Seasonality: economic data and model estimation

Seasonality is a pattern that more or less repeats itself each year, although this pattern may drift or change in amplitude over time; the study of seasonality also is linked to the study of business cycles

Ted Jaditz

Ted Jaditz is an economist in the Office of Prices and Living Conditions, Bureau of Labor Statistics.

What is seasonality? This is a surprisingly difficult question, for which there is no simple answer. At the most basic level, there is the intuition that seasonality is an approximately cyclical pattern in a time series that more or less repeats itself each year. Complicating matters is the possibility that the pattern will drift or change in amplitude from year to year.

The consensus among economists is that three basic exogenous factors give rise to seasonality in economic data.[1] The first factor is the weather: temperature, hours of daylight, and the likelihood of severe storms. All change somewhat predictably with the calendar and affect the costs of doing many types of business. Second, predictable and regular calendar events, such as Christmas, the Federal tax payment deadline on April 15, and the Independence Day holiday, affect production and consumption decisions. Third, social conventions have an impact on the timing of certain activities. For example, families with school-aged children time their vacations with the school calendar.

Businesses and consumers smooth over or heighten these exogenous factors as they plan their activities. For instance, a firm might time a shutdown for retooling to accommodate the vacation plans of employees. In response, the firm's suppliers and customers also might shut down at that time, causing what amounts to a seasonal slump in the industry.

In addition, changes in production techniques or preferences can accentuate or diminish seasonal patterns. For example, improvements in transportation between the Northeast and California might dampen seasonal patterns in produce prices in the Northeast.

Why bother with seasonality?

Historically, the study of seasonality has been tied to the study of business cycles. The business cycle is a pattern of boom and bust that is apparent in economic data over long periods, particularly in measures of output. A typical business cycle lasts about 48 months, although a cycle may extend for as little as 2 years or as long as 8 years. Predicting business cycles, or even determining where we are in a particular cycle, is important to business and government. As a result, many analysts use current data to make inferences about changes in the overall economy. The aim is to identify changes in the trend of economic activity from movements in certain indicators, such as data on prices or interest rates, or some other index of economic activity that is reported very frequently.

In this context, a seasonal pattern can complicate inferences about the business cycle. For example, industrial production drops significantly in the first quarter of the year, whether the economy is in an expansion or a recession. Analysts must judge whether a first-quarter dip is caused by seasonal factors that will disappear next quarter or whether the decline is a signal of a change in the business cycle from boom to bust. Decisions such as whether to hire additional workers and whether to invest in new plant and equipment will depend on a correct reading of the causes of economic changes.

In this arena, the traditional view of seasonality is that it is overlaid onto economic data and that it has no intrinsic interest, particularly when compared with the business cycle component:

> Since the seasonal pattern, by and large, is much the same in years of "good" business and years of "bad" business, our analysis of cyclical movements can be facilitated by putting the seasonal fluctuations provisionally out of sight. The effects on business enterprise of an increase in activity that is expected to last at most a few months are very different from an increase that is expected to continue for years.[2]

The seasonal pattern in economic data is "noise" in the "signal" that obscures and conceals the important features of the economy, which are the components of trends and cycles. This view of seasonality is at the very core of the approach taken by government statistical agencies toward seasonal adjustment of their data.

Implicit in such a view is a model of how economic data are generated. Arthur F. Burns and Wesley Clair Mitchell argue that an economic time series,

$$y_t = y_t^{TC} + y_t^S + e_t$$

can be broken down into y_t^S, the seasonal variable at time t, and y_t^{TC}, the trend in the time series and/or business cycle component.[3] (It is common also to include an irregular "noise" component e_t that is unrelated to the trend, cycle, or seasonal variable.) Implicit in Burns and Mitchell's view is the idea that the analyst is most interested in estimating trend-cycle relationships in the data.

Christopher A. Sims has analyzed the problem of trying to estimate relationships between time series under conditions of seasonality, using the Burns-Mitchell paradigm.[4] In the Sims model, one is interested in estimating a relationship between the unobservable trend-cycle components of two time series. Sims discusses a true regression model,

$$y_t^{TC} = x_t^{TC} b + u_t,$$

in which u_t is a regression disturbance term uncorrelated with x_t^{TC}. However, one does not observe y_t^{TC} or x_t^{TC}. Instead, one observes

$$y_t = y_t^{TC} + y_t^S + v_t$$

and

$$x_t = x_t^{TC} + x_t^S + w_t$$

where y_t^S and x_t^S are seasonal "noises," which are perhaps correlated with each other according to the equation

$$y_t^S = x_t^S c + z_t.$$

In these equations, v_t, w_t, and z_t are irregular "noises" not correlated with each other or with any other feature of the model. In this model, the trend-cycle component and the seasonal component are unrelated, are uncorrelated, and have different structural explanations.

Suppose we ignore the seasonal "noise" and estimate the model

$$y_t = x_t b' + e_t.$$

One might ask, What is the relationship between the estimated parameter b' and underlying model parameter b? In general, the two will not be the same: the parameter estimate that one obtains by regressing the observed y_t on x_t is not going to be equal to the estimated parameter that one would obtain if one regressed y_t^{TC} on x_t^{TC}. One can further show that, in general, the parameter b' is a weighted average of two parameters, one of which is the model parameter b and the other of which is the relationship with the seasonal components, c. Thus, b' is a biased estimate of b.[5] If the seasonal components of the two time series are unrelated, the net effect will be that the estimated parameter will be attenuated. On the other hand, it is possible—indeed, likely—that the seasonal components will be correlated across the series at hand. In this case, the bias can be quite large.

If we ignore seasonality, our estimates of the model parameters may be polluted by seasonal "noise." Thus, predictions and inferences may be incorrect. One might then ask whether it is possible to filter out the seasonal variable to get better estimates of the true model parameters. A common approach is to deal with seasonality by applying the X-11 filter. (For an examination of seasonality as applied to the BLS Consumer Expenditure Survey using X-11 methodology, see the article by Thomas G. Moehrle, pp. 38–46, this issue.)

Defining X-11

The X-11 filter is a general-purpose univariate method for filtering time series to attenuate the seasonal component. It was developed by Julius Shiskin and associates at the Bureau of the Census in the 1950's and 1960's. Assuming that a time series can be broken down as above into a trend-cycle component, seasonal component, and irregular component, the idea behind X-11 is to extract the various components with a series of symmetric moving-average filters.

While numerous options affect the procedure, the outline of X-11 can be described as follows. Starting with monthly data, the first step is to take a 12-term centered moving average of the data.[6] This is used as a first estimate of the trend-cycle component. The difference between the centered moving average and the original series is a first estimate of the sum of the seasonal and irregular components. Then, a five-term moving average is applied separately to each month, to extract the seasonal factors. The seasonal factors are further smoothed and are subtracted out to yield an estimate of the irregular components. At this point, extreme values of the irregulars are downweighted, and another series of moving-average filters is applied to obtain refined estimates of each of the components. The initial estimate of the trend is further smoothed, and the seasonals and irregulars are reestimated. The seasonally adjusted series is obtained by subtracting the final estimate of the seasonal from the raw series.

Why not just filter all data with X-11 before estimating a regression? This approach has two problems. First, filtering series that lack a seasonal component can induce spurious correlations in the adjusted series. Second, problems of hypothesis testing associated with X-11 filtering are surprisingly difficult to resolve.

The first point is made by Kenneth F. Wallis, who conducted simulations that show how the X-11 filter affects the autocovariance structure of several common models.[7] Wallis performs the exercise of generating data from a known model, filtering it with X-11, and trying to recover the parameters of the original model. Not surprisingly, he finds that it is difficult to recover the original coefficients from the filtered data.

For example, suppose one picks a value for b and generates data using the model

$$y_t = by_{t-1} + e_t.$$

Let y_t^a be the filtered data that we get from passing the original data through the X-11 procedure. Estimate the AR(1) model

$$y_t^a = b'y_{t-1}a + e_t,$$

and compare the estimates of b' with the known values of b that generated the simulated data. In this particular example, the model parameter b' estimated from the filtered data will be an *overestimate* of the true value of the parameter b.

Now consider a "white noise" series—a time series for which there are no significant autocorrelations in the data. Wallis' experiments show that the X-11 filtered data exhibit spurious "significant" positive and negative autocorrelations at various leads and lags.

For data with a strong periodic component, the simulation results are more surprising. Consider, for example, the AR model

$$y_t = by_{t-4} + e_t.$$

This model is consistent with, for example, stochastic seasonality in quarterly data. The autocorrelation plot for the $\{y_t\}$ time series shows peaks at lags 4, 8, 12, and so on and zeros at the other lags. The autocorrelation plot for the X-11 filtered data will show that correlation at the seasonal lags is reduced. However, filtering induces spurious positive correlations on all other lags. The conclusion is that X-11 filtering tends to scramble some univariate time series properties of the data.

Wallis addresses the question of how the relationship between a pair of time series is affected by filtering. Suppose the true model is

$$y_t = x_t b + e_t,$$

but instead we estimate the model with "adjusted" series obtained by linear filtering:

$$y_t^a = c_1 y_{t-1} + c_2 y_{t-2} + \ldots = \Sigma c_i L^i y_t = C(L)y_t$$
$$x_t^a = d_1 x_{t-1} + d_2 x_{t-2} + \ldots = \Sigma d_i L^i x_t = D(L)x_t.$$

Here, L is the lag function $Lx_t = x_{t-1}$. After filtering, we find that the true model becomes

$$y_t^a = C(L)y_t = C(L)(x_t b + e_t)$$
$$= (C(L)/D(L))x_t^a b_t + C(L)e_t$$
$$= \Sigma b_i x_{t-i} a + \Sigma c_i e_{t-i}.$$

What starts out as a simple, direct linear relationship in the raw data is greatly complicated by the filter: a simple first-order linear relationship becomes a complicated relationship involving a number of lags of the dependent variable. Note, however, that most of the complications are avoided if both series are run through the same filter, that is, if

$$C(L) = D(L).$$

Because the X-11 filter is nearly data independent, we can argue that this is usually the case. Then, the only ill effect associated with X-11 filtering is to induce autocorrelation in the error term. Although this complicates inference and hypothesis testing, in principle, one can take it into account when doing empirical work.

The second problem has to do with model estimation using the adjusted data. For example, estimating and extracting seasonal factors is analogous to extracting means from the data before calculating a regression. The estimation that is involved should affect hypothesis tests in some way, through, for instance, a degrees-of-freedom correction. Unfortunately, identifying the appropriate statistical adjustments is an unsolved problem of statistics.

This makes it difficult for analysts to interpret their empirical results. If one estimates a model using seasonally adjusted data, the test statistics that are generated by a statistical package are not entirely reliable. Thus, one is not sure how to interpret tests of hypotheses about model parameters:

> When the objective is to estimate regression coefficients, the ultimate appraisal of any seasonal adjustment procedure must be based on whether it improves the properties of the parameter estimates. From this, it becomes obvious that the adjustment procedure should be evaluated in the context of the econometric model in which it is used, and it must, therefore, be expected that whether a seasonal adjustment method improves the properties of the parameter estimates depends on the data, the method of parameter estimation, and the characteristics of the econometric model.[8]

Ideally, then, a seasonal filter should be estimated in conjunction with an economic model, and the analyst should take explicit account of the effects of seasonal adjustment in conducting hypothesis tests on estimated coefficients.

Some common filters

Three main approaches are available for modeling seasonality. One can model seasonality as a deterministic seasonal process, a stationary seasonal process, or a seasonally integrated process. The first of these assumes that the seasonal component has a purely deterministic explanation that does not vary in shape. This type of seasonality is modeled with seasonal dummy variables by the equation

$$y_t^S = \mu_t + e_t,$$

where e_t is "white noise" and

$$\mu_t = \Sigma_{i=1,p}\, m_i\, I_i(t),$$

in which p is the number of periods in the year (that is, 4 or 12, depending upon whether one is working with quarterly or monthly data), m_i is the mean for observations in the ith period of the year, and $I(t)$ is an indicator variable ($I_i(t) = 1$ if observation t falls into the ith period of a year, and $I_i(t) = 0$ otherwise). The seasonal pattern is represented by allowing each month or quarter to have a different mean. This is perhaps the most common approach to seasonal adjustment in empirical work.

The second approach is to model seasonality as a stationary stochastic process, generated by a stationary autoregression

$$y_t^S = \Sigma_{j=1,p}\, m_j\, y_{t-j}^S + e_t = B(L)\, y_t^S + e_t,$$

where e_t is "white noise." This corresponds to a seasonal process for which the magnitude or sign of the seasonal effects vary slowly through the sample.

Numerous refinements to these basic models are possible. One example is an increasingly popular method of modeling seasonal variation in which parameters are allowed to vary smoothly through time as a means of capturing time-varying seasonality. In a variation of this, we may wish to allow a model with a combination of seasonal dummies and stochastic seasonality in which all the parameters are slowly varying through time.[9] However, the basic building blocks of the time-varying parameter specifications are still a seasonal dummy component and a stochastic seasonal component.

Some work has been done on the problem of determining which of these two approaches fits the data better. Robert B. Barsky and Jeffrey A. Miron have evaluated the relative importance of deterministic seasonality and stochastic seasonality in U.S. macroeconomic data. They found that seasonal dummies seem to be more important than stochastic seasonal effects in modeling U. S. macroeconomic time series.[10] They also found that the estimated seasonal dummy coefficients appear to be quite stable over time and that the R^2 in the regression of detrended economic data on seasonal dummies is typically quite high. Seasonal dummies commonly capture nearly 85 percent of the short-term variation apart from the trend.

These findings suggest that the stationary stochastic model is perhaps less important as a description of seasonality in macroeconomic data. The stationary stochastic approach models seasonality as an AR process applied to zero mean innovations; hence, the means for the seasonal periods each have the same unconditional expectation. The finding of a consistent pattern in the seasonal dummy coefficients is contrary to what one might expect if the stochastic model were the primary description of seasonality.

The preceding findings must be qualified by the possibility that *neither* the deterministic *nor* the stationary stochastic model constitutes an appropriate model of seasonality. Accordingly, a third approach, the integrated seasonal process, is gaining much attention in the literature on applied economics. Seasonal integration occurs if there is a unit root in the autoregressive representation of the seasonal process. For example, for quarterly data, we might have a model such as $y_t^S = y_{t-4}^S + e_t$, where e_t is "white noise." That is, the seasonal part of our time series has the features of a random walk. The seasonally integrated model has properties that are quite different from the properties of the other two models. For example, it exhibits a "long memory" property, in which a single shock may permanently affect the observed seasonal pattern. Another important property of the model is that the implied variances of both the overall process time series and the seasonal component increase at a rate linear with time.

An important reason why one should be aware of seasonally integrated models is that the econometrics of nonstationary unit root processes are quite different from the econometrics of stationary processes. In general, the usual techniques of hypothesis testing give misleading results when applied to nonstationary data. The problem is that test statistics fail to converge to their expected limiting distributions. As a result, one can easily draw wrong conclusions on the nature of the model underlying one's data. For example, Philip Hans Franses, Svend Hylleberg, and Hahn S. Lee showed that if one estimates a seasonal dummy model on nonstationary data, statistical tests will tend to give spurious results.[11] Tests will appear to indicate that seasonal dummies are significant and that the model fits the data very well. The observed relationship is spurious because the seasonal dummy model will predict the future evolution of the data very poorly. A test for seasonal unit roots is given and applied in the literature.[12]

Application to economic models

All of the foregoing filtering methods rest on the assumption that one can decompose an economic time series into a trend-cycle component and an independent seasonal component. Several recent articles and papers have looked at whether the trend-cycle/seasonal breakdown is justified. Barsky and Miron's article was one of the first to focus on seasonal fluctuations in their own right, examining the role they played in economic data.[13]

Barsky and Miron found patterns in the seasonal dummy models estimated on macroeconomic data. One of their key findings was that there is a stable pattern in the seasonality of real output variables. The typical seasonal pattern is an increase in growth rates in the second and fourth quarters, a mild decrease in the third quarter, and a very large decrease in the first quarter. This pattern seems to be stable in the data over the postwar period, and it also holds for other countries.[14] For a typical country, gross domestic product peaks in the fourth quarter, rising between 4 and 5 percent. It then *falls* by between 5 percent and 10 percent during the subsequent first quarter.

Several articles followed Barsky and Miron in using seasonal dummy variables to examine seasonal patterns in macroeconomic data. Miron has written a summary of the findings that emerged from this literature.[15] In the business cycle, the production of goods in all the major sectors of the economy moves together. Similarly, the seasonal cycle seems to exist across various sectors of the economy, affecting each more or less in the same manner. This is somewhat counterintuitive, because analysts might expect a differential impact of seasonality across industry sectors.

Another key feature of the business cycle is the absence of production smoothing: production and sales move together, and inventories do not appear to be used to accommodate changes in demand. The same pattern also appears to hold in seasonal data. A third feature of the business cycle is that changes in the money stock and changes in output are correlated; one sees a similar pattern of correlation in money and output in the seasonal cycle, in which periods of high production are periods where the money stock is high, and periods of low production are periods in which money holdings are low. Finally, labor productivity is procyclical over both the seasonal and the business cycles.

The reliability of these results depends partly on one's ability to extract the trend from a time series before estimating the seasonal component. (If this is not done, spurious results are possible. For example, if a time series exhibits a secular downward trend, then unless the trend is extracted, seasonal dummies will extract it and will appear to be statistically significant.) Rather than depending on estimates of the trend component of a time series, we can estimate a model of business cycles that allows for seasonal effects. Several papers and texts have taken just this tack, and the results further tend to confirm the importance of seasonal factors in business cycle relationships.

Eric Ghysels has found that seasonal means are different in periods of expansion, compared with periods of recession.[16] (During a recession, the winter drop is slightly sharper, and the fall boom is not as marked.) Ghysels estimated a stochastic regime-shifting model and found that the transition probabilities varied with the seasons.[17] In particular, recessions are less likely to begin during the fourth-quarter boom and more likely to begin during the first-quarter downturn.

Ghysels' work presents problems for seasonal adjustment. At the core of seasonal adjustment methods is the assumption that the seasonal pattern is independent of patterns in economic trends and cycles. If seasonal components and cycles are related, seasonal adjustment may end up filtering out information that is useful and important in describing the economy.

When is seasonal adjustment appropriate?

Are seasonal adjustments still a good idea? To this question, we give the economist's typical answer: it depends. We have a collection of findings that suggests that, first, empirical findings may be suspect if seasonality is a factor but is ignored. Second, if seasonality is not present, and we attempt to filter it out anyway, we have additional problems. Third, there is probably no one best model of seasonality, and taking a specific approach to modeling the seasonal component involves making assumptions about the characteristics of the underlying process that must be carefully thought out and defended.

Finally, we question whether it makes sense to assume that we can break down a time series into a trend-cycle component and a seasonal component. Before conducting empirical work, economists should consider whether this approach to the time series is reasonable and whether the assumptions involved are appropriate for the model being examined. For example, we can imagine empirical estimation problems in which we are interested only in the total effect of one time series on another, and the decomposition of this total into a seasonal and a model component is not necessary to answer the empirical question being considered. In any event, economists should be aware of the tradeoffs that are implicit in seasonally adjusting economic time series. □

Footnotes

[1] Svend Hylleberg, *Seasonality in Regression* (Orlando, FL, Academic Press, Inc., 1986).

[2] Arthur F. Burns and Wesley Clair Mitchell, *Measuring Business Cycles* (Cambridge, MA, National Bureau of Economic Research, 1946).

[3] *Ibid.*

[4] Christopher A. Sims, "Seasonality in Regression," *Journal of the American Statistical Association*, September 1974, pp. 618–27.

[5] Using vector notation, let $x = x^{TC} + x^{S}$ and $y = y^{TC} + y^{S}$. We have assumed that the trend-cycle and seasonal components for x and y are mutually uncorrelated. Then the least squares estimate of the parameter vector b' is

$$b' = (x'x)^{-1}x'y = (x'x)^{-1}\,(x^{TC}\,y^{TC})\,b + (x'x)^{-1}\,(x^{S}\,y^{S})c.$$

That is, b' is a weighted average of the trend-cycle parameter b and the seasonal parameter c.

[6] The idea behind a moving average is to replace a given term in a time series with an average including that term and the observations on either side. To illustrate, suppose we observe the following time series:

$$x_1\ x_2\ x_3 \ldots$$

We first define a three-term moving average. This would replace each term by a simple average of the term itself, the one before it in the list, and the one after it. So we would replace term two above with the term $\hat{x}_2 = (x_1 + x_2 + x_3)/3$. A five-term moving average would similarly replace a term with the average of the term itself and the two terms on either side. Thus, observation number three would be replaced with an average,

$$\hat{x}_3 = (x_1 + x_2 + x_3 + x_4 + x_5)/5.$$

Moving averages of an even number of terms are a bit tougher. For technical reasons, we would like to keep our moving averages symmetric, that is, to have the same number of terms on either side of our observation included in the average. *Centering* refers to the manner in which this is accomplished.

Here are the steps in calculating a four-term centered moving average around observation three. First, calculate the average of the first four terms. The result is

$$x_{3a} = (x_1 + x_2 + x_3 + x_4)/4.$$

Then calculate the average of terms two through five:

$$x_{3b} = (x_2 + x_3 + x_4 + x_5)/4.$$

We would replace term three with the average of these two averages:

$$\hat{x}_3 = (x_{3a} + x_{3b})/2.$$

Another way to describe this, then, is to say that a four-term centered moving average is a two-term average of a four-term moving average. Similarly, a 12-term centered moving average is a 2-term average of a 12-term moving average.

[7] Kenneth F. Wallis, "Seasonal Adjustment and Relations between Variables," *Journal of the American Statistical Association*, March 1974, pp. 18–31.

[8] Hylleberg, *Seasonality in Regression*, p. 36.

[9] See, for example, Estela B. Dagum and Benoit Quenneville, "Dynamic Linear Models for Time Series Components," *Journal of Econometrics*, January–February 1993, pp. 333–51, for an overview of these methods.

[10] Robert B. Barsky and Jeffrey A. Miron, "The Seasonal Cycle and the Business Cycle," *Journal of Political Economy*, June 1989, pp. 503–34.

[11] Philip Hans Franses, Svend Hylleberg, and Hahn S. Lee, *Spurious Deterministic Seasonality*, Report 9355/A (Rotterdam, Erasmus University, 1993).

[12] See Svend Hylleberg, C.W. Engle, B.S. Granger, and B.S. Yoo, "Seasonal Integration and Co-integration," *Journal of Econometrics*, May 1992, pp. 215–38; Philip Hans Franses, *Testing for Seasonal Unit Roots in Monthly Data*, Report 9032/A (Rotterdam, Erasmus University, 1990); and J. Joseph Beaulieu and Jeffrey A. Miron, "Seasonal Unit Roots in Aggregate U. S. Data," *Journal of Econometrics*, January–February 1993, pp. 305–28.

[13] Barsky and Miron, "The Seasonal Cycle and the Business cycle."

[14] Beaulieu and Miron, "Seasonal Unit Roots."

[15] Jeffrey A. Miron, *The Economics of Seasonal Cycles*, Working Paper 3522 (Cambridge, MA, National Bureau of Economic Research, 1990).

[16] Eric Ghysels, *On Seasonal Asymmetries and Their Implications for Stochastic and Deterministic Models of Seasonality*, Working Papers, Department of Economics (Montreal, University of Montreal, 1991).

[17] Eric Ghysels, *A Time Series Model with Periodic Stochastic Regime Switching* (Montreal, University of Montreal, 1993). See also J.D. Hamilton, *Time Series* (Princeton, NJ, Princeton University Press, 1994).