Seasonal adjustment in the ECI and the conversion to NAICS and SOC

As part of the conversion of ECI industry and occupation classifications to NAICS and SOC, the Bureau of Labor Statistics used the converted classifications to estimate the seasonally adjusted ECI; in addition, the Bureau improved the methodology and processing of seasonally adjusted estimates

E. Raphael Branch
and
Lowell Mason

E. Raphael Branch is a labor economist in the Division of Compensation Data and Estimation, Office of Compensation and Working Conditions, Bureau of Labor Statistics. Lowell Mason is an economist in the Program Standards Group in the same office. E-mail: branch.raphael@bls.gov

Over the course of a year, rates of change in the cost of wages and benefits, as measured in the Employment Cost Index (ECI), reflect events that follow a more or less regular pattern. These events include expansions and contractions of economic activity that occur in specific periods of the year, such as increased work in the construction industry during warm weather and heightened activity associated with the beginning of the school year in the education industry. Such regular patterns in an economic time series are typically referred to as *seasonal effects*. The process of estimating and removing these effects from an economic series is called *seasonal adjustment*. Seasonal adjustment makes it easier for analysts to observe the longrun and other movements in an economic time series, exclusive of seasonal effects. Economists and other researchers are particularly interested in observing cyclical and longrun movements of economic series to better understand the economic behavior of various sectors of the economy.

The ECI is a time series—a quarterly fixed-weight index of changes in the cost of employment compensation—published since 1975.[1] The ECI includes index and percent-change estimates for employer costs per hour worked, including the cost of total compensation, wages and salaries, and benefits. As a time series with repeated quarterly measurements, the ECI has been analyzed for seasonal adjustment, and seasonally adjusted ECI estimates are available from December 1990 onward, in addition to estimates that are not seasonally adjusted.

Responding to a mandate from the Office of Management and Budget (OMB) of the Executive Office of the President, the Bureau of Labor Statistics (BLS, the Bureau) converted the ECI industry and occupation classifications from the Standard Industrial Classification (SIC) system and the Occupational Classification System (OCS) to the North American Industry Classification System (NAICS) and the Standard Occupational Classification (SOC) system.[2] The first NAICS- and SOC-based ECI estimates were released for the March 2006 reference period. At the same time, seasonally adjusted data were issued on the basis of NAICS and SOC.

The OMB mandate necessitated three changes to the ECI seasonal adjustment process. First, it was necessary to construct 10-year time series based on the new classification systems. (Ten years is the span of data employed for seasonal adjustment of the ECI.) The construction of these historical series in turn required the reclassification of industries and occupations not previously coded with NAICS and SOC. Second, the Bureau made several methodological enhancements, including updating fixed weights used in the ECI.[3] The change in the weights made it necessary to revise the weights used in the seasonal adjustment calculation. Last, the Bureau improved both the seasonal adjustment methodology and the processing used to derive seasonally adjusted estimates. This article gives an overview of the ECI seasonal adjustment process and discusses these changes and improvements to the adjustment of the ECI.

Seasonal adjustment of the ECI

Seasonal adjustment of the ECI is accomplished both directly and indirectly. In direct seasonal adjustment, an index that has not been seasonally adjusted is divided by the seasonal factor for the series. The seasonal factors for directly adjusted series are estimated by applying seasonal adjustment techniques directly to historical data for the series. In indirect seasonal adjustment, indexes that have been directly seasonally adjusted are weighted and then summed.[4] In effect, indirect seasonal adjustment yields a weighted average index in which the weights sum to unity.[5] The choice of direct or indirect seasonal adjustment is related to the level of aggregation of the data. The methodology used for the ECI follows standard BLS practice, which is based on practices that are used widely across U.S. Federal statistical programs.

Direct seasonal adjustment. At the lowest level, ECI series are seasonally adjusted directly, according to the formula

$$\hat{A}_{it}^{\text{direct}} = \frac{O_{it}}{\hat{S}_{it}},$$

where $\hat{A}_{it}^{\text{direct}}$ is the estimated directly seasonally adjusted index for ECI series i at time t; O_{it} is the original, not seasonally adjusted index for ECI series i at time t; and $\hat{\text{SI}}_{it}$ is the estimated seasonal factor for ECI series i at time t.

1. *Deriving seasonal factors.* Seasonal factors for the direct seasonal adjustment of an ECI series are estimated with the X–12–ARIMA program developed by the Time Series Staff of the Research Division of the Census Bureau of the U.S. Department of Commerce.[6] The estimated seasonal adjustment factors for the most recent year of available data are used as projected seasonal factors for the coming year, under the assumption of static expectations. (In other words, the estimated seasonality for the last year of available data is assumed to be the same in the next year.) Using projected factors is necessary because data for the current year are not available. The *data span* used in ECI seasonal adjustment (that is, the range of data used in estimating seasonal factors) is the most recent 10 years; thus, for 2006, the data span is 1996 through 2005.

An integral part of the X–12–ARIMA program is the X–11 seasonal adjustment method, used to derive the seasonal factors.[7] X–11 is a seasonal adjustment method that uses moving averages, also known as filters. This method assumes that a time series can be decomposed into three components: trend, seasonal, and irregular. These components then can be combined in an additive or a multiplicative way. For the ECI, as with most economic series, the multiplicative model is used. That is, the original series is specified as the product of the trend, seasonal, and irregular components:

$$O_{it} = T_{it} \times S_{it} \times I_{it}.$$

Here, O_{it} and S_{it} are as before, T_{it} is the trend for series i at time t, and I_{it} is the irregular component of series i at time t. A direct seasonally adjusted series is obtained by dividing the original series by the estimated seasonal factors, or

$$\hat{A}_{it} = \frac{O_{it}}{\hat{S}_{it}} \approx T_{it} \times I_{it},$$

where $\hat{A}_{it}$ is the estimated seasonally adjusted time series i at time t and O_{it}, S_{it}, T_{it}, and I_{it} are as before.

A. Trend component. The trend (or, more precisely, the trend cycle, which includes both secular and cyclical changes) is the long-term movement in the series, without the seasonal and irregular effects. The trend represents the underlying index levels of the series. The ECI is primarily a measure of trends in compensation costs. The trends in the ECI are dominant, and all ECI series have a long-term positive trend that is the result of various economic influences, such as (1) employers granting wage increases to offset inflation and (2) the rising costs of employer-provided health insurance benefits. The ECI is seasonally adjusted to provide estimates of long-term movements in compensation costs for a better understanding and analysis of the U.S. economy.

B. Seasonal component. The seasonal component of a time series consists of effects that follow a more or less regular pattern each year and arises from systematic, calendar-related influences, including the following:

- Natural conditions, such as weather fluctuations that are representative of a season.
- Business and administrative practices, such as changes in health plan benefits and rates, which usually occur at the beginning of the calendar year.
- Social and cultural events, such as Thankgiving and Christmas, which often result in in the hiring of part-time and temporary labor in the retail trade industry.

Construction industry data from the ECI offer a clear example of seasonality in an economic time series. The demand for labor in the construction industry increases during the spring and summer, when the weather is temperate, and decreases in the fall and winter, when the weather is less suitable for outdoor work. Wage and benefit costs in the industry increase and decrease as the demand for labor shifts. Construction wage increases are lowest in December, gradually increase beginning in March as the weather improves, and finally reach a peak in June, when the weather is most conducive to construction.

Table 1 lists the estimated seasonal factors for the published ECI wage series for 2005—factors that also are used as projected seasonal factors for 2006. Despite the small size of these factors, they have an impact on the index. For instance, when the not seasonally adjusted index for wages in the nondurable-goods-manufacturing industry was 100.0 in the December 2005 base period, the seasonally adjusted index was 100.3.

C. Irregular component. The residual component of a time series, called the irregular component, is what is left of the series after the trend and seasonal components have been removed. The residual component represents the short-term fluctuations in the series that are neither systematic nor predictable. Abnormal weather conditions (such as the substantial hurricanes of 2005) and increases in gas prices are examples of irregular effects. Both sampling and nonsampling error also can cause irregular fluctuations in a time series.

2. Identifying seasonal series. Once the seasonal-factor estimation of all ECI series is completed, the results are evaluated to determine those series recommended for seasonal adjustment. This evaluation is necessary because not all ECI series exhibit pronounced seasonal effects. For example, under NAICS and SOC, no seasonality was found in wages and salaries, or in the cost of benefits, for service occupations; consequently, these series (as well as the total-compensation series for service occupations) are not seasonally adjusted in the March 2006 data.

Three quality control statistics are used as guidelines in deciding whether to seasonally adjust an ECI series. These three statistics are the F statistic for stable seasonality (F_s), the M7 statistic, and the Q statistic. To determine whether a series exhibits seasonality, the analyst uses the F_s statistic. F_s is a one-way analysis-of-variance test that measures the degree of stability of the seasonal component of a time series. In particular, F_s summarizes quarterly differences among means of the seasonal-irregular (SI) ratios. The null hypothesis of the test is that all four quarterly seasonal means are equal. If the means are *not* equal, as measured by F_s exceeding a critical value of 7.0, then the series is considered seasonal and will be seasonally adjusted if the M7 and Q statistics are acceptable.[8]

Even if a series is seasonal, as indicated by the F_s test, it may not be possible to obtain a statistically efficient estimate of its seasonal component. That is, the series' seasonality, though present, may not be *statistically identifiable*. To determine whether it is in fact so identifiable, analysts use the M7 and Q statistics. M7 compares the F statistics for moving seasonality with the F statistic for stable seasonality. *Moving seasonality* refers to whether seasonal movements change over time and is measured with a two-way analysis-of-variance test. If there is too much moving seasonality, the estimate of the series' seasonal component could be erroneous.

The Q statistic is a weighted average of 11 M statistics (including M7) that test for different types of problems that may affect the overall quality of the seasonal adjustment, such as large changes, large variances, and the absence of randomness in the irregular component.[9] The M statistics are normalized to 1.0. Values less than 1.0 for any M statistics indicate that certain properties of the seasonal and irregular components are acceptable. A value of M7 less than 1.0 and Q less than 1.0 together indicate that a series' seasonality is identifiable. A value of M7 greater than 1.0 indicates that the series has too much moving seasonality to estimate the seasonal component accurately. A value of Q greater than 1.0 indicates a variety of possible difficulties, including large changes, large variances, and the absence of randomness in the irregular component or too much change in the seasonal component. In either case, the seasonality of the series cannot be measured.

Table 1. Seasonal factors for published direct seasonally adjusted wage series, Employment Cost Index, private industry, 2005

Series	March	June	September	December
Construction	99.83343	99.98706	100.28369	99.90260
Durable-goods manufacturing	99.94400	100.06148	100.11964	99.87821
Nondurable-goods manufacturing	100.13595	100.16893	100.02010	99.68659
Retail trade	99.92927	100.12733	100.07867	99.87055

NOTE: Although durable- and nondurable-goods-manufacturing series are scheduled to be discontinued in March 2007, they are presented here as representative of the March 2006 estimates.

If an ECI series has both stable and identifiable seasonality on the basis of the F_s, M7, and Q tests, the series is seasonally adjusted.[10] Otherwise, the series will not be seasonally adjusted. However, due to the volatility of some ECI series and the small size of seasonal factors in the ECI, the quality control statistics for a series may be inconsistent from year to year. In anticipation that the quality control statistics might swing in favor of seasonal adjustment the next year, a series is not immediately removed from seasonal adjustment the first time it fails the quality control tests. Instead, the series is no longer seasonally adjusted if seasonality is not found for 3 consecutive years.

Indirect seasonal adjustment. After seasonally adjusting lower level series by the direct method, the analyst seasonally adjusts aggregated series indirectly. Indirect seasonal adjustment is expressed by the formula

$$\hat{A}_{it}^{\text{indirect}} = \sum_{i}^{n} \left(W_{it} \hat{A}_{it} \right),$$

where $\hat{A}_{it}^{\text{indirect}}$ is the estimated indirect seasonally adjusted ECI aggregate series j at quarterly period t, W_{it} is the weight for an ECI indirect series component i at quarterly period t,[11] $\hat{A}_{it}$ is the directly estimated seasonally adjusted component series i at quarterly period t, and n is the number of component series i in ECI aggregate series j.

For example, given seasonally adjusted indexes and aggregation weights for durable- and nondurable-goods industry wages and salaries in December 2005, the two components of manufacturing wages and salaries are as shown in the following tabulation:

Component series	*Direct seasonally adjusted index*	*Cost-employment weight*
Durable goods	100.1	0.657619
Nondurable goods	100.3	.342381

The estimated indirect seasonally adjusted index for manufacturing wages and salaries is calculated as

$$\begin{aligned}\hat{A}_{t}^{\text{indirect}} &= \sum_{s} \left(W_{st} \hat{A}_{st} \right) \\ &= (.657619)(100.1) + (.342381)(100.3) \\ &\approx 100.2.\end{aligned}$$

For indirect seasonal adjustment, the directly adjusted component series are weighted because they often contribute unequally to the aggregate. In the preceding example, the durables industry wage bill[12] accounts for about 66 percent of the manufacturing wage bill and the nondurable-goods industry wage bill accounts for the remaining 34 percent.

The indirect method is preferred for aggregated series such as compensation (which is an aggregation of wages and benefits), because it reduces the chances of inconsistencies between aggregate and component series, either in the seasonally adjusted indexes or in the percent changes based on them. In addition, if component series have very different seasonal patterns, indirect seasonal adjustment is preferable, primarily because, when the data are combined, as would occur in direct seasonal adjustment, the seasonal patterns may "wash out," making them more difficult to estimate.[13] However, even when component series have the same pattern, estimating an aggregate series by the direct method can lead to anomalies, particularly in the case of the ECI, in which the seasonal factors are small and indexes may fluctuate, due in part to rounding. For example, it is possible for a directly adjusted aggregate series to move in a direction opposite that in which its components are moving.

1. Three-month percent changes. Three-month percent changes are calculated from the published seasonally adjusted indexes. Let R_t represent the rate of change in any seasonally adjusted compensation, wage, or benefit series. Then the 3-month percent change is calculated as

$$R_t = \left(\frac{\hat{A}_t - \hat{A}_{t-1}}{A_{t-1}} \right) \times 100,$$

where $\hat{A}_t$ is an estimated seasonally adjusted ECI for quarterly period t and $\hat{A}_{t-1}$ is an estimated seasonally adjusted ECI for the preceding quarterly period $t - 1$.

2. Revisions. In addition to initial seasonal adjustment of the (appropriate) ECI estimates, revisions to seasonally adjusted historical indexes and to related historical 3-month percent changes are conducted annually after the December quarterly production cycle is completed. The revisions are necessary because the averaging methods used in deriving seasonal factors yield imprecise initial estimates of the seasonal component for the year of interest.[14] Accordingly, the estimates are incrementally corrected with subsequent estimates that make use of more recent data. The BLS practice is to make five revisions before the seasonal-factor estimates become final. Therefore, as part of the seasonal adjustment methodology, the Bureau annually revises the seasonally adjusted indexes and 3-month percent-change estimates of the most recent 5 years.[15] (See the appendix for a discussion of why revisions are necessary.)

NAICS and SOC conversion

Methodology changes. Under NAICS and SOC, the basic direct and indirect seasonal adjustment methodology is the same as in the past. Differences in direct seasonal adjustment between the two systems are mainly in the classification of series and the precision of data inputs, the combination of which allows for more accurate seasonal-factor estimates. With the introduction of the X–12–ARIMA program[16] for the ECI in 1999, many more series could be seasonally adjusted than in the past. As a result, many more seasonally adjusted series became available for use as components in indirect seasonal adjustment than were initially available when seasonal adjustment was introduced into the ECI in 1990.

Taking advantage of the increased number of direct seasonally adjusted series, analysts now make greater use of unpublished data in constructing indirect adjusted series. Almost all detailed ECI series are potential components in the indirect seasonal adjustment of a more aggregate series. This fact offers a distinct advantage over the past, when substantially fewer ECI series could be indirectly seasonally adjusted because of processing limitations and the need to wait for new production systems to become operational.

Under NAICS and SOC, all compensation series are seasonally adjusted indirectly, as a practical consideration in order to eliminate potential inconsistencies.[17] For instance, as of December 2005, the Bureau had published 38 seasonally adjusted series classified by SIC and OCS, 27 of them directly seasonally adjusted and 11 indirectly adjusted. By comparison, the March 2006 tables show 50 seasonally adjusted NAICS and SOC series, 4 directly and 46 indirectly. For example, under SIC–OCS, the total-compensation series for construction, wholesale trade,[18] and retail trade were seasonally adjusted by the direct method, primarily because the benefit series for those industries were not published. Under NAICS and SOC, however, these compensation series are seasonally adjusted by the indirect method, using their wages and salaries and benefits components. Exhibit 1 shows which series in the March 2006 data release are adjusted by the direct or indirect method.

In addition to absorbing changes related to the conversion to NAICS and SOC, as well as methodological improvements available with the use of X–12–ARIMA processing, the new ECI quarterly production processing system improves the Bureau's flexibility to drop and add seasonally adjusted series when necessary. The new ECI quarterly production system overcame limitations in the earlier system that prevented the analyst from implementing selected changes in seasonal adjustment.

Data

The ECI seasonal adjustment methodology calls for a 10-year data span for X–12–ARIMA estimation. (See the appendix for a discussion of the size of the data span.) For the December 2005 base-period ECI and the March 2006 estimates, NAICS and SOC series were required for each quarter from March 1996 through December 2005. ECI-sampled establishments and occupations were classified by NAICS and SOC beginning in March 2000. Because NAICS and SOC codes were not available before 2000, these codes were assigned by a variety of methods. If ECI-sampled establishments and occupations for 1996 through 1999 also were in the March 2000 ECI sample (when data collectors first assigned NAICS and SOC codes), then the codes assigned also applied to the 1996–99 sample data. This procedure accounted for roughly 44 percent of the NAICS and SOC classifications of the observations for those years. For establishments and occupations that were not part of the ECI sample as of March 2000, NAICS and SOC classifications were determined by several other methods. Forty percent of the NAICS classifications were determined on the basis of State unemployment filings, and 22 percent of the missing SOC classifications were determined by matching similar occupational classifications in the OCS. The remaining NAICS and SOC classifications were imputed with the use of a "nearest-neighbor" approach similar to the approach used to estimate other missing values.[19]

Publication plans

ECI data are published in quarterly news releases and historical listings of total compensation, wages and salaries, and costs of benefits, each by economic sector (civilian, private, or government), industry, and occupation. As of March 2006, the tables published in the news release and those in the historical listings no longer show estimates for series that are not seasonally adjusted. The Bureau anticipates publishing additional series beginning in March 2007.

As in the past, the news release tables of seasonally adjusted data contain indexes for the current quarter and the previous quarter, as well as historical 3-month percent changes for the most recent eight quarters. However, historical NAICS and SOC 3-month percent changes calculated prior to the official estimates of March 2006 were included in the news release to provide a historical context for the NAICS and SOC seasonally adjusted ECI series. These historical data were included for information purposes only; the official estimates for periods prior to March 2006 are SIC and OCS estimates, which reside among the SIC and OCS archived data. The SIC and OCS estimates have been rebased, so percent changes may differ from the original estimates due to rounding. NAICS- and SOC-based seasonal factors are published separately on the BLS Web site.

Revisions of historical seasonally adjusted data for the most recent 5 years also were published on the BLS Web site prior to publication of the news release and appear in the regular historical listing as well. As part of the conversion to

Exhibit 1. Method of seasonally adjusting series in the Employment Cost Index, March 2006

Economic sector, industry, and occupational group	Total compensation[1]	Wages and salaries	Cost of benefits
All civilian workers			
All workers	Indirect	Indirect	Indirect
Private-industry workers			
All workers	Indirect	Indirect	Indirect
By occupation:			
White-collar occupations	Indirect	Indirect	Indirect
Blue-collar occupations	Indirect	Indirect	Indirect
Service occupations	([2])	([2])	([2])
By industry:			
Goods-producing industries	Indirect	Indirect	Indirect
Construction	Indirect	Direct	([3])
Manufacturing	Indirect	Indirect	Indirect
Durable goods	Indirect	Direct	([3])
Nondurable goods	Indirect	Direct	([3])
Service-providing industries	Indirect	Indirect	Indirect
Trade, transportation, and utilities	Indirect	Indirect	([3])
Wholesale trade	Indirect	([2])	([3])
Retail trade	Indirect	Direct	([3])
Information	Indirect	Indirect	([3])
Financial activities	Indirect	Indirect	([3])
Professional and business services	Indirect	Indirect	([3])
Education and health services	Indirect	Indirect	([3])
Educational services	Indirect	Indirect	([3])
Healthcare and social assistance	Indirect	Indirect	([3])
Leisure and hospitality	Indirect	Indirect	([3])
Other services, except public administration	Indirect	([2])	([3])
State and local government workers			
All workers	Indirect	Indirect	Indirect

[1] Includes wages, salaries, and employer costs for employee benefits.

[2] No seasonality was found in the series.

[3] Series is not published.

NAICS and SOC, the first available historical listing contains seasonally adjusted data from March 2001 through March 2006. For information purposes, the 10-year historical NAICS and SOC ECI (not seasonally adjusted) series that were used in estimating seasonal factors are available on the BLS Web site.

The SIC–OCS seasonally adjusted ECI series underwent their last seasonal adjustment revision, and the results were published in the usual historical listings and archived on the BLS Web site in April 2006. Seasonal factors for SIC- and OCS-based series were last published in March 2005. Thus, for the March 2006 data release, there are two historical listings with seasonally adjusted (as well as not seasonally adjusted) data: a listing of archived SIC and OCS data for the period through December 2005 and a listing that is the first of ongoing historical listings by NAICS and SOC classification.[20] □

Notes

ACKNOWLEDGMENTS: Thanks go to James A. Buszuwski, Thomas G. Moehrle, Stuart Scott, and Claire Gallagher of the Bureau of Labor Statistics and Kathleen M. McDonald-Johnson of the U.S. Census Bureau for providing insightful comments on previous drafts of this paper.

1 Selected ECI series for private-industry wages and salaries were first published for September–December 1975. Selected private-industry total compensation and benefit series began with December 1979 data. Government series were introduced in June 1981.

2 The SIC was prepared by the Technical Committee on Industrial Classification of the Office of Management and Budget; see *Standard Industrial Classification Manual 1987* (Executive Office of the President, Office of Management and Budget, 1987). The OCS was developed for use in the 1990 decennial census; see *Occupational Classification System Manual*, on the Internet at **www.bls.gov/ncs/ocs/ocsm/commain.htm**. For NAICS, see *North American Industry Classification System, United States 2002* (Executive Office of the President, Office of Management and Budget, 2002). For the SIC system, see *Standard Occupational Classification Manual 2000* (Executive office of the President, Office of Management and Budget, 2000).

3 The Bureau introduced 2002 fixed employment weights to replace the 1990 weights used from 1995 through 2005. The Bureau also updated its methods of imputation. The accompanying articles in this issue discuss these changes in more detail. (See Stephanie Costo, "Introducing 2002 weights for the Employment Cost Index," which reviews the reweighting of the ECI and discusses its effects; and Song Yi, "Accounting for missing data in the Employment Cost Index," which discusses imputation procedures for estimating missing values.)

4 If a component of an indirectly seasonally adjusted series is not seasonal, it will enter the indirect seasonal adjustment calculation as a not seasonally adjusted series.

5 The weighted sum of directly adjusted indexes can be viewed as a weighted sum of ratios of indexes that are not seasonally adjusted, divided by their respective seasonal factors.

6 When the Bureau of Labor Statistics first began seasonally adjusting ECI data with its December 1990 estimates, the staff of the BLS ECI Data Estimation Branch used X–11–ARIMA/88, developed by Estela Dagum of Statistics Canada. X–11–ARIMA/88 offered ARIMA forecasting and diagnostic tools for evaluating estimated seasonal factors for the ECI. During the early development of ECI seasonal adjustment, BLS staff visited Statistics Canada, and that agency's staff participated in the development of methodology and in the evaluation of the ECI series for seasonal adjustment. The X–11–ARIMA/88 seasonal adjustment estimation program contains the X–11 seasonal adjustment program, developed by the U.S. Census Bureau (see Julius Shiskin, Allan H. Young, and John C. Musgrave, "The X–11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper No. 15 (U.S. Census Bureau, 1967)), plus enhancements to the original X–11–ARIMA program developed at Statistics Canada in 1980. X–12–ARIMA offers additional features to improve the forecasting and diagnostics and is the standard seasonal adjustment package used for BLS programs. All of these programs improved the quality of ECI estimates.

7 X–11 is a seasonal adjustment method that uses moving averages to smooth an economic series. The addition of ARIMA modeling offers better forecasting of the seasonal component and helps improve the quality of the smoothing. In 1999, the Bureau first published seasonal-factor estimates based on the use of X–12–ARIMA, which includes further enhancements to the X–11 methodology and additional diagnostics, as well as several autoregressive estimation techniques for improving the quality of the seasonal-factor estimates. X–11 is the most widely used technique for seasonal adjustment, and X–12–ARIMA is used extensively in U.S. Federal statistical programs.

8 In practice, some assumptions of the standard F-test just described are violated; for example, the irregular component may be autocorrelated with the seasonal component. For this reason, a high critical value of F_s is used to analyze stable seasonality. In particular, an F_s greater than 7.0 suggests stable seasonality (that is, the null hypothesis is rejected and the series is considered seasonal), and consequently, the series will be seasonally adjusted. By contrast, an F_s less than 7.0 would lead to acceptance of the null hypothesis, indicating that the variation in the series, measured across quarterly means—and therefore the series itself—may not be seasonal or that seasonality cannot be adequately measured. In this case, the series will not be seasonally adjusted. (For a detailed explanation of the F statistic for stable seasonality, see Dominique Ladiray and Benoit Quenneville, *Seasonal Adjustment with the X–11 Method* (New York, Springer-Verlag, 2001), pp. 57–58, 135–36.)

9 The purpose of the M-statistics is primarily to evaluate estimates of the irregular and seasonal components. To measure the seasonal component of a series accurately, it is desirable that the irregular component of the series be statistically random and neither too large nor too small relative to the remaining components and the series as a whole. It is also desirable that the seasonal component be stable over the data span. In general, six of the M-statistics measure changes, variances, and the size of the irregular component relative to the entire series, as well as the trend-cycle and seasonal components. Excluding M7, the remaining M-statistics measure the size and movements in the seasonal component over the entire series and for the most recent 3 years. For details, see Ladiray and Quenneville, *Seasonal Adjustment with the X–11 Method.*

10 This procedure is subject to at least one caveat: an ECI series with stable and identifiable seasonality might not be seasonally adjusted if there is a consistency or production system consideration that would preclude following the guideline. Just such a consideration did in fact arise in selected instances in the SIC–OCS seasonally adjusted series. With the introduction of new seasonal adjustment methods and a new production system, this kind of problem has been essentially eliminated under NAICS and SOC.

11 For more details about the ECI base-period cost weights, see Costo, "Introducing 2002 weights."

[12] The wage bill is the average wage of workers in an occupation within an industry, multiplied by the number of workers represented by that occupation.

[13] For a detailed discussion of aggregation in seasonal adjustment, see Stuart Scott and Peter Zadrozny, "Aggregation and Model-based Methods in Seasonal Adjustment of Labor Force Series," *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* (Alexandria, VA, American Statistical Association, 1999), pp. 156–61.

[14] For a further discussion about seasonal adjustment revisions, see James A. Buszuwski, "Alternative ARIMA Forecasting Horizons when Seasonally Adjusting Producer Price Index Data with X–11–ARIMA in Concurrent Mode," *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* (Alexandria, VA, American Statistical Association, 1987), pp. 488–93.

[15] Seasonally adjusted percent changes are reviewed for rounding anomalies. Percent changes are based on rounded indexes so that the percent changes can be replicated. In calculating percent changes, indexes are rounded to one decimal place, a practice that sometimes causes percent changes in compensation indexes to fall outside the range of their wage and benefit components. When this happens, the percent change is referred to as "unbounded." Unbounded percent changes are a potential problem in any calculation in which rounded indexes are used. The problem may be more pronounced in the seasonal adjustment percent changes because rounding occurs more than once in the calculation.

[16] For the ECI, the batch version of X–12–ARIMA is used, allowing multiple series to be processed in a single computer job.

[17] Quality control diagnostics also can be used to decide which method is best by comparing improvements in the quality control statistics of one method over the other. However, this course of action was not undertaken due to time constraints and requirements needed for the new production system to begin operation. In addition, past experience with the two estimation methods often provided conflicting results, due largely to rounding.

[18] The wholesale trade compensation series is seasonally adjusted because the wholesale trade benefit cost series (which is unpublished) is seasonal, whereas the wholesale trade series is not.

[19] "Nearest neighbor" is one of the procedures used to estimate or impute a missing value. Under that approach, imputation classes (cells) are formed on the basis of information that is known about all units, such as economic sector (that is, private industry, or State or local government), industry, major occupational group, collective bargaining status, region, and full- or part-time status. Within each cell, a unit that is missing the characteristic of interest (a unit that is unusable and that therefore will be a recipient of an imputed value) takes the value of the characteristic of a unit that is "usable" (a donor of an imputed value) and that is "nearest" to the recipient. "Nearest" is defined by the similar auxiliary data the donor and recipient may share. There may, of course, be many donors from the same cell. If so, the specific tie breaker—determining which donor the recipient will use—is based on the establishment size. The donor with the establishment size closest to that of the recipient will be chosen for imputation. In cases where donors do not have all of the same characteristics, the nearest-neighbor procedure drops one of the matching variables and tests whether a full match now can be obtained. If no full match can be obtained even after one variable is dropped, another matching variable will be dropped until a full match can be found.

[20] A detailed discussion of time-series analysis and seasonal adjustment can be found on the U.S. Census Bureau Web site, **www.census.gov/srd/www/x12a/**. Additional information is available on the Australian Bureau of Statistics Web site, **www.abs.gov.au/Ausstats**.

APPENDIX: X–12–ARIMA and its application to the ECI

X–12–ARIMA was developed by the Time Series Staff of the U.S. Census Bureau and is used to estimate seasonal *autoregressive integrated moving-average* (ARIMA) models to help produce good estimates of seasonal factors and other components of a time series. X–12–ARIMA incorporates the features of its predecessors, developed at the Census Bureau and Statistics Canada.[1] Among the many features of X–12–ARIMA are its capability to provide for (1) user-supplied extreme-value adjustments (also called outlier adjustments or prior adjustments), (2) forecasts of the series, (3) automatic choices of ARIMA models, and (4) sliding-spans analysis. For the estimation of seasonal factors in the ECI, all four of these features, as well as a logarithmic transformation of the data, are used.

Prior adjustment factors are applied to the original indexes, and the series with the prior adjustments included is used to derive forecasts.[2] In general, prior adjustments limit the effects of extreme short-term economic influences and other factors. Applying prior adjustment factors to the original ECI series helps improve estimation of the series for subsequent steps of the seasonal adjustment process and, ultimately, improves the quality of the seasonally adjusted series.

For purposes of estimation, the ECI series (with prior adjustments) is forecasted 1 full year (four quarters) beyond the last actual data. Forecasting the original series improves its subsequent X–11 decomposition. This technique makes it possible for X–11 to estimate moving averages that cannot be obtained from the original data alone. Using these forecasts and revising the estimates for 5 years makes it possible to provide the best estimates of seasonal factors available with the X–11 method.[3]

The *automodel* option offers a choice of five ARIMA models that often are used in forecasting an economic series. For the ECI, the first model that meets the criteria for acceptance is selected from among the five models offered.[4]

Sliding-spans analysis was used to determine the size of the data span—the length of the time series used to derive seasonal component estimates. This feature provides for seasonal factor estimation of subsets of a data span, including overlapping periods. It thereby becomes possible to compare seasonal adjustment diagnostics for multiple adjacent data spans.[5]

X–11 decomposition

An X–11 program within X–12–ARIMA processes the estimated time series with forecasts, decomposing the series into trend, seasonal, and irregular components. The decomposition involves three basic steps,[6] which are executed many times to obtain final estimates of the seasonal factor. These steps, which highlight how X–11 estimates the seasonal factors under the assumptions outlined earlier (see text, page 13), are as follows:

1. Estimate the trend $\hat{T}_{it}$ from the original series O_{it}.
2. Estimate the seasonal irregular component, SI, from the original series and the trend component $\hat{SI}_{it} = O_{it} \div \hat{T}_{it}$.
3. Estimate the seasonal component $\hat{S}_{it}$ by applying a moving average to the initial seasonal irregular (SI) component:

$$\hat{S}_{it} = \mathrm{SI} \xrightarrow{\mathrm{MA}} \hat{S}_{it}$$

Then calculate the seasonally adjusted series: $\hat{A}_{it} = O_{it} \div \hat{S}_{it}$.

The decomposition to final component estimates is iterative. Moving-average procedures are repeated many times to obtain successively more precise estimates of each component. Moving averages, called *filters*, successively operate on a shifting time span of data. That is, moving averages filter out certain cycles in the original series. The trend is removed from the original series by filtering out longer, nonseasonal cycles. The seasonal component is removed by filtering out shorter, irregular cycles. Historically, this procedure has been found to isolate the seasonal effects well.

Graphic analysis

The analysis of seasonal adjustment results for the ECI includes a review of time-series and spectrum graphs. Overlay graphs of the original and seasonally adjusted series are reviewed. Because the ECI is predominantly a measure of a trend, the differences between the seasonally adjusted series and the original, not seasonally adjusted series are small.

To assist in evaluating the quality of seasonal adjustment, spectrum graphs of the original and seasonally adjusted series are compared.[7] Seasonal peaks in the spectrum graphs for quarterly data are measured at frequencies of 0.25 and 0.5 on the *x*-axis. Graphs of the seasonal factors by quarter also are reviewed. Seasonal factors for all periods in the data span are plotted around an axis at their mean.

Also, the spectrum graph of the irregular component is used to observe whether any seasonality remains in the seasonally adjusted ECI series. This kind of seasonality, called *residual seasonality*, can result from limitations in the seasonal adjustment procedure or from difficult-to-estimate seasonal effects in the series. The absence of a peak at frequencies of 0.25 and 0.5 is an indication that a series has no residual seasonality. If a peak were present at a frequency of 0.25 or 0.5 in the irregular spectrum, and if diagnostic reports indicated the presence of residual seasonality, the series would be investigated further. Whenever there is evidence of residual seasonality, the seasonally adjusted ECI series is tested by running it through X–12–ARIMA to see if any significant seasonality is found, as indicated by the F_s statistic. (If it is, then steps are taken to identify and correct any problem that may exist and to reestimate the seasonal factors.)

Notes to the appendix

[1] For details on the development of X–12–ARIMA, see David F. Findley, Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen, "New Capabilities and Methods of the X–12–ARIMA Seasonal Adjustment Program," *Journal of Business and Economic Statistics*, April 1998; on the Internet at **www.census.gov/srd/www/sapaper.html/**.)

[2] The prior adjustment factors are estimated outside of the X–12–ARIMA program and are introduced with the prior adjustment option provided in the program. Using estimates from an initial execution of X–12–ARIMA to obtain preliminary extreme-value adjustments when extreme-value weights range from zero to 20 percent, the analyst takes a ratio of the unadjusted seasonal-irregular component (table D8 in the X–12–ARIMA results), divides by the estimated replacement values for the extreme values (table D9 in the X–12-ARIMA results), and multiplies by 100. The resulting prior adjustment values are entered into a second X–12–ARIMA estimation in which the original series is prior adjusted.

[3] For more details on forecasting in seasonal adjustment estimation, see James A. Buszuwski, "Alternative Seasonal Adjustment Forecasting Horizons and Methods for the Producer Price Indexes," *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* (Alexandria, VA, American Statistical Association, 1986), pp. 373–78.

[4] The typical model chosen is the $(011\ 011)_4$ model, in the $(p\ d\ q)(P\ D\ Q)_s$ notation for the nonseasonal and seasonal operators, respectively, of a seasonal ARIMA model. For a detailed explanation of ARIMA models, see George E. P. Box and Gwilym M. Jenkins, *Time Series Analysis: Forecasting and Control*, rev. ed. (San Francisco, Holden-Day, 1976), chapter 9, pp. 300–33.

[5] From 1990 through 2001, the data span was determined by adding a year of data at each revision. By the time of the 2001 revision, the data span had reached 19 years. With the 2002 seasonal adjustment revision, a 10-year data span was introduced. Due to time and resource constraints, the size of the data span was not revisited for the NAICS and SOC conversion.

[6] The U.S. Census Bureau has an extensive set of documentation, including manuals and technical articles, as well as the X–12–ARIMA program, available as a free download on the Internet at **www.census.gov/srd/www/X-12a/**.

[7] The *spectrum*, or *spectral density*, graph measures relative contributions of frequencies to overall fluctuations in the series. The *x*-axis measures time, in cycles per quarter. Seasonal effects in quarterly data can be observed at frequencies of 0.25 and 0.5 cycle per quarter. The *y*-axis, or ordinate, is 10 times the logarithm of the spectrum amplitudes for the first difference of the series. (For details on the spectrum diagnostics in X–12–ARIMA, see Findley, Monsell, Bell, Otto, and Chen, "New Capabilities and Methods," pp. 127–77.)