Thus, they did not improve with the passage of time—that is, as the forecast horizon became smaller. On the other hand, as the forecast horizon declined, so did the errors in the forecasts of the participation rates.

Employment by industry

The questions asked about the employment-by-industry projections are presented in exhibit 2. These questions were discussed in both the original forecast and in the subsequent

Table 4.	Errors in labor force and participation rate projections for 2000, various horizons				
	Errors made in—				
Projection	1986	1988	1990	1992	1994
Labor force	−1.5	.2	1.5	.7	−.6
Participation rate:					
All	.6	1.8	1.5	1.0	−.2
Men	.0	1.2	1.3	.6	−.7
Women	1.3	2.4	1.8	1.4	.4

Exhibit 2. Questions about the employment forecasts

Questions	Accuracy measure	Problem with questions and/or accuracy measure	New question and/or measure
How many people will be employed in each industry?	Percentage error, mean absolute percent error	No standard of comparison; gives equal weight to large and small industries	Standard of comparison: rates of growth equal to previous rates of growth; mean weighted percent error
Which industries would have highest (lowest) employment growth rates?	Compare the number of industries projected to grow the fastest (slowest) with those that did grow fastest (slowest).	No standard of comparison; no analysis of all industries' projected and actual growth rates.	Standard of comparison: forecasts of fastest (slowest) growing industries from naïve model; Spearman rank correlation coefficient for all industries
What is the distribution of employment by industry?	Dissimilarity Index	No standard of comparison	Standard of comparison: same share as in 1988 and shares based on previous growth rates
What were the sources of the industry employment forecasts errors?	Model simulations	None	—

evaluation of that forecast. These include: What is employment by major industry group? What is employment by industry? Which industries are expected to grow the fastest? In which industries will employment decline? These are all questions that involve structural change and should be evaluated correspondingly. The measures that previously have been used in the evaluations and some of the limitations of these measures are also listed in exhibit 2.

Mean absolute percent error. How accurate were the employment-by-industry projections? The accuracy of the employment-by-industry projections has conventionally been evaluated by calculating the mean absolute percent error (MAPE). Again, an evaluation of a single prediction should not be based on the magnitude of the error regardless of whether it is measured in absolute or percentage terms. Nevertheless, we use this statistic in order to note that it can be calculated in two ways: (1) the simple average of the absolute percentage error of the forecast for each industry; or (2) a weighted average of the industries' errors, with the weights equal to the industry's share of employment.[14] The second measure reduces the weight of small industries, which might have large percentage errors.

The standard of comparison was a naïve forecast. It was assumed that the employment growth rate in each industry between 1988 and 2000 would be the same as the one that

had occurred between 1976 and 1988. The mean absolute percent errors (MAPEs) of the BLS projections and of the naïve model are presented in the following tabulation:

	BLS		Naïve	
	Unweighted	*Weighted*	*Unweighted*	*Weighted*
Major industry sectors	11.9	7.4	11.8	7.2
174 disaggregated industries ...	18.8	13.6	24.6	14.4

They are divided into several categories: unweighted and weighted MAPEs for the 12 major industry sectors and the 174 disaggregated industries. For the 12 major sectors, there is very little difference between the errors of the BLS projections and those of the naïve model used as the standard of comparison. The MAPEs of the BLS projections for the 174 disaggregated industries are less than those of the naïve standard of comparison. This is especially true for the unweighted MAPEs.

The BLS projections correctly predicted the direction of employment change in 135 of 174 industries. The naïve model made a larger number of mistakes, 49. Because most industries grew during this period, a better measure is to de-

termine whether industries in which employment was expected to grow rapidly (or slowly), actually experienced this type of growth. (This is done later where rank correlation coefficients and contingency tables are used.)

Measuring structural change: dissimilarity indexes. As with the labor force projections, we use dissimilarity indexes to determine whether the structural employment changes that occurred were forecast accurately. These dissimilarity indexes were for the 12 major industry sectors as well as for the 174 different smaller industries—and for the benchmarks used as the standards of comparison. Two naïve models serve as benchmarks. The first assumes that each industry's share of total employment would be the same in 2000 as it had been in 1988. The second assumes that each industry's share (s_i) of total employment would increase from 1988 to 2000 by the same amount as occurred between 1976 and 1988—that is,

$$s_{i2000} = s_{i1988} + (s_{i1988} - s_{i1976}).$$

The results are mixed. The following tabulation shows the dissimilarity indexes for BLS projections and naïve model estimates:

	BLS projections	Dissimilarity indexes for naïve models	
		No change from 1988	Same change 1988–2000 as 1976–88
12 major industries	3.75	6.03	3.60
174 industries	6.89	10.38	8.09

The dissimilarity indexes associated with the BLS projections are smaller than those derived from the naïve standard that assumed the shares of industrial employment would remain constant between 1988 and 2000. In comparison with the naïve growth model, the BLS projections were better for the disaggregated industries, but slightly worse for the 12 major groupings. These results suggest that the BLS projections were able to capture the structural changes that occurred in industry employment as least as well as would have been obtained from a simple extrapolation.

Structural change: Spearman rank correlations. Another way to determine whether the projections captured the structural changes that occurred is to compare the forecasted growth rates of industrial employment with the actual growth rates. The original projections had listed the 20 industries expected to have the largest employment growth rates and the 20 industries that were expected to show the largest employment declines.[15] Only 12 of the industries that were expected

to experience the fastest growth actually did so. It is difficult to interpret these results without a standard of comparison. The naive extrapolation model is again used as the benchmark. That model actually identified 13 of the 20 industries that had the highest employment growth rates. (See table 5). Similar results were obtained for the 20 slowest growing industries. BLS and the naïve model identified 8 and 7 of those 20 industries, respectively. Consequently, we conclude that the BLS projections of the fastest and slowest growing industries were not substantially different from the forecasts generated by a naïve model.

Rather than merely focus on the 20 industries in the two tails of the distributions, the Spearman rank correlation coefficient between the predicted and actual growth rates for all 174 industries were also calculated. This coefficient was 0.64 for both the BLS projections and the naïve extrapolation model. This result indicates that both sets of forecasts were able to forecast many of the structural changes that occurred, but there was no difference between the BLS projections and the extrapolations obtained from a naïve model.

Structural change: contingency table. The Spearman rank correlation coefficient provides an overall assessment of the rankings of the predicted and actual employment growth rates. Another method for demonstrating the same result is to construct a contingency table. After the employment growth rates of the 174 industries had been ranked, they were divided into quintiles, with the 35 industries having the highest growth rates placed in the first quintile, and so forth. This procedure was applied to both the projected and actual growth rates, and a 5X5 contingency table was constructed. We then tested the null hypothesis that there was no relationship between the predicted and actual growth rates of these quintiles. A similar procedure was used to evaluate the forecasts of the naïve extrapolation model. The contingency tables are presented in table 6. While the null of independence is clearly rejected, less than half of the observations of both the BLS and naïve model projections lie on the main diagonal (indicating that the forecasted and actual growth rates were in the same quintile). However, a majority of the remaining observations lie in the adjacent cells. These results are consistent with those obtained from the Spearman rank correlation coefficients.

The main result is that the BLS projections do not differ significantly from those obtained from the naïve extrapolative model.[16]

What were the sources of error in the employment projections? Arthur Andresassen used computer simulations and factor analysis to determine why the employment errors oc-

curred.[17] He showed that there were two basic errors that offset each other. The low projection of gross domestic product was offset by inaccurate employment-output relationships. Our analysis did not attempt to replicate this analysis.

Occupational employment

The questions discussed in the occupation projections and in the subsequent evaluation are analogous to those of the

Table 5.	Rank of the 20 industries exhibiting the highest employment growth rates, 1988–2000 by rank of BLS and naïve model projections			
		Rank		
Industry		Actual	BLS Projection	Naïve model projection
Computer and data processing services		1	9	2
Personnel supply services		2	10	1
Health services		3	3	5
Amusement and recreation services		4	39	44
Miscellaneous transport services		5	16	28
Residential care		6	11	3
Individual and miscellaneous social services		7	14	13
Research and testing		8	38	12
Water and sanitation		9	6	22
Security and commodity brokers and exchanges		10	19	6
Commercial sports		11	94	69
Credit agencies and investment offices		12	1	15
Motion pictures and video tape rentals		13	92	42
Miscellaneous business services		14	22	17
Job training and related services		15	43	10
Child daycare services		16	131	29
Oil and gas field services		17	4	95
Personal services		18	5	9
Miscellaneous equipment rental and leasing		19	2	4
Air transportation		20	104	8

Table 6.	Relationship between ranks of predicted and actual growth rates of employment by industry quintiles, BLS projections and Naïve model				
	Actual growth				
Projected growth	1–35	36–70	71–105	106–140	141–174
BLS projection[1]					
1–35	18	13	2	1	1
36–70	9	12	10	2	2
71–105	5	5	11	7	7
106–140	1	4	10	12	8
141–174	2	1	2	13	16
Naïve model[2]					
1–35	21	11	3	0	0
36–70	6	12	6	4	7
71–105	6	7	11	7	4
106–140	1	4	11	10	9
141–174	1	1	4	14	14

[1] $x^2 = 93.61$
  $P = 0$
[2] $x^2 = 92.22$
  $P = 0$

Exhibit 3. Questions about occupational forecasts			
Questions	**Accuracy measure**	**Problem with questions and/or accuracy measure**	**New question and/or measure**
How many people will be employed in each occupation?	Absolute error, absolute percent error	No standard of comparison; gives equal weight to large and small occupations	Standard of Comparison: Naïve model: same growth; mean weighted percent error
Which occupations will grow fastest?	Compare the number of occupations projected to grow the fastest with those that did grow fastest; distribution of growth rates by growth adjectives	No standard of comparison; analysis of all occupations' projected and actual growth rates	Spearman rank correlation coefficient; standard of comparison not possible due to definitional changes
Which occupations will have the largest job growth?	Compare the number of occupations that were projected to have largest job growth with those that did	No standard of comparison	Standard of comparison not possible due to definitional changes
What is the distribution of employment by occupation?	Absolute percent error	No standard of comparison	Dissimilarity Index: comparison with naïve model
What were the sources of errors?	Model simulations	None	—

employment by industry estimates. (See exhibit 3.) In presenting the projections, the analysis included the occupation's share of employment; the occupations that are likely to grow the fastest or decline; and the occupations that are likely to have the largest number of new jobs. The evaluation by Andrew Alpert and Jill Auyer considered the absolute percent change of actual and projected occupational employment; the numerical change in these categories; and the share of employment of each occupational group. Again, these are appropriate measures, but it is possible to use additional measures and compare them with a standard of comparison. The benchmark is the naïve model in which it is assumed that the growth in each occupation between 1988 and 2000 was equal to the growth rate that occurred between 1976 and 1988.[18]

The following tabulation shows MAPEs for BLS projections and naïve model estimates:

	BLS		Naïve	
	Unweighted	*Weighted*	*Unweighted*	*Weighted*
Major occupations....	5.86	5.29	13.8	11.8
338 disaggregated occupations....	45.2	15.0	–	–

The mean absolute percent errors of the projections of the major occupational groups are substantially smaller than those of the naïve model. Moreover, the BLS projections correctly predicted the direction of change for eight of the major occupational groupings, with agriculture being the exception.[19] On the other hand, the unweighted MAPE for the 338 smaller occupational groups is substantial (45.2 percent), indicating that there were substantial errors in many of these occupational groups. Because the weighted MAPE is smaller (15.0 percent), the larger percentage errors occurred in the smaller occupational groups.

Structural change: dissimilarity indexes. As with the projections of the other two variables, dissimilarity indexes are used to determine whether the structural employment changes that occurred were forecast accurately. These dissimilarity indexes were for the 9 major occupational groups, as well as for the 338 different occupational classifications and for the benchmarks used as the standards of comparison. For the major occupational groups, two naive models are used as benchmarks. The first assumes that each occupation's share of total employment would be the same in 2000 as it had been in 1988. The second assumes that each occupation's share (s_i) of total employment would increase from 1988 to 2000 by the same amount as occurred between 1976 and 1988—that is, $s_{i2000} = s_{i1988} + (s_{i1988} - s_{i1976})$. There is only one naive benchmark for the 338 occupa-

Table 7. Relationship between ranks of predicted and actual growth rates of occupational employment, quintiles, BLS projections and actual growth

Projected growth	Actual growth				
	1–67	68–135	136–203	204–271	272–338
1–67	32	14	8	7	6
68–135	12	17	16	15	8
136–203	12	17	15	13	11
204–271	8	13	19	16	12
242–338	3	7	10	17	30

tions because the definition of some of the occupations changed, and, thus, it was not possible to construct the comparable growth rates. Consequently, for these 338 occupations, we only used the first of these naïve models—that is, the distribution of occupational shares would be the same in 2000 as it had been in 1988.

The results in the following tabulation indicate that the BLS projections captured the structural changes that occurred in occupational employment better than the naïve models did. The dissimilarity indexes associated with those projections were substantially smaller than those of the naïve models. Shown below are the dissimilarity indexes for BLS projections and the naïve models.

		Dissimilarity indexes for naïve models	
	BLS projections	*No change (2000=1988)*	*Same change 1988– 2000 as 1976–88*
Major occupations	2.12	3.12	4.53
338 disaggregated occupations	7.64	8.43	–

Structural change: Spearman rank correlation. Similar to the employment-by-industry projections, the BLS estimates listed the 20 occupations that were expected to grow the fastest. Of these 20 occupations, only 6 actually had the fastest growth. Instead of focusing on the occupations that were in the tail of the distribution, we calculated the Spearman rank correlation coefficient between the predicted and actual growth rates for all 338 occupations. That coefficient is 0.43; it is statistically significant, but because there is no comparable benchmark, there is no basis of comparison. We can only note that this Spearman coefficient is substantially less than the comparable coefficient for the employment by industry data.

Structural change: contingency table. In presenting the distributions of the actual and projected growth rates of the 338

occupations, Alpert and Auyer divided them into six growth categories, ranked from declining to growing much faster than average. While they did not test the null hypothesis that there is no relation between the projected and actual growth rates, this hypothesis can be rejected. We also constructed a contingency table (see table 7), but it is based on the quintiles of each distribution rather than on growth categories. The hypothesis that there is no relationship between the projected and actual growth rates is also rejected, but it should be noted that there are many observations that are not on the main diagonal or in the adjacent cells. This result indicates that the projections for many occupations were clearly inaccurate and explains why the Spearman rank correlation coefficient is only 0.43.

What were the sources of error in the occupation projections? Alpert and Auyer identified some of the sources of error in the occupation projections. Some of the errors were attributable to assumptions made about technological changes that were expected to occur between 1988 and 2000. These included increases in automation that did not occur in many occupations, thus accounting for larger increases in employees than was anticipated. In addition, these authors ran simulations to show that in some cases inaccurate staffing patterns were the source of the errors, while in other cases the misestimates could be attributed to the mistakes made in the industry projections. We did not replicate this analysis.

Overall conclusions

This study established a set of procedures for evaluating BLS projections of the labor force, industry employment, and occupational employment. These procedures were then used to evaluate the projections for 2000 that were published in 1989. The projections were compared with benchmarks derived from naïve models. Our results showed that in most cases, the accuracy of the BLS projections were comparable to estimates obtained from naïve extrapolative models. □

Notes

NOTE: This article was written under contract with the Bureau of Labor Statistics to explore current projection evaluation techniques and to suggest new approaches for effective evaluation of the Bureau's long-term projections of employment. BLS published its own evaluation of the 2000 projections relative to actual outcomes in the October 2003 *Review*. The current paper suggests additional evaluation approaches, including a comparison with what a simple extrapolation would have produced and the use of contingency tables. BLS intends to employ the new techniques suggested here in addition to all of its more traditional evaluation techniques in its examination of future employment projections.

[1] A governing principle of such an evaluation is that the forecaster should not be penalized for external errors nor benefit if the external errors offset the internal mistakes.

[2] See Howard N Fullerton, Jr., "New labor force projections, spanning 1988 to 2000," *Monthly Labor Review*, November 1989, pp. 3–11; Valerie Personick, "Industry output and employment: a slower trend for the nineties," *Monthly Labor Review*, November 1989, pp. 25–41; and George Silvestri and John Lukasiewicz, "Projections of occupational employment," *Monthly Labor Review*, November 1989, pp. 42–65.

[3] See Howard N Fullerton, Jr., "Another look at the labor force," *Monthly Labor Review*, November 1993, pp. 31–40; Andrew Alpert and Jill Auyer, "Evaluating the BLS 1988–2000 employment projections," *Monthly Labor Review*, October 2003, pp. 3–12; and Arthur Andreassen, "An evaluation of the 2000 employment by industry projections," Bureau of Labor Statistics mimeo.

[4] Because the projections that are being analyzed in this article were prepared in 1988, the forecast period is 12 years in length. Consequently, the change from 1976 to 1988 was used as the basis for this benchmark.

[5] H.O. Stekler, "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, Vol. 7, 1991, pp. 375–84.

[6] The problem is not due to the length of the forecast horizon because long-run population forecasts have been evaluated using these error measures. Rather, this problem occurs when there are a small number of observations. See Stanley K. Smith and Terry Sinich, "Evaluating the Forecast Accuracy and Bias of Alternative Population Projections for States," *International Journal of Forecasting*, Vol. 8, 1992, pp. 495–508.

[7] However, if a *set* of such *single* forecasts is available, the quantitative measures could then be used to evaluate the set of these predictions. It is thus imperative to also calculate the quantitative statistics so that eventually an entire set of forecasts can be evaluated.

[8] H.O. Stekler (1991) and R. Fildes and Stekler (2002) noted that these error measures are descriptive statistics and do not inform whether the forecasts are "good" or "bad". To make such a judgment, a benchmark is required. It is then possible to determine whether the forecasts being evaluated are more accurate than those generated by the benchmark procedure. In addition, it is desirable to be able to use a statistical test to determine whether the two sets of forecasts are *significantly* different. See H.O. Stekler, "Macroeconomic Forecast Evaluation…"; and R. Fildes and H.O. Stekler, "The State of Macroeconomic Forecasting," *Journal of Macroeconomic Forecasting*, Vol. 24, 2002, pp. 435–468.

[9] Howard N Fullerton, Jr., "Evaluating the BLS labor force projections to 2000," *Monthly Labor Review*, October 2003, pp. 3–12.

[10] The BLS forecast is also more accurate than a naïve model that would have predicted the growth rate from 1988–2000 would be identical to that observed from 1976–1988. That naïve model would have predicted the labor force to grow at an annual rate of 2 percent. See Howard N Fullerton, Jr., "New Labor Force Projections, Spanning…"

[11] The errors occurred in the youngest and oldest age groups of both men and women.

[12] R.A. Kolb and H.O. Stekler, "Information Content of Long-Term Employment Forecasts," *Applied Economics*, 1992, pp. 593–96.

[13] See Howard N Fullerton, Jr., "Labor force projections: 1986 to 2000," *Monthly Labor Review*, September 1987, pp. 19–29; "New labor force projections, Spanning…"; "Labor force projections: the baby boom moves on," *Monthly Labor Review*, November 1991, pp. 31–44; "Another look at the labor force…"; and "The 2005 labor force: growing, but slowly," *Monthly Labor Review*, November 1995, pp. 29–44.

[14] Even though the MAPE is not useful in evaluating a single forecast, this statistic can be used when single forecasts are combined into a set of forecasts.

[15] Valerie Personick, "Industry output and employment…"

[16] This result suggests that extrapolative methods may have been the basis of the BLS projections.

[17] Arthur Andreassen, "An Evaluation of the 2000 Employment…"

[18] This benchmark could only be used for the major occupational groupings because the definitions of the disaggregated groupings have changed.

[19] The naïve model generated identical results.