12-1987

What Is the Effect of Random Variation in State Unemployment Rates?

Edward W. Hill
Cleveland State University, e.hill@csuohio.edu

What is the effect of random variation in State unemployment rates?

State and local users of the data may tend to assume that the rates have low levels of dispersion; however, a closer analysis reveals large variances attributed to sample size

Edward W. Hill

The reported monthly unemployment rate from the Current Population Survey (CPS) is the best point estimate of labor market activity available by State and local labor market areas. Because of its timeliness, wide coverage, and comprehensiveness, it is used by governments, planners, corporations, and the media. However, statements are often made about fluctuations in the unemployment rate which are unwarranted due to the variance of the data series.

The inverse of the unemployment rate is commonly used as a proxy for gross regional product. It is also used intraregionally, as a coincident indicator of the local business cycle. Interregionally, it is used as a sign of the relative strength of local economies. The unemployment rate is also an important instrument in public policy decisions. This is especially true at the State and local levels where announcements in the rate can trigger political activity. The annual rate is used by the Federal Government to redistribute funds to the States. In many States, the rate is used as part of formulae to redistribute funds from State to local governments. It is also used to extend or contract the length of time people are eligible for unemployment benefits.

Most of these uses of the unemployment rate for States and localities assume that it has low levels of dispersion and that month-to-month movements in the rate are meaningful. Because users usually do not pay attention to error attributed to random variation in sampling, they may be using the unemployment rate to make inferences, decisions, resource allocations, or policy statements which are unwarranted.

The first section of this article examines national CPS data to indicate the impact which sample size has on the standard error of subpopulations in the sample and to show how these errors can influence policy conclusions.

The second section examines the unemployment rate cross-sectionally for the 11 States for which data are available from the April 1986 CPS.[1] These data demonstrate that the monthly unemployment rate should not be used to make finely drawn distinctions between the States. This is especially true if the data are used to make inferences about the relative aggregate economic well-being of the States.

The third section uses monthly time series data, from January 1982 to December 1986, for the State of Ohio. These data are employed to examine the extent to which movements in the reported monthly unemployment rate are statistically significant.

Statistical error in the CPS

Reported differences in the variance for specific national CPS subpopulations are largely caused by relative subsample sizes. For instance, the expected coefficient of variation for the civilian labor force and the number employed will be lower than the coefficient of variation for the number unemployed and, correspondingly, for the unemployment rate. Relative errors for demographically distinct subpopulations also vary with size. It is shown in table 1 that as the size of

Edward W. Hill is Research Director of the Economic Development Program, College of Urban Affairs, Cleveland State University.

the population decreases, the coefficient of variation and the resulting confidence interval increases.

The CPS unemployment rate was 7 percent in April of 1986; with a coefficient of variation of 1.7 percent, the 95-percent confidence interval ranged from 6.76 percent to 7.24 percent. (The normal confidence level used for these by BLS is 90 percent.) It is interesting to note that the levels of dispersion for subpopulations, with which social policy has been historically concerned, are of much greater magnitude than those found for the sample as a whole. The reported unemployment rate for black men was 13.4 percent, and the 95-percent confidence interval was from 12.04 percent to 14.69 percent. The rate for black teens was 40.7 percent, and the 95-percent confidence interval ranged from 36.40 percent to just under 44.68 percent. These are wide error bands and are cause for concern if the rates are being used for reasons other than business cycle analysis.[2] Seemingly large changes in the unemployment rate for these groups would actually not be significant. They could be a fluke of the specific month's sample.

It is instructive to calculate what the unemployment rate would have to be in May to be significantly different, at the 95-percent confidence interval, from the April figures. This can be done by using the standard error of month-to-month variation in the unemployment rate. The overall unemployment rate must either exceed 7.24 percent, or drop below 6.76 percent. The rate for nonteenage white men would need to fall outside of the 5.05-percent to 5.67-percent range, and the rate for nonteen black men would be outside of the 12.04-percent to 14.69-percent range. The range for black teens is from 36.40 percent to 44.69 percent.[3] In each case, the May rate fell inside of the confidence interval, which implies that we cannot say with statistical certainty that the May rates are different from those of April.

National data demonstrate how relatively small sample sizes can influence the utility of the unemployment rate as a social indicator. This is a task for which the metric is frequently used. Dispersion caused by small sample sizes makes movements in the monthly unemployment rate for minority subpopulations nearly meaningless.

Cross-sectional variations

Few attempts are made to gauge the precision of the States' monthly unemployment estimates. However, the CPS is designed to ensure that reported unemployment levels have a coefficient of variation of 8 percent or less, at a 6-percent unemployment rate.[4] This standard applies to monthly unemployment rates which are reported for the 11 States with populations large enough to yield an adequate sample (these will be referred to as "survey States" in this article). It also applies to the annual unemployment estimates for all of the States and the District of Columbia. The remaining 39 States and the District of Columbia use a nonsurvey method to estimate their monthly and quarterly rates.

There are large differences in the estimated unemployment rates among the States. However, finely drawn distinctions among them may be misleading. This is especially apparent when the data are viewed within the context of the "common wisdom." This wisdom holds that States on the coasts have fared well in the current recovery, but the midsection of the country is faring less well. This wisdom can be questioned when variations in State estimates are considered.

Table 2 lists the estimated unemployment rates, the coefficients of variation, and the 95-percent confidence intervals for the survey States. There are substantial differences in the levels of variation. The coefficient of variation is higher for States with smaller populations and lower unemployment rates; the average coefficient of variation is 7.11 percent. The table contains two measures of relative dispersion, the coefficient of variation and the range of the confidence interval as a percentage of the estimate. The latter measure divides the difference between the extremes of the 95-percent confidence interval by the reported unemployment rate. It is a measure of the relative width of the interval. The average of this measure is 27.4 percent, indicating that the interval is extremely wide.

A t-test of the difference in the unemployment rates between any two of the survey States was conducted to examine whether the differences were statistically significant. As table 3 indicates, in several cases they were not.

The States can be placed into four groups. Massachusetts' reported unemployment rate is significantly different from New Jersey's and it constitutes the first group. The second group consists of New Jersey, North Carolina, and Florida.

Table 1. Month-to-month variation in the unemployment rates for subpopulations in the Current Population Survey, April 1986

Characteristic	Estimated unemployment rates	Coefficient of variation[1] (percent)	95-percent critical values[2]	
			Minimum	Maximum
Total, 16 years and older	7.0	1.70	6.76	7.24
White:				
Men, 20 years and older	5.4	2.59	5.05	5.67
Women, 20 years and older	5.2	2.50	4.86	5.45
Both sexes, 16–19	15.7	3.69	14.44	16.83
Black:				
Total, 16 years and older	14.6	4.04	13.42	15.76
Men, 20 years and older	13.4	4.92	12.04	14.69
Women, 20 years and older	12.0	4.50	10.82	13.06
Both sexes, 16–19	40.7	4.98	36.40	44.68

[1] The coefficient of variation is calculated by dividing the standard error (s) by the mean (x) and multiplying the result by 100, ((s/x)•100).

[2] The 95-percent confidence interval of the unemployment rate is calculated by multiplying the standard error by 1.96 and adding or subtracting, that number from the reported unemployment rate (which is the estimate of this distribution), x±(1.96•s).

SOURCE: The standard errors were obtained and calculated from *Employment and Earnings*, May 1986, tables A-6, C, and G. All calculations were made by the author.

The estimated mean unemployment rate of each State is not statistically different from the other States in this group. California, New York, and Pennsylvania constitute the third group. When one puts aside glorified stories of the economic renaissance on the west coast, it appears that there is no significant difference between California and Pennsylvania in terms of their mean levels of unemployment. The hypothesis that Pennsylvania's rate of 7 percent is not different from Ohio's 7.9 percent cannot be rejected. But it appears that Ohio is more closely associated with the high unemployment group: Ohio, Illinois, Texas, and Michigan.

It is unwise to use monthly unemployment rates unaccompanied by other data to make finely drawn distinctions among the States. Cross-sectional data indicate that statistical uncertainty, which is inherent in monthly State unemployment rates, results in confidence intervals that are nearly 28 percent as large as the estimated unemployment rate.

Ohio's time-series variation

Monthly data for Ohio are examined to determine the frequency of significant differences in the reported unemployment rates. Seasonally adjusted time-series data from January 1982 to December 1986 are used to examine whether month-to-month changes in Ohio's unemployment rates are significant.

The 59 months of data plotted in chart 1 constitute a particularly good period to examine movements in Ohio's monthly unemployment rate because of the wide range—a high of 14.2 percent in January 1983, to a low of 7.4 percent in March 1986. This was an especially difficult time for Ohio. The people of the State experienced the usual cyclical swings of an economy dependent on capital goods production. In addition, they had to contend with accelerated secular change partially due to offshore competition.

Table 2. Reported unemployment rates for 11 CPS survey States, by levels of variation, April 1986

Area	Estimated unemployment rate[1]	Coefficient of variation[2]	Range of 95-percent confidence[3]	95-percent confidence interval[4]	
				Minimum	Maximum
Massachusetts	3.8	9.62	36.8	3.1	4.5
New Jersey	4.7	8.55	34.0	3.9	5.5
North Carolina	5.1	8.38	31.4	4.3	5.9
Florida	5.4	7.86	29.6	4.6	6.2
California	6.7	5.46	20.9	6.0	7.4
New York	6.7	5.58	20.9	6.0	7.4
Pennsylvania	7.0	7.04	28.6	6.0	8.0
Ohio	7.9	6.63	25.3	6.9	8.9
Illinois	8.2	6.56	25.6	7.1	9.2
Texas	8.2	6.27	24.4	7.2	9.2
Michigan	9.1	6.27	24.2	8.0	10.2
Average, 11 States	—	7.11	27.4	—	—

[1] Not seasonally adjusted.

[2] Coefficient of variation is calculated by dividing the standard error (s) by the mean (x) and multiplying the result by 100. ((s/x)*100).

[3] Range of percent of employment estimate: [(95-percent confidence interval maximum − 95 percent confidence interval minimum)/(Unemployment rate)]•100.

[4] The 95-percent confidence interval of the unemployment rate is calculated by multiplying the standard error by 1.96 and adding or subtracting that number from the reported unemployment rate (which is the estimate of this distribution), x ± (1.96•s).

SOURCE: Unemployment rate: *Employment and Earnings*, table D-1 (Bureau of Labor Statistics, 1986). Data and formulae to calculate standard errors: Charles D. Jones, "CPS Variances—Parameters Needed to Calculate State, Census Region, and Division Variances." All calculations made by author.

Table 3. Differences between estimated unemployment rates of the 11 survey States

State	Group 1	Group 2			Group 3			Group 4			
	MA	NJ	NC	FL	CA	NY	PA	OH	IL	TX	MI
MA	------	[1]−1.65									
NJ	------	-----	−0.68	−1.21	[1]−3.67						
NC	------	-----	----	−0.50	[1]−2.82						
FL	------	-----	----	----	[1]−2.32						
CA	------	-----	----	----	-----	0.00	−0.49	[1]−1.88			
NY	------	-----	----	----	-----	--	−0.49	[1]−1.88			
PA	------	-----	----	----	-----	--	----	−1.26	[1]−1.65		
OH	------	-----	----	----	-----	--	----	-----	−0.40	−0.41	−1.55
IL	------	-----	----	----	-----	--	----	-----	-----	0.00	−1.15
TX	------	-----	----	----	-----	--	----	-----	-----	-----	−1.18
MI	------	-----	----	----	-----	--	----	-----	-----	-----	----

[1] The reported unemployment rate for the State listed in the row is significantly different from that of the State listed in the column, using a one-tailed t-test at the 95-percent critical value.

NOTE: Reported numbers are the value of the t-test on the difference between two means:

$$[u_i - u_j]/[(s_i^2 + s_j^2)^{.5}]$$

where: u_i is the reported unemployment rate for State i,
u_j is the reported unemployment rate for the State with the next highest rate (State j),
s_i is the standard deviation of the rate of State i,
s_j is the standard deviation of the rate of State j.

SOURCE: See table 2.

To get a feeling for the amount of variance in the series, measures of dispersion and central tendency were developed.[5] Normally, economists and planners use the monthly unemployment rate as if each observation has no variation. But as the series is constructed with monthly samples, each observation has its own measure of dispersion.

The average monthly coefficient of variation of the unemployment rate over the time period was 5.9 percent. This metric, in turn, had a coefficient of variation of 9.6 percent, which indicates that there was a range of statistical error, or imprecision, in the data series. However, each month's reported unemployment rate is an efficient point estimator and the best unemployment data available for Ohio. It remains to be determined if the dispersion is sufficiently low to justify the robust way in which monthly changes in the unemployment rate are used.

The values of the t-ratios of the difference in each months' unemployment rate over time are plotted in chart 1. The t-test used is slightly biased in favor of finding that each month's rate is not different from the previous month's rate This is attributed to the fact that the correlation of the month-to-month variances used in the computation of the t-test is for the levels of unemployment, rather than the unemployment rates.[6]

The 66-percent and 95-percent critical values of the two-tailed t-test are displayed; they are ±1.00 and ±1.96, respectively. If the ratio has a value which lies outside of the range ±1.00, then there are at least 2 chances out of 3 that the reported rate is significantly different from the previous month's rate; if it exceeds the range ±1.96, then there are 95 chances out of 100 that the actual rates are different in the 2 months. It is evident that most of the observations fall within the ±1.00 range. The 95-percent test is very stringent; in fact, only two observations exceed the boundaries. This means that reported unemployment rates were statistically different from the previous month's rates only twice over this time period.

The 66-percent critical values appear to be a more sensible standard, especially as the test is biased in favor of finding no relationship.[7] The reported unemployment rate was significantly different from the previous month's rate, with 66-percent confidence, 12 times out of a total of 59, or 1 month out of every 5. The reported rate exceeded the upper bound 5 times and the lower bound, 7 times.

Much of the reported movement in the unemployment rate is not statistically significant. As a rule of thumb, the reported unemployment rate in Ohio must change by, plus or minus, 0.7 percent before it is considered to be significantly different from the previous month's rate with 66-percent confidence. The same figure, with 95-percent confidence, is ±1.3 percent.[8]

The CPS State unemployment rates are important data; they are provided on a regular and timely basis and are the best available point estimates of the capacity of a State's labor market. Despite the large amount of random error in each month's estimates, they also provide information about the direction in which a State's economy is heading. A moving average of the rate provides very reliable information about the trend of the State's business cycle. But the rate suffers as an indicator of social distress because it does not include people who are not part of the labor force and it weighs all employment equally (from 1 hour per week to 40 hours per week).

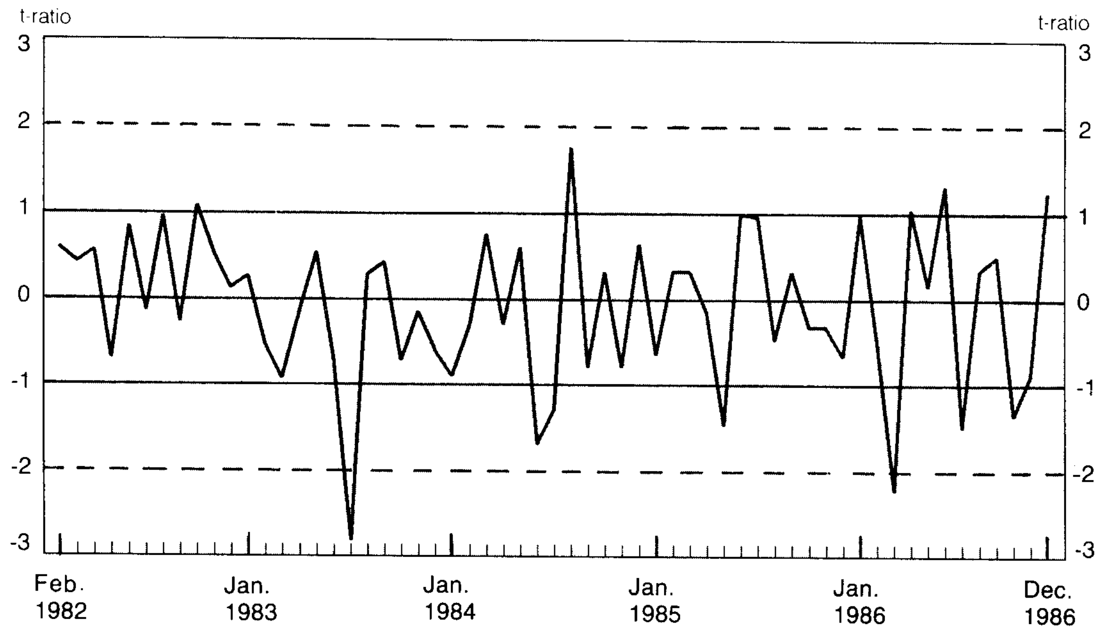
Conclusion

Small sample sizes for specific subpopulations in the national CPS yield relatively large variances for the reported unemployment rates. This can lead to a problem in using the rates as indicators of aggregate economic distress because changes in the rate which look large may be attributed to sampling error. This is an especially acute problem in using the reported unemployment rates for minority teens.

The analysis of the reported unemployment rates for the 11 survey States for April 1986 indicates that economists and planners should not use the unemployment rate to make finely detailed distinctions among the States. Confidence intervals are too wide to place much weight on finely drawn differences between States.

The analysis of longitudinal data for the State of Ohio indicates that most of the movement in the unemployment rate is spurious. In Ohio, the rate must change from 0.7

Chart 1. Difference between Ohio's month-to-month unemployment rates using values of two-tailed t-tests, February 1982-December 1986



NOTE: Dashed grid indicates 95 percent critical values, solid grid indicates 66 percent critical values percent to 1.3 percent before it can be called statistically significant. This would be a minimum for States with either smaller populations or lower unemployment rates.

The CPS showed that 423,000 Ohioans were unemployed in April 1986. The coefficient of variation indicates that there are 2 chances in 3 that the unemployment rate was in a range from 7.4 percent to 8.6 percent.[9] The reported rate in Ohio was 8.0 percent. If the next month's rate was within this range, then the new rate would not be statistically different from the old. This means that the change in the unemployment rate would have to exceed ±0.6 percent for the new rate to lie outside of April's interval (the May rate was 8.1 percent).

The unemployment rate remains the best point estimate of local labor market activity, but it should be used cautiously. A large amount of the change in the monthly unemployment rate appears, both cross-sectionally and longitudinally, to be attributed to random error. There is nothing wrong with the definition of unemployment that has been captured by the unemployment rate, or in the way data are collected by the CPS. The problem is with the way in which the rate is used and interpreted. □

—— *FOOTNOTES* ——

ACKNOWLEDGMENT: I thank John Bushery, Bureau of Census; Dixie Sommers and Roger Flemming, Ohio Bureau of Employment Services' Labor Market Information Division; Timothy Kinsella, Harry Margulis, Mark Salling, and Michael Spicer, College of Urban Affairs at Cleveland State University; and John Clair Thompson, University of Connecticut, for their help with this report. I also thank Cuyahoga County Commissioner Mary O. Boyle for demonstrating the importance of this issue on intergovernmental transfers within Ohio. While the help of these people is greatly appreciated, this article does not represent their views or the views of their agencies. This report was prepared with the financial support from a University Center Grant from the U.S. Department of Commerce, Economic Development Administration, and from the Ohio Board of Regents through its Urban University Program.

[1] A procedure developed by the Bureau of the Census is used to calculate the standard errors of the reported unemployment rates for the 11 survey States. See Charles D. Jones, "CPS Variances—Parameters Needed to Calculate State, Census Region, and Division Variances" (Bureau of the Census, 1985), unpublished memorandum. The 11 States are California, Florida, Illinois, Massachusetts, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, and Texas. See Kathleen Creighton and Robert Wilkinson, "Redesign of the Sample for the Current Population Survey," *Employment and Earnings*, April 1984, pp. 7–10.

The Current Population Survey is used to calculate annual labor market statistics for all of the States and the District of Columbia. The annual figures can usually be found in the May issue of *Employment and Earnings*. Unofficial estimates of annual averages of employment, unemployment, and the unemployment rate for metropolitan areas and a few central cities are published in *Geographic Profile of Employment and Unemployment, 1985*, Bulletin 2266 (Bureau of Labor Statistics, 1986).

[2] The utility of labor market data is judged by three standards: the ability to (1) measure labor market capacity; (2) estimate the position of the economy in the business cycle; and (3) provide information on aggregate economic distress. See Glen C. Cain, "The unemployment rate as an economic indicator," *Monthly Labor Review*, March 1979, pp. 24–35; and Julius Shiskin, "Employment and unemployment: the doughnut or the hole?" *Monthly Labor Review*, February 1976, pp. 3–10. Cain provides persuasive evidence that the unemployment rate performs best as a coincident cyclical indicator. As a cyclical indicator, change in the rate is more important than its absolute position. Others have indicated that it performs least well by the third standard. For example, see Terry F. Buss, "Unemployment Rates and Their Implications for Human Resource Planning," *Journal of Economic and Social Measurement*, No. 14, 1986, pp. 1–18; John C. Ries, "Unemployment in 1982: Beyond the Official Labor Force Statistics," *New England Economic Review*, May-June 1984, pp. 29–37; and Diane Wernecke, "Measuring Economic Hardship in the Labor Market," *American Economic Review*, May 1979, pp. 43–47.

[3] These results were obtained using a t-test for the difference between two means, evaluated at the 95-percent critical value. The standard error for month-to-month change in the unemployment rate was used in the denominator of the statistic. Solve the following for u_1:

$$[u_1 - u_2]/s = \pm 1.96$$

where: u_1 is the next month's rate,
u_2 is the current month's rate, and
s is the standard error of month-to-month change in the rate.

[4] See Creighton and Wilkinson, "Redesign of the Sample."

[5] Each month's reported unemployment rate has its own standard deviation and coefficient of variation (CV). To determine if the amount of dispersion was relatively constant over the period, the mean level of dispersion was measured by calculating the average coefficient of variation over the period. To measure the amount of variance in the standard error over the time series, the CV of each month's CV was calculated. This last measure assumes that each month's rate is independent from the previous rates. This is not strictly true, as unemployment rates are serial correlated. The CV of each month's CV should be read as a rough indication of the amount of month-to-month dispersion in the data.

[6] The correlation coefficient of the variance of month-to-month changes in the unemployment rate will be larger than that of month-to-month changes in the level of unemployment due to the behavior of entrants to the labor force. The number employed is fairly stable over the business cycle, compared with the number unemployed. Monthly fluctuations in the unemployment rate are more heavily influenced by flows into, or out of, unemployment from not-in-the-labor-force than into, or out of, employment. This implies that changes in the unemployment rate will be partially dampened by the relative stability of the number employed in the denominator of the statistic. This, in turn, implies that the monthly variances of the unemployment rate will be more closely correlated than those of the number unemployed.

However, it is expected that the difference in the two correlation coefficients will be extremely small. Two pieces of evidence are offered. First, if movements in the variance of the unemployment rate are dampened by the presence of the employed in the denominator, the average monthly coefficient of variation and the coefficient of variation of the monthly coefficients of variation of the rate would differ from that of the level of unemployment. Monthly Ohio data indicate that this is not true:

	Unemployment rate	*Number unemployed*	*Number employed*
Average CV	5.90	5.99	1.15
CV of CVs	9.64	9.34	3.33

The average of the monthly coefficients of variation for the unemployment rate is very close to that of the number unemployed, as is the coefficient of variation of the monthly coefficients of variation. Secondly, the standard deviation of changes in the monthly unemployment rate and the deviation of levels of the rate for the United States are equal. This implies that standard errors for the levels are close substitutes for changes.

The t-test used was of the form:

$$[u_1 - u_2]/[var_1 + var_2 - 2r(var_1 * var_2)^{.5}]^{.5}$$

where: u_1 is the month's unemployment rate,
u_2 is the previous month's rate,
var_1 is the variance in the month's unemployment rate,

var_2 is the variance of the previous month's rate, and
r is the correlation of the variances of the monthly levels of unemployment.

See Ohio Bureau of Employment Services, *Labor Market Review* (various issues).

[7] Conceding that the t-test used is biased in favor of accepting the null hypothesis, the data can be reexamined to see the effect of lowering the critical values. It has little impact on the results. If the critical value were lowered from 1.96, the 95-percent level, to 1.90, no additional observations would become significant. If the critical value were reduced from 1.00, the 66-percent value, to 0.95, four additional observations would become significant. In both cases, the results are below those expected if the events were purely random. If the behavior were random, we would expect to see significant results in 3 observations out of 59, with 95-percent confidence. This is equivalent to 1 month out of 20. Instead, the rates in only 2 months were significantly different from the previous month's rate, 1 in 30. The same is true at the 66-percent level. If the data were random, between 19 and 20 observations would be significant, 1 month in 4. Instead, only 12 are observed, 1 month in 5.

[8] The upper and lower critical values, at both 95- and 66-percent levels of confidence, were calculated for each month using the t-test of the difference in means, using the formula shown in footnote 6. The average of the difference between the upper bound and the reported unemployment rate was calculated.

[9] These results were obtained using the formula shown in footnote 6. The 66-percent critical value can be interpreted as meaning that if the observed rate exceeds the critical rate, there are 2 chances out of 3 that the observed rate is different from the previous month's rate. This corresponds to plus or minus one standard deviation from the observed rate. If the 95-percent critical values were used, the range would be from 6.9 to 9.1 percent and the change in the next month's unemployment rate would have to exceed ±1.1 percent.

A note on communications

The *Monthly Labor Review* welcomes communications that supplement, challenge, or expand on research published in its pages. To be considered for publication, communications should be factual and analytical, not polemical in tone. Communications should be addressed to the Editor-in-Chief, *Monthly Labor Review*, Bureau of Labor Statistics, U.S. Department of Labor, Washington, DC 20212.