The Review *of* Economics *and* Statistics

VOL. LXXXVII NOVEMBER 2005 NUMBER 4

SYMPOSIUM ON REGIONAL ECONOMIC INDICATORS

James H. Stock, Harvard University

Regional economists regularly confront the practical problem of using disparate data on economic activity to obtain a coherent picture of the state of the economy and the factors that influence regional economic growth. These data typically come from different sources, each with their own conceptual limitations, measurement frequencies, and historical spans. Thus an important practical challenge facing regional economists is combining these different sources of data to provide a timely and accurate measure of regional economic activity. The articles in this symposium make significant advances in the construction, dissemination, and use of monthly indexes of economic activity for the U.S. states. As the articles illustrate, these new indexes introduce new possibilities for forecasting and analyzing regional economic business cycles.

These papers were originally presented at a conference organized by the Federal Reserve Bank of Philadelphia on September 4–5, 2003.

CONSISTENT ECONOMIC INDEXES FOR THE 50 STATES

Theodore M. Crone and Alan Clayton-Matthews*

Abstract—In the late 1980s James Stock and Mark Watson developed for the U.S. economy an alternative coincident index to the one now published by the Conference Board. They used the Kalman filter to estimate a latent dynamic factor for the national economy and designated the common factor as the coincident index. This paper uses the Stock-Watson methodology to estimate a consistent set of coincident indexes for the 50 states. These indexes provide researchers with a comprehensive monthly measure of economic activity that can be used to examine a number of state and regional issues.

I. Introduction

IN the late 1980s James Stock and Mark Watson developed a coincident index for the U.S. economy as an alternative to the one published at that time by the Department of Commerce.[1] Stock and Watson's alternative index is the latent factor estimated in a dynamic single-factor model using the Kalman filter. State versions of the Stock-Watson index have been developed for the New England states, New York, Pennsylvania, New Jersey, Delaware, and Texas. This paper develops a consistent set of Stock-Watson coincident indexes for all 50 states. Besides their use in monitoring state economies, these indexes are useful in comparing the length, depth, and timing of recessions at the state level. They can also be useful in time series analysis as a composite measure of monthly economic activity.

A number of economic indicators, such as real gross state product, real personal income, or payroll employment, are commonly used to compare state economies. None, however, is completely satisfactory for business cycle analysis. Among the commonly used indicators, real gross state product is the most comprehensive measure of economic activity in a state, but it is available only annually and with a considerable lag. Though the real gross state product is a good metric for trend growth in a state's economy, the annual frequency of the data makes it an unsatisfactory indicator of state business cycles. At the national level a recession is characterized as a contraction in many economic activities, and the duration and depth of the contraction are factors in determining official recessions.[2] Turning points in national business cycles (peaks and troughs) are dated by months, and at least two official recessions have occurred within the span of a calendar year. Thus, the appropriate metric for defining state business cycles is a monthly indicator or set of monthly indicators. The advantage of a Stock-Watson-type index is that it combines several monthly indicators in a single measure of the state's economy.

Received for publication October 29, 2003. Revision accepted for publication May 26, 2004.

* Federal Reserve Bank of Philadelphia and University of Massachusetts Boston, respectively.

The views expressed here are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. We thank Jason Novak for excellent research assistance on this project. Any errors are the authors' alone.

[1] The traditional index is now published by the Conference Board.

[2] See Zarnowitz (1992).

II. The Dynamic Single-Factor (Stock-Watson) Model

The basic model discussed in this paper was developed by Stock and Watson (1989, 1991, and 1993). The structure of the model as applied here is

$$\Delta \mathbf{x}_t = \gamma(L)\, \Delta c_t + \mu_t, \tag{1}$$

$$\mathbf{D}(L)\mu_t = \varepsilon_t, \tag{2}$$

$$\phi(L)\, \Delta c_t = \eta_t, \tag{3}$$

where time series variables are subscripted with an index "t" indicating the time period with which they are associated. $\Delta \mathbf{x}_t$ is a $G \times 1$ vector of observable stationary series, usually series that have been logged and first-differenced to achieve stationarity. We use four state-level indicator series: nonagricultural employment, the unemployment rate, average hours worked in manufacturing, and real wage and salary disbursements. The unemployment rate is first-differenced, while the other three are logged and first-differenced. Each differenced indicator series is normalized by subtracting its sample mean and dividing by its sample standard deviation, so $\Delta \mathbf{x}_t$ has a zero mean, and each component has a unit variance. These indicator series are monthly except for real wage and salary disbursements, which are quarterly. We estimate a monthly model, treating the quarterly wage series as a running three-month sum with two out of every three months missing. The details of how this is done are explained in section IV.

Δc_t is a scalar latent stationary series that is common to the G observable series and, in this context, can be interpreted as deviations from the average growth rate of the economy. The vector μ_t consists of G mutually uncorrelated, mean-zero, stationary autoregressive (AR) processes. The $G \times 1$ vector ε_t and the scalar η_t comprise $G + 1$ mutually uncorrelated white-noise processes. The symbol L stands for the lag operator, that is, $L^k x_t = x_{t-k}$. All polynomials in the lag operator are of finite order, and all but the polynomials in $\gamma(L)$ are one-sided.

The parameters of the model can be expressed as follows:

$$\gamma(L) \equiv [\gamma_1(L), \gamma_2(L), \ldots, \gamma_G(L)]', \tag{4a}$$

where

$$\gamma_g(L) \equiv \sum_{s=a}^{b} \gamma_{gs} L^s, \tag{4b}$$

$$\mathbf{D}(L) \equiv \mathrm{diag}[d_1(L), d_2(L), \ldots, d_G(L)]', \tag{5a}$$

where

$$d_g(L) \equiv 1 - d_{g1}L - d_{g2}L^2 - \cdots; \tag{5b}$$

$$\phi(L) \equiv 1 - \phi_1 L - \phi_2 L^2 - \cdots; \tag{6}$$

$$\sum \equiv \mathrm{cov}([\varepsilon_t', \eta_t]') = \mathrm{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_G^2, \sigma_\eta^2]. \tag{7}$$

According to equation (4b), the common component may enter equation (1) with one or more leads ($a < 0$) or lags ($b > 0$). The lag polynomial matrix $\mathbf{D}(L)$ is assumed to be diagonal, so that the μ_t's in different equations in (2) are contemporaneously and serially uncorrelated with one another. In equations (3) and (6), we use a second-order autoregressive process (AR) for the common state. The orders of the other lag polynomials differ from state to state and are addressed in the discussion on model specification and estimation below.

The model can be interpreted as a time series version of a factor analysis model, where the first difference of the unobserved state, Δc_t, represents the common factor in the indicators, $\Delta \mathbf{x}_t$. The differenced state, hereinafter simply referred to as the common state, follows an AR process. The μ_t, components of the observed series in equation (2) comprise what are called the idiosyncratic portions of the observed series, or, in factor analysis, the unique factors — as opposed to the common factor. The zero-mean normalization of the $\Delta \mathbf{x}_t$ series, typical in factor-analysis models, eliminates the need for constants in equations (1) and (3). The unit-variance normalization of the indicator series in $\Delta \mathbf{x}_t$ is not necessary to identify the model but is a convenience that scales the data and therefore the parameters. The scaling may also aid in the numerical optimization used to estimate the parameters. The scale of the $\gamma(L)$ coefficients is fixed by setting the variance of η_t to unity, that is, $\sigma_\eta^2 = 1$, and the timing of the coincident index is fixed by setting all but one of the elements of $\gamma(L)$ to 0 in one of the equations in (1).

Maximum likelihood estimation of the parameters of the system in equations (1)–(3) and estimation of the smoothed state is accomplished by representing the system in state-space form and using the Kalman filter.[3] There are two equivalent ways to form the state-space system. They yield the same estimates and likelihood but differ in computation time. One way is to treat equation (1) as the measurement equation, and equations (2) and (3) as the state equation, with the idiosyncratic components μ_t as part of the state vector along with Δc_t. The second way is to eliminate equation (2) from the system by multiplying both sides of equation (1) by $\mathbf{D}(L)$ and solving for $\Delta \mathbf{x}_t$. This can greatly reduce the dimension of the state vector, decreasing the computation time significantly. The formation of the state-space system is described in detail by Stock and Watson (1991). Maximum likelihood estimation of the state space

[3] If the assumption of normality of ε and η is incorrect, then the estimates are not efficient, but they are still consistent.

model is described in several sources, including Hamilton (1994).

Often, the purpose of "running" the model is to get estimates of the common state or forecasts of the common state. $\Delta\hat{c}_{t|s}$ refers to the estimate of Δc_t with information through time period s. These estimates are linear functions of the observables $\Delta\mathbf{x}_t$. The coefficients of these linear functions, called scoring coefficients in factor analysis, are called filters in time-series analysis. This linear function may be written as:

$$\Delta c_{t|s} = \sum_{g=1}^{G} {}_{t,s}m_g(L)\ \Delta x_{g,t}, \qquad \text{where}$$

$$ {}_{t,s}m_g(L) = \sum_{k=t-s}^{t} {}_{t,s}m_{g,k}\, L^k. \tag{8}$$

The prefix subscript "t,s" on the filter indicates that the coefficients of the filter depend on both t and s, although if $\Delta\mathbf{x}_t$ is stationary, the filter depends only on the difference $s - t$ as $t \rightarrow \infty$. Furthermore, the filter converges to a symmetric two-sided filter as $|s - t| \rightarrow \infty$. In practice, the convergence of the filter in these ways is quite rapid as $t \gg 0$ and $|s - t| \gg 0$. The term "Kalman filter" refers not only to the set of recursive equations used to calculate the common state but also to the filter that yields $\Delta\hat{c}_{t|t}$, while "Kalman smoother" refers to the filter that yields $\Delta\hat{c}_{t|T}$, where T refers to the full set of information. We use the Kalman smoother for estimates of the common state and the Kalman filter for both evaluating the relative contributions of the indicator variables and retrending the common state to grow at the same rate as the real gross state product.

III. Retrending the Common State

Integrating (reverse-differencing and exponentiating of) the common state gives an estimate of the underlying state of the economy.[4] By construction, without constants in equations (1) and (3), this state would be driftless. Furthermore, the identification restriction $\sigma_\eta^2 = 1$ affects the scale of the growth rate of the underlying state. If the model estimates are used to construct economic indexes, the estimates of the common state given by the Kalman filter need to be recalibrated in some useful way.

Two common recalibration approaches use a linear transformation of the Kalman filter estimates or Kalman smoother estimates of the state variable, for example, $a + b\Delta\hat{c}_{t|t}$ or $a + b\Delta\hat{c}_{t|T}$, which is then integrated (and exponentiated if appropriate). One approach (Stock & Watson, 1991), analogous to the method used by the Conference Board and developed at the Bureau of Economic Analysis (Green & Beckman, 1993), gives the resulting index both a trend rate of growth and a variance around that trend that is a weighted average of the trends of the indicators, with weights proportional to the contributions of the indicators in the Kalman filter or smoother. A second approach (Clayton-Matthews & Stock 1998/1999) chooses the linear transformation that gives the resulting index the same trend and variance around the trend as some other series of interest. The trend in the state indexes developed in this paper is set equal to the trend in each state's gross state product (GSP). Here we adopt a blend of the two approaches, resulting in an index with a trend rate of growth equal to that of each state's real gross state product over the period of the estimation, and a variance around that trend that is a weighted average of the indicators' contributions to the common state. Thus, the growth in the retrended index can be interpreted as the underlying growth rate of the economy.

In the linear function of the trendless state, the constant "a" is the average monthly growth rate of the state's real gross state product over the estimation period. The coefficient "b" is defined in the following manner. Define $B_g = m_g(1)/\sigma_g$, where $m_g(1)$ is the sum of the Kalman filter coefficients for the gth index variable and σ_g is the standard deviation of the logged-differenced indicator variable used in the normalization to calculate Δx_g. Then the coefficient "b" is given by $b = 1/\sum_g B_g$. As explained in detail in Clayton-Matthews and Stock (1998/1999), this scheme mimics the approach developed by the U.S. BEA and used by the Conference Board in forming the U.S. current and leading indexes.

IV. Handling Mixed Monthly and Quarterly Frequencies

The original Stock and Watson model was developed for variables, all of which had the same frequency. In some applications, the indicator series may be of mixed frequencies. In our case, three of the series are monthly, and one, real wage and salary disbursements, is quarterly. One way to handle this situation would have been to collapse the monthly series into quarterly observations. This would have entailed some loss of information and, perhaps even more important, would have resulted in indexes that were less timely in giving up-to-date information on the state of the states' economies. Instead, we treated the quarterly series as a moving three-month sum of an unobserved monthly series, that is, $x_t = z_t + z_{t-1} + z_{t-2}$, a strategy described by Zadrozny (1990). The quarterly change in wage and salary disbursements, $\Delta x_t \equiv x_t - x_{t-3}$, is related to the monthlychange in the underlying monthly series, $\Delta z_t \equiv z_t - z_{t-1}$, in the following way:

[4] The result is an index that we set to 100 in July 1992.

$$\begin{aligned}\Delta x_t &= (z_t + z_{t-1} + z_{t-2}) - (z_{t-3} + z_{t-4} + z_{t-5}))) \\ &= (z_t - z_{t-3}) + (z_{t-1} - z_{t-4}) + (z_{t-2} - z_{t-5}) \\ &= \Omega(L)\ \Delta z_t,\end{aligned}$$

where $\Omega(L) = 1 + 2L + 3L^2 + 2L^3 + L^4$. (Note that $z_t - z_{t-3} = \Delta z_t + \Delta z_{t-1} + \Delta z_{t-2}$, etc.) The measurement equation for z in equation (1) could be expressed as $\Delta z_t = \gamma(L)\ \Delta c_t + \mu_t^*$, where μ_t^* represents z's monthly idiosyncratic component. Then applying $\Omega(L)$ to both sides gives

$$\Delta x_t = \gamma(L)\Omega(L)\ \Delta c_t + \mu_t, \qquad (1')$$

where $\mu_t = \Omega(L)\mu_t^*$. The coefficients on the common state in the wage and salary models are the estimates of the $\gamma(L)$ parameters, which estimate the effect of the common state, Δc_t (and its lags, if any), on the unobserved monthly series, Δz_t. The idiosyncratic component of the quarterly change in wage and salary disbursements, μ_t, on the other hand, is modeled as a quarterly series, and so its estimated autoregressive structure in equations (2) and (5a,b) above should be interpreted accordingly.

Defined in this way, the measurement series Δx_t is observed every third month and is missing two out of every three months. The Kalman filter is modified to handle missing data by omitting the measurement equation for wage and salary disbursements in the months for which the data are missing. The procedure, described in Zadrozny (1990), is implemented simply by changing the dimensions of the relevant state-space matrices month by month during the Kalman filter recursion, as needed.[5]

V. A Consistent Set of Indexes for the 50 States

If state coincident indexes are used to compare business cycles across states, a certain degree of consistency must be imposed on their construction. At a minimum, a set of consistent state indexes should meet the following criteria:

1. The indexes should be constructed from the same set of indicators for each state.
2. In models with leads or lags, the timing of the index (the latent dynamic factor) should coincide with the timing of the same indicator variable in each state. This implies that for every state the measurement equation for one indicator variable (the same one for each state) includes *only* the contemporaneous value of the common factor.
3. The trend for each state index should correspond to the trend of the same variable for each state or to a weighted trend of several variables where the weights for the corresponding variables are the same for each state.[6]

Most previous composite state indexes have been based on three or four indicators.[7] All the state indexes developed to date include monthly payroll employment and the state unemployment rate. Other monthly variables include average hours worked in manufacturing (Clayton-Matthews, Kodrzycki, & Swaine, 1994; Crone, 2000; & Orr, Rich, & Rosen, 1999) and state withholding taxes and sales taxes (Clayton-Matthews & Stock, 1998/1999). Quarterly variables used in Stock-Watson-type state indexes include real earnings (Orr, Rich, & Rosen, 1999) and real personal income minus transfer payments (Crone, 2000).

We have identified three monthly indicators and one quarterly indicator available for inclusion in the coincident indexes for all 50 states. The three monthly series are available on a consistent basis for most states since 1978, so our state indexes are estimated from that year.

Nonagricultural payroll employment. This series is produced by the Bureau of Labor Statistics (BLS) in cooperation with the individual states. It is the most reliable employment series published for all the states. Its most obvious drawback is that it does not include the self-employed or farmworkers. Therefore, it may be a less reliable indicator of economic activity in states whose economies are dominated by agriculture. The measurement equations for nonagricultural employment in all our state models include only the contemporaneous value of the common factor. Thus, the timing of the final index is set to coincide with the timing of the employment series.

Unemployment rate. This series is also produced by the BLS in cooperation with the states. It is published on a seasonally adjusted basis from 1978 for all the states except California. The California unemployment rate series published by the BLS begins in 1980. Data for the previous two

[5] Any missing data can be handled in this fashion. For example, state quarterly wage and salary disbursements lag the other indicators by several months, and so they are treated as missing in the last several months of each index update. The software we used to estimate these models, developed by Clayton-Matthews (2001), handles both monthly and quarterly mixed frequencies and missing data.

[6] Crone (1998/1999) produced a set of indexes for the 48 contiguous states based on payroll employment, the unemployment rate, and average hours worked in manufacturing. These indexes satisfied the first two criteria, but not the third. The indexes in that study had two major limitations: The underlying data series were all related to employment rather than to income or value added, and the trend for each state index was calculated according to the original Stock-Watson model, that is, the weighted average trend of the components. Because the weights differed by state, the calibration of the trend was not consistent across states.

[7] See Clayton-Matthews and Stock (1998/1999); Crone (2000); and Orr, Rich, and Rosen (1999). The state indexes in Crone (1998/1999) were constructed using only three indicator variables, as were the indexes for Connecticut, Maine, New Hampshire, Rhode Island, and Vermont in Clayton-Matthews, Kodrzycki, and Swaine (1994).

years were obtained from the California Employment Development Department (www.calmis.ca.gov). The data used to produce state unemployment rates are from the current population survey, the payroll employment survey, state population estimates, and unemployment claims. Because the peak of the unemployment rate often lags the trough in economic activity at the national level, the measurement equation for the unemployment rate in several of our state models includes lags of the common factor as well as the current value. Also, the unemployment rate is entered in the measurement equation as the standardized first difference rather than the standardized log difference.

Average hours worked in manufacturing. Stock and Watson's national index and the traditional coincident index published by the Conference Board include industrial production. There is no comparable measure of industrial output at the state level, but average hours worked in manufacturing, from the same survey as payroll employment, is used in our model as an indicator of industrial activity at the state level. These data are not published on a seasonally adjusted basis, so we seasonally adjusted the series for each state.[8]

Real wage and salary disbursements. Personal income and its components are available at the state level on a quarterly basis from the Bureau of Economic Analysis (BEA). The state indexes estimated for this project include real wage and salary disbursements, the major component of personal income.[9] The quarterly wage and salary disbursements reported by the BEA on a seasonally adjusted basis are deflated by the national CPI-U to obtain real wages and salaries. We do not include in our state models proprietors' income or rent, interest, and dividends. In several farm states in our sample period, farm income represented more than 50% of proprietors' income in some years, and farm income can have a very irregular pattern due in part to government price support programs. Therefore we excluded proprietors' income from the income measure in our models. Rent, interest, and dividends, unlike wages and salaries and proprietors' income, are reported by state of residence rather than by the state in which the income is generated. Because coincident indexes are meant to track economic activity or output *in the state,* rent, interest, and dividends were also excluded from our income measure. Because wages and salaries are reported only on a quarterly basis, the lag structure in the measurement equation for this variable is adjusted as described above and in Clayton-Matthews (2001).

VI. Seasonal Adjustment and Smoothing of the Data

The data used in a Stock-Watson-type model are assumed to be seasonally adjusted if the original data have a seasonal component. The unemployment rate and quarterly wages and salaries are seasonally adjusted by the BLS and the BEA, respectively. The BLS publishes a seasonally adjusted series of nonagricultural employment for the states beginning in 1982; prior to that time only nonseasonally adjusted data are available. Since our indexes are estimated from 1978, we have independently adjusted the nonagricultural employment data for seasonal variation using the X-11 procedure. The BLS does not publish a seasonally adjusted series for average hours worked in manufacturing at the state level. Therefore, we also seasonally adjusted those data.

No smoothing of the data (beyond seasonal adjustment) is theoretically necessary before estimating a Stock-Watson-type model, because the Kalman filter procedure smooths the series over time. However, data with high-frequency noise might require a large number of lags or leads in the measurement and/or error equations in the model to adequately estimate a common factor. For this reason, Clayton-Matthews and Stock (1998/1999) used a bandpass nine-period moving-average filter on some series before estimating their Massachusetts model.

Two series in our state models also required some pre-estimation smoothing beyond the normal seasonal adjustment to satisfactorily estimate the common factor. Average hours worked in manufacturing are estimated from a survey taken during one week in each month. The estimates are particularly susceptible to unusual weather, work stoppages, and other exceptional factors; and the series exhibits a good deal of high-frequency noise. Therefore, we smoothed the manufacturing-hours series in the process of seasonal adjustment. The X-11 seasonal adjustment uses a moving-average procedure to decompose the monthly data into three components—trend-cycle, seasonal, and irregular. The seasonally adjusted series includes both the trend-cycle and irregular components. One option for smoothing a data series in the X-11 procedure is to weight the extreme values of the *irregular* component based on their distance from the mean in standard-deviation units.[10] We used this option to smooth the manufacturing hours series. In 18 states we also used the X-11 option to produce a nonfarm employment series smooth enough to estimate the common factor. (See Appendix.)

[8] In Kansas the data on average hours worked in manufacturing are not available prior to 1979, and in Indiana the data are not available prior to 1989.

[9] The national coincident indexes include monthly personal income minus transfer payments. Transfer payments are not considered payment for current production.

[10] The irregular component in the multiplicative version of the X-11 seasonal adjustment procedure has a value that varies around 1 and is used to adjust the trend-cycle component, measured in units of the original data series.

For the series on average hours worked in manufacturing, the switch from the Standard Industrial Classification System (SIC) to the North American Industrial Classification System (NAICS) in January 2001 introduced a discontinuity. This issue was resolved by treating hours worked in January 2001 as missing data. Because the model is estimated using the log difference of manufacturing hours, this results in the omission of the hours data in calculating the common factor for January and February 2001. In Washington state weather-related and strike-related declines in hours worked in 1980 and 1989 also resulted in coding certain months in those years as missing data. (See Appendix.)

Several anomalies also appeared in the wage and salary data and in the unemployment series in some states. Nationally, wages and salaries declined more than 4% in real terms in the first quarter of 1993. That was more than four times greater than the average absolute percentage change since 1978 and more than 50% greater than the next highest absolute percentage change. This quarterly decline in wages and salaries is attributable in part to the movement of bonuses and other one-time compensation into 1992 in anticipation of an income tax increase proposed by the new administration.[11] This did not represent a shift in economic activity but rather a shift in the timing of compensation and, therefore, should not be reflected in the coincident index. The impact of this shift was concentrated in 12 states in which the quarterly decline in real wages and salaries was greater than 4%. The wage and salary data for the first quarter of 1993 were eliminated (that is, designated as missing data) in those states before estimating the coincident index model.[12] Wage and salary data were also eliminated for two quarters in West Virginia (1978:I and 1981:II) because strikes by coal miners reduced real wage and salary disbursements more than 10% in those two quarters. (See Appendix.)

In eight states there were one or more shifts of 1.5 percentage points or more in the level of the unemployment rate in a single month. Some of these shifts may have been the result of a change in the current population sample. In two instances the shift was reversed in 6 to 10 months.[13] In the eight states combined we eliminated a total of 12 months of unemployment data before estimating the coincident index models. (See Appendix.)

The preestimation smoothing of some data series and the elimination of some anomalous data points prevent outliers from distorting the estimation of a relatively smooth coincident index from the Stock-Watson model.

[11] For a discussion of this issue see Feldstein and Feenberg (1996).

[12] In New York state real wages and salaries declined more than 10% in the first quarter of 1993 and increased more than 4% in the fourth quarter of 1992; so the fourth-quarter 1992 data were also eliminated from the wage and salary series in New York state.

[13] Opposing shifts appear in the data for Oklahoma in June 1980 and December 1980 and in the data for Georgia in March 1991 and January 1992.

Table 1.—Contributions of Indicator Variables to State Indexes

Index	Relative Contribution of Indicator Variable (Number of States) <10%	10%–25%	>50%	Median Relative Contribution (%)
Employment	0	2	30	57.4
Unemployment rate	7	33	2	17.7
Mfg. hours	40	9	0	4.9
Real wages and salaries	13	24	0	14.3

VII. Model Specification and Estimation

Except for the restriction that the measurement equation for employment must contain only the contemporaneous value of the common factor, no a priori restrictions were placed on the lag structure of any of the measurement equations in our state models.[14] We began our search for the most appropriate model for each state with a parsimonious specification that included only the contemporaneous value of the common factor in each of the measurement equations. Except for the employment equation, we then added leads and lags in the measurement equations to improve the model based on four criteria.

First, because we assume that the underlying state of the economy is relatively smooth, we required that the final coincident index be smooth. At a minimum this requirement rules out a negative first-order autocorrelation in Δc_t. In all our final state indexes the first-order autocorrelation of the change in the index is positive and significant; and the autocorrelation coefficients range from 0.62 to 0.98.

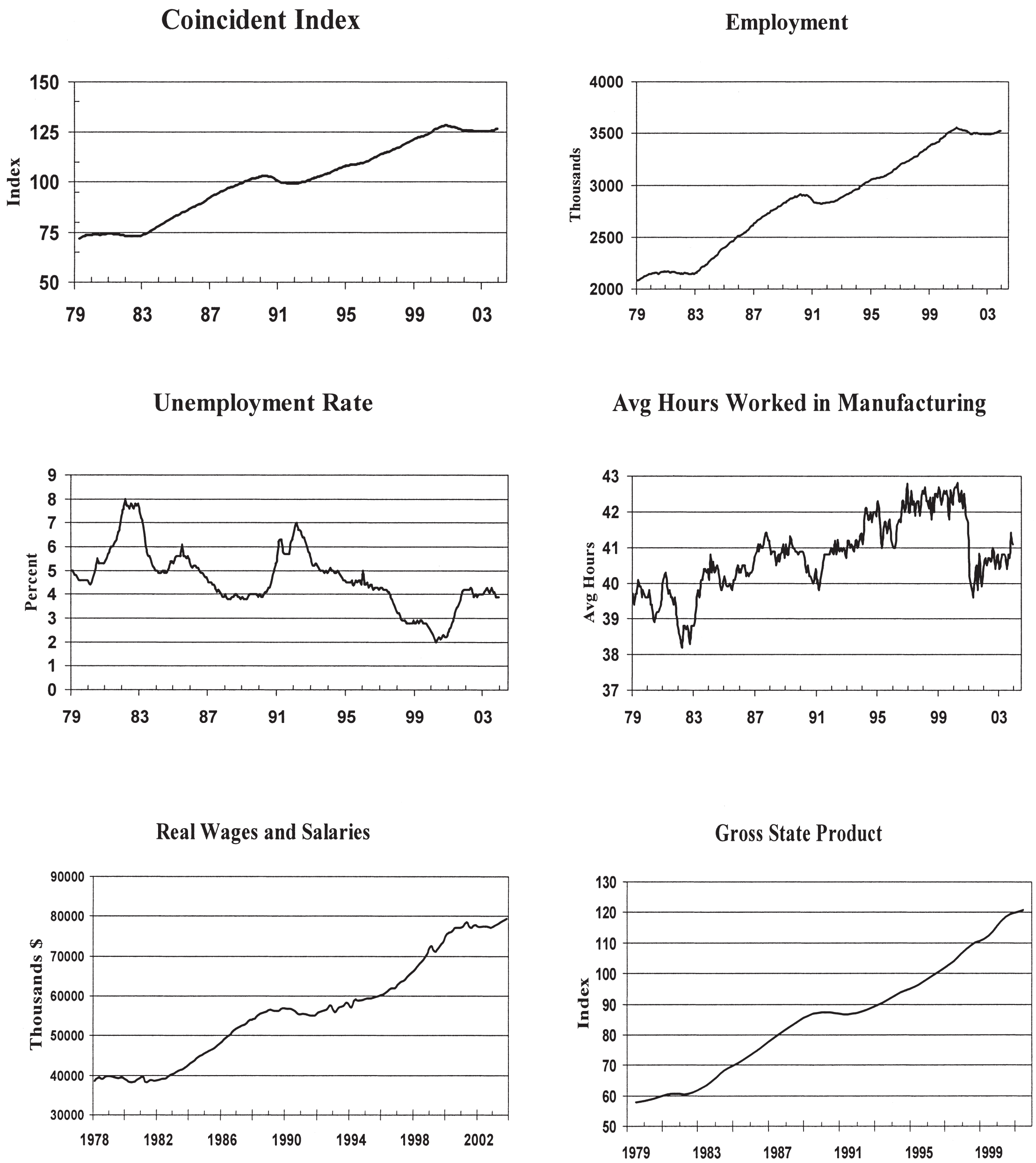
Second, to the extent possible we chose models in which the coefficients on the common factor were statistically significant. About 86% of the coefficients in the measurement equations were significant at the 5% level, and about 88% were significant at the 10% level. In the final models, 15 states had one or more lags of the common factor in the measurement equation for unemployment, and 12 states had lags of the state variable in the measurement equation for wage and salary disbursements. In the measurement equation for average hours worked in manufacturing, 25 states had one or more leads and six states had one or more lags of the common factor.[15]

Third, we chose models in which the contribution of the indicator variables to the estimated change in the common factor was not too heavily concentrated on only one of the indicator variables. We defined the contribu-

[14] For consistency, we specified the autoregressive process for the common factor as an AR(2) process in the model for each state.

[15] See the tables in Crone and Clayton-Matthews, 2004 (Appendix B).

Figure 1.—Virginia Coincident Index and Components



tion of the gth indicator as $|m_g(1)|/\Sigma_{g=1}^{G}|m_g(1)|$, where $m_g(1)$ is the sum of the Kalman filter coefficients for the gth indicator. We used the absolute value because the unemployment rate enters the estimation of the common state negatively. Thus defined, these contributions add to 1. In most states the relative contributions of the observed variables to the monthly change in the common factor were well distributed. Changes in employment were generally the largest contributor to monthly changes in the common factor (Δc_t). In 30 states the change in employment contributed more than 50% to the change in the common factor; the median contribution of employment was about 57%. (See table 1.) In most states the relative contribution of the unemployment rate to the change in the common factor was between 10 and 25%, and the median contribution was almost 18%. In approximately half the states real wages and salaries contributed between 10 and 25% to the change in the common factor, and the median contribution was approximately 14%. Hours worked in manufacturing was almost always the least important contributor to the change in the common factor; the median contribution of hours worked was approximately 5%.

Fourth, given the satisfaction of the first three criteria, we chose the model for each state that came closest to satisfying the test for a *single* dynamic factor that was used by Stock and Watson (1989, 1991). They check the assumption

TABLE 2.—RESULTS OF THE ESTIMATION OF THE LATENT DYNAMIC FACTOR FOR VIRGINIA

Measurement Equation	Variable	Coefficient	Asymptotic Standard Error	*t*-Statistic
Estimated Coefficients of the Common Factor *C* with Leads (+) and Lags (−)				
Employment	C_t	0.458	0.065	7.02
Unemployment rate	C_t	−0.305	0.060	−5.05
Average weekly mfg. hours	C_t	0.059	0.028	2.12
Wages and salaries	C_t	0.131	0.041	3.24
	C_{t-1}	−0.102	0.040	−2.54
Estimated Coefficients for Lags in the Autoregressive Equations for the Error Terms				
Employment	(1)	−0.069	0.103	−0.67
Unemployment rate	(1)	0.010	0.062	0.16
	(2)	0.193	0.061	3.17
	(3)	0.153	0.062	2.47
Average weekly mfg. hours	(1)	−0.332	0.059	−5.66
	(2)	−0.136	0.062	−2.21
	(3)	0.008	0.062	0.12
	(4)	0.026	0.059	0.43
Wages and salaries	(1)	−0.217	0.107	−2.02
	(2)	0.095	0.107	0.88
	(3)	0.148	0.110	1.34
	(4)	0.191	0.107	1.79
Estimated Coefficients in the Autoregressive Equation for the Common Factor				
Lag 1		0.349	0.103	3.39
Lag 2		0.521	0.107	4.88

F-statistics for Tests of Single-Index Model*

	Error Employment	Error Un. Rate	Error Avg. Weekly Mfg. Hours	Error Wages and Salaries
Error(employment)	1.447	0.509	0.963	0.971
Error(unemployment rate)	0.223	1.393	1.224	0.834
Error(mfg. hours)	0.580	1.470	0.091	0.704
Error(wages and salaries)	1.143	0.739	0.983	0.807
Employment	1.918	1.088	0.538	0.814
Unemployment rate	0.285	1.343	1.353	1.141
Average weekly mfg. hours	0.788	1.482	0.101	0.866
Wages and salaries	1.210	0.723	2.478	0.759

Relative Contribution of Observed Variables to Monthly Changes in the Common Factor Based on the Proportion of the Sum of the Cumulative Dynamic Multipliers Due to the Variable

	Employment	Unemployment Rate	Average Weekly Mfg. Hours	Wages and Salaries
Relative contribution (%)	70.48	10.87	5.08	13.56

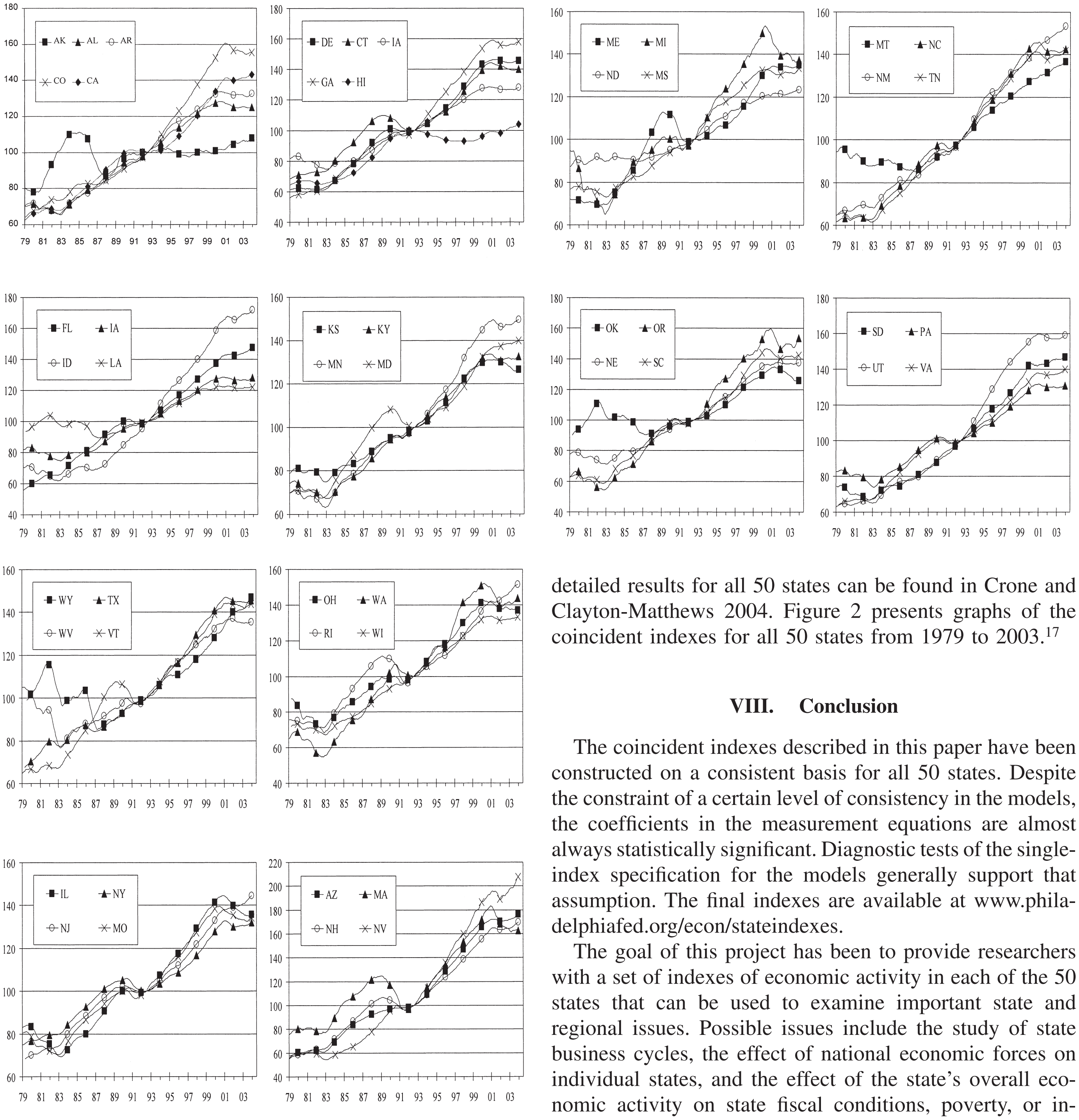
*F-statistics for the hypothesis that the coefficients are jointly zero in regressions of errors in the measurement equation against six lags of the errors from the various measurement equations or six lags of the measurement variables.

of a single common factor by testing whether the disturbances in the measurement equations can be predicted by past values of the indicator variables or past values of the errors from the measurement equations. In a series of tests, they regress the errors from each measurement equation sequentially on a constant and six lags of the errors from each of the measurement equations and six lags of the indicator variables. If the single-index model is the proper specification, the coefficients on the lags should jointly equal 0. We applied the same test to our state models and report the *F*-statistics for rejecting the hypothesis that the coefficients are zero. [See tables in Crone & Clayton-Matthews, (2004, Appendix B).] Of the 1600 *F*-statistics reported for the 50 states, 1,510 indicate that the assumption of a single common factor is correct.

Figure 1 presents graphs of the coincident index for Virginia and the components of the index.[16] The results of the estimation of the equations for the Virginia index are reported in table 2. All the coefficients in the measurement equations for Virginia are significant at the 5% level. With one exception, the tests for a single common factor do not suggest the presence of a second factor. The one exception is the regression of the error term from the equation on

[16] The discontinuity in the series on average hours worked in manufacturing in January 2001 is clearly visible in the relevant panel in figure 1.

FIGURE 2.—COINCIDENT INDICES FOR 50 STATES



average weekly hours on lags of real wages and salaries. The *F*-statistic from this regression does not reject the presence of a second factor. In terms of the contribution of the indicator variables to the coincident index, nonfarm employment has a higher contribution than the median for all 50 states, and the unemployment rate has a somewhat lower contribution. Each of the indicator variables contributed 5% or more to the change in the common factor. The detailed results for all 50 states can be found in Crone and Clayton-Matthews 2004. Figure 2 presents graphs of the coincident indexes for all 50 states from 1979 to 2003.[17]

VIII. Conclusion

The coincident indexes described in this paper have been constructed on a consistent basis for all 50 states. Despite the constraint of a certain level of consistency in the models, the coefficients in the measurement equations are almost always statistically significant. Diagnostic tests of the single-index specification for the models generally support that assumption. The final indexes are available at www.philadelphiafed.org/econ/stateindexes.

The goal of this project has been to provide researchers with a set of indexes of economic activity in each of the 50 states that can be used to examine important state and regional issues. Possible issues include the study of state business cycles, the effect of national economic forces on individual states, and the effect of the state's overall economic activity on state fiscal conditions, poverty, or in-migration. The indexes have already been used to group states based on the similarities of their business cycles (Crone, 2004) and to compare the timing of state business cycles (Crone, 2003; Owyang et al., 2003). The indexes have also been used to estimate the effect of regional economic activity on bank conditions (Daly et al., 2004).

[17] Because some state indexes were very similar, it was not possible to present the states in alphabetical order or by region.

And a similar index has been used to estimate the effect of state economic activity on tax revenues (Bram et al., 2004).

In the absence of monthly GSP data, these indexes are the most comprehensive measure of economic activity available for all 50 states. They provide a consistent measure of economic activity across the states and a single measure of economic output at the state level.

REFERENCES

Bram, Jason, Andrew Haughwout, James Orr, Robert Rich, and Rae Rosen, "The Linkage between Regional Economic Indexes and Tax Bases: Evidence from New York," Federal Reserve Bank of New York staff report no. 188 (2004).

Clayton-Matthews, Alan, *DSFM Manual* (version 4/17/01), University of Massachusetts Boston mimeo, (2001).

Clayton-Matthews, Alan, Yolanda K. Kodrzycki, and Daniel Swaine, "Indexes of Economic Indicators: What Can They Tell Us about the New England Economy?" *New England Economic Review, Federal Reserve Bank of Boston* (November/December 1994), 17–41.

Clayton-Matthews, Alan, and James H. Stock, "An Application of the Stock-Watson Index Methodology to the Massachusetts Economy," *Journal of Economic and Social Measurement* 25 (1998/1999), 183–233.

Crone, Theodore M., "Using State Indexes to Define Economic Regions in the US," *Journal of Economic and Social Measurement* 25 (1998/1999), 259–275.

———, "A New Look at Economic Indexes for the States in the Third District," *Business Review, Federal Reserve Bank of Philadelphia* (November/December), 2000, 3–14.

———, "A Redefinition of Economic Regions in the U.S.," Federal Reserve Bank of Philadelphia. Working paper 04-12 (September 2004).

———, "The Long and the Short of It: Recent Trends and Cycles in the Third District States," *Business Review, Federal Reserve Bank of Philadelphia* (Third Quarter 2003), 29–37.

Crone, Theodore M., and Alan Clayton-Matthews, "Consistent Economic Indexes for the 50 States," *Federal Reserve Bank of Philadelphia* working paper no. 04-9 (2004) (available at www.philadelphiafed.org/econ/stateindexes).

Daly, Mary, John Krainer, and Jose A. Lopez, "Does Regional Economic Performance Affect Bank Conditions? New Analysis of an Old Question," Federal Reserve Bank of San Francisco working paper 2004-01 (2004).

Feldstein, Martin, and Daniel Feenberg, "The Effect of Increased Tax Rates on Taxable Income and Economic Efficiency: A Preliminary Analysis of the 1993 Tax Rate Increases" (pp. 89–117), in James M. Poterba (Ed)., *Tax Policy and the Economy*, Vol. 10 (NBER, MIT Press, 1996).

Green, G. R., and B. A. Beckman, "Business Cycle Indicators: Upcoming Revision of the Composite Indexes," *Survey of Current Business* (October 1993), 44–51.

Hamilton, James D., *Time Series Analysis* (Princeton: Princeton University Press, 1994).

Orr, James, Robert Rich, and Rae Rosen, "Two New Indexes Offer a Broad View of Economic Activity in the New York–New Jersey Region," *Current Issues in Economics and Finance, Federal Reserve Bank of New York,* 5: 14 (October 1999).

Owyang, Michael T., Jeremy Piger, and Howard J. Wall, "Business Cycle Phases in U.S. States," Federal Reserve Bank of St. Louis working paper no. 2003-011D (2003).

Stock, James H., and Mark W. Watson. "New Indexes of Coincident and Leading Economic Indicators" *NBER Macroeconomics Annual,* 1989, pp. 351–394.

———, "A Probability Model of the Coincident Economic Indicators" in K. Lahiri and G. H. Moore, eds. *Leading Economic Indicators: New Approaches and Forecasting Records.* (Cambridge: Cambridge University Press, 1991, pp. 63–89).

———, "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience," James H. Stock and Mark W. Watson, eds. in *Business Cycles, Indicators, and Forecasting*. (Chicago: University of Chicago Press, 1993).

Zadrozny, Peter A., "Estimating a Multivariate ARMA Model with Mixed-Frequency Data: An Application to Forecasting U.S. GNP at Monthly Intervals," Federal Reserve Bank of Atlanta, working Paper 90-6 (1990).

Zarnowitz, Victor, *Business Cycles: Theory, History, Indicators, and Forecasting* (Chicago: University of Chicago Press, 1992).

APPENDIX

Seasonal Adjustment and Preestimation Smoothing of the Data

We seasonally adjusted the average hours worked in manufacturing and the nonfarm employment data for all the states, because these two series are not available on a seasonally adjusted basis for the time period over which we estimated our models. The preestimation smoothing of the data is described in this appendix.

1. Hours Worked in Manufacturing

The X-11 seasonal adjustment uses a moving average procedure to decompose the monthly data into three components—trend-cycle, seasonal, and irregular. One option for smoothing a series using X-11 is to weight the extreme values of the *irregular* component based on their distance from the mean in standard-deviation units. The default for this option in SAS is to give full weight to those irregular components that are less than 1.5 standard deviations from the mean and gradually reduce the weights to 0 when the irregular component is more than 2.5 standard deviations from the mean. In all states except those listed below, we used the SAS default option to smooth the data. In the following states the irregular components that were between one standard deviation and two standard deviations from the mean were given a weight between 1 and 0. Those more than 2 standard deviations from the mean were given a weight of 0: Alabama, Alaska, Delaware, Florida, Georgia, Hawaii, Kentucky, Louisiana, Michigan, New Hampshire, North Carolina, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Dakota, Tennessee, Texas, Washington, West Virginia, and Wyoming.

In Washington state the hours worked in manufacturing were treated as missing data for January 1980 because a severe storm in the week of the survey reduced the hours worked more than 5%. The entries for October and November 1989 were also treated as missing data because a strike at Boeing reduced manufacturing hours worked more than 14% from the months immediately preceding. Because of the discontinuity in the hours data with the introduction of the North American Industrial Classification system (NAICS) in January 2001, the hours data were coded as missing in that month for all states.

2. Nonfarm Payroll Employment

In addition to the normal seasonal adjustment, in the following states the irregular components that were between 1.5 and 2.5 standard deviations from the mean were given a weight between 1 and 0. Those more than 2.5 standard deviations from the mean were given a weight of 0: Hawaii, Michigan, Nebraska, Nevada, Oregon, Rhode Island, Virginia, and West Virginia. In the following states the irregular components that were between 1 standard deviation and 2 standard deviations from the mean were given a weight between 1 and 0. Those more than 2 standard deviations from the mean were given a weight of 0: Delaware, Iowa, Kentucky, Minnesota, Montana, North Dakota, Ohio, South Dakota, Utah, and Washington.

3. Unemployment Rate

For the months indicated in the following states, the unemployment rate was treated as missing data because of an anomalous change of more than 1.5 percentage points in a single month:

Delaware (September 1990)
Georgia (March 1991, January 1992)

Illinois (November 1993)
Michigan (April 1980, December 1981, February 1991)
New Jersey (May 1992)
North Carolina (October 1981)
Oklahoma (June 1980, December 1980)
Rhode Island (December 1989)

4. Wage and Salary Disbursements

Wage and salary disbursements were treated as missing data in the first quarter of 1993 for the following states because the shift of bonuses and other compensation into 1992 in anticipation of a tax increase resulted in a decline of more than 4% in real wages and salaries for the quarter: California, Connecticut, Delaware, Illinois, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, and Washington.

Because of the shift of compensation to 1992, wage and salary disbursements in New York increased more than 4% in the fourth quarter of 1992, so data in that quarter were also treated as missing for New York state.

Strikes by coal miners reduced real wage and salary disbursements in West Virginia more than 10% in the first quarter of 1978 and the second quarter of 1981, so wage and salary data for those quarters were treated as missing data.