

SEASONALITY: PROJECT TEMPLATE

(The attached PDF file has better formatting.)

Many candidates examine seasonality in their time series student projects. Your project can focus on seasonality, have a section on seasonality, or include a test for seasonality.

This project template summarizes seasonality issues that can be used in student projects. Most of this project template explains the types of seasonality and adjustment methods. The last page outlines a student project comparing the adjustment methods for CPI.

Take heed: We do *not* require the optimal seasonality adjustment for a student project. The sophisticated seasonal adjustments actually used for most economic time series are not covered in this course. We explain here the methods discussed in the textbook, and we indicate when each is appropriate.

Scope: Some student projects focus on seasonality, examining methods of identifying and correcting for seasonality or comparing two time series for seasonality. An analysis of ARIMA models with and without seasonal adjustments can form a good student project.

We examine seasonality in almost every time series. Even time series with no obvious seasonality, such as stock prices, over-night interest rates, inflation, claim severities, and business profits, may have seasonal patterns.

Recommendation: Include a section on seasonality in your student project. Form graphs of monthly averages (or quarterly averages) and 12 month sample autocorrelations, and explain whether they show seasonality.

Data: For a student project focusing on seasonality, choose data with a seasonal pattern.

- ~ The seasonality of interest rates is offset by the FED's monetary policy.
- ~ The global interest rate patterns of the past two decades have weaker seasonality, since many parts of the world do not share America's holiday shopping.

Interest rates are not a suitable time series for a student project on seasonality. But a good choice for a project on seasonality is the CPI indices on the NEAS web site.

- The non-seasonally adjusted index shows seasonality.
- The seasonally adjusted index does not show much seasonality.

Your student project might look at the following time series:

- The two CPI indices on the NEAS web site.
- The first differences of each CPI index.
- The non-seasonally adjusted index with a seasonality adjustment.

- The first differences of this seasonally adjusted time series.

The data collection is minimal, since all the data are on the NEAS web site.

Weather: Projects on daily temperature and other weather patterns deal with seasonality. For daily temperature and rainfall, de-seasonalize the data. The project template on daily temperature discusses the procedures in more detail. Hourly temperature forms a good project on seasonality, since it overlays two patterns: 24 hour pattern and 365 day pattern.

Take heed: An excellent student project topic is rainfall in Los Angeles or Chicago or any other large city with heavy smog from weekday traffic. Some people say that the build-up of smog particles during weekdays and the dispersion of smog during week-ends causes a seven day seasonality in rainfall. You might search for smog levels in various cities: many large cities provide daily for hourly smog reports. You can see if smog has a seven day seasonality and fit an ARIMA process to a regression of rainfall on smog levels.

ARIMA MODELING FOR CPI INDICES

The CPI index is not stationary. Your student project can fit ARIMA processes to the CPI, its first differences, or the first differences of its logarithms. You can use several methods to de-seasonalize the index.

Take heed: The CPI index has an exponential trend. Instead of first differences, take logarithms and first differences. Alternatively, take ratios of CPI index values in successive periods and logarithms of these ratios. The ratios have a lognormal distribution, and the logarithms of the ratios have a normal distribution.

Your student project explains how to test for stationarity, identify the type of trend, and derive the stationary time series.

- ~ Examine the seasonality in each time series: the CPI indices, your adjusted CPI index, and the first differences of each. You can also use the seasonally adjusted CPI, to see if that gives a better ARIMA fit.
- ~ Fit an ARIMA process to each time series.
- ~ Compare the ARIMA process for the seasonally adjusted index with the ARIMA process for the index with a simple seasonality adjustment.

The seasonal adjustments actually used for the CPI is more sophisticated than the simple seasonality adjustment in the textbook. An ARIMA process should fit better to the seasonally adjusted CPI than to the other indices. Compare the goodness-of-fit with the Box-Pierce Q statistic and Bartlett's test.

This is a simple student project, using data solely from the NEAS web site and the time series techniques in the textbook. For a time series of CPI price level figures, we expect an ARIMA (1,2,0) or an ARIMA (2,2,0) process.

- One difference converts the price level to the inflation rate. Don't forget to use logarithms if you see an exponential trend.
- If the inflation rate is a random walk, we take differences to form a white noise process.
- If the inflation rate has a trend, we take differences to eliminate the trend.

INFLATION AND THE MONEY SUPPLY

For a more sophisticated student project, you can examine the non-seasonally adjusted index divided by the money supply. The FED adjusts the money supply to remove the seasonality in interest rates. The change in the money supply causes a proportional change in the price level, though the effect is not immediate. Because the economic effects are not immediate and we have only rough monthly figures, you won't get strong relations.

INTERVALS

Much seasonality in the time series used in student projects relates to end of the year shopping, summer vacations, and weather. With monthly data, you can identify spikes in January, August, December, or any other relevant month.

- For end of the year shopping, quarterly data are fine. Many business time series have quarterly figures, which differentiate higher sales in November/December from lower sales in January through March. For inflation, weather, and other business patterns, quarterly data doesn't provide fine enough information.
- For daily temperature, use days. See the project template on daily temperature for an example of fitting an ARIMA process to a highly seasonal time series.

Many time series have seasonal patterns: vacation travel in August, student job seekers in May through July, and weddings in June. Graph the monthly figures for your time series.

Insurance time series like claim frequency and severity have strong seasonal patterns in many lines of business. An analysis of seasonality for loss cost trends in personal auto or workers' compensation is valuable to insurers. Your student project will benefit your employer in addition to getting VEE credit.

GRAPHS

Graph the time series. Many monetary time series have an inflationary trend, so the seasonality may be more evident in the first differences than in the original series. If the trend is exponential, take logarithms and first differences. Form graphs of each series: the initial time series, its logarithms, and the first differences.

If the data are stochastic, fluctuations in the time series may obscure the seasonal pattern. We have two methods of identifying seasonality in stochastic data, corresponding to the two method of correcting for seasonality: multi-year averages and correlograms.

Illustration: We examine home sales by month in a small state. The values are stochastic, and home sales depend on economic conditions and interest rates, so the seasonality in any year is unclear. We use a 20 year average or a correlogram of 20 years of data to see the seasonality.

METHODS

Seasonality has several forms. The proper method of dealing with seasonality depends on the type of seasonality. Your student project can focus on the optimal method of correcting for seasonality. We contrast three scenarios, covering most time series.

Scenario #1: The current value depends on the time of year, not on the value at the same date one year ago.

Illustration: The expected daily temperature depends on the time of the year. It may be 25° on February 15 and 95° on August 15. It does not depend on the daily temperature one

year ago. If the daily temperature is 45° on February 15, 20X8, we expect it to be 25° on February 15, 20X9, not 45°.

If we do not deseasonalize the data, we get high autoregressive (ϕ) parameters and sub-optimal ARIMA models. The illustration below shows the rationale.

Illustration: Suppose the proper model is an AR(1) process, where μ is the average daily temperature for that day, ϕ_1 is 20%, the average daily temperature varies greatly through the year (from 25° F to 95° F), and σ is high.

The δ parameter varies from 20° F to 76° F, depending on the time of year. Instead of a δ parameter, the ARIMA process might use a centered moving average of twenty values.

An ARIMA model may have $\phi_1 = 20\%$ and $\phi_2 = \phi_3 = \dots = \phi_{10} = \phi_{356} = \phi_{357} = \dots = \phi_{366} = 4\%$. We actually want $\delta = 80\% \times$ expected temperature for that day, which we estimate as a centered moving average of 20 surrounding days.

A better ARIMA model may have $\phi_1 = 20\%$ and values of 1% for 80 other lags, using 20 day moving averages from four years. An even better model may have $\phi_1 = 20\%$ and values of 0.1% for 800 other lags, using 20 day moving averages from 40 years.

This is wrong. Don't use autoregressive parameters to de-seasonalize the data.

- Separate the seasonality and the autoregressive process by deseasonalizing the data.
- Then fit an AR(1) or AR(2) process.

Scenario #2: Autoregressive Seasonality Parameter

The current value depends on the value N periods back, not on the time of the year.

Illustration: We model monthly auto insurance policy counts with an ARIMA process. The auto policy has a six month term, and 90% of policyholders renew their policies.

We don't expect policy counts to be higher or lower in any month. (This is not strictly true. There is a slight seasonality in auto insurance policy counts stemming from high auto sales in the fall and higher home sales in the summer, but the effect on policy counts is small.) But if policy counts were high in month j , they are high in month $j+6$.

Illustration: Auto insurance policy counts have no seasonality. Insurer ABC had a sales competition in September 20X6, with double commissions to agents with 40 or more new policies. Policy counts in September 20X6 were 20% higher than in other months.

With six month policies, renewals in March 20X7 will be 20% higher than in other months. This is not inherent seasonality: policy counts in March 20X6 were no higher than usual.

The policy counts have no (significant) seasonality. The long-term monthly averages are about equal. The simple seasonal adjustment has no effect. But the autoregressive parameter for a lag of six months is high. We model the seasonality by the ARIMA process.

PRINCIPLES FOR COMPANY SALES VOLUME, BY TYPE OF PRODUCT:

For products without renewal purchases, we expect an autoregressive process with a high ϕ_1 parameter or a moving average process with a negative θ_1 parameter, depending on (i) the type of product and (ii) company sales vs industry sales. See the module postings for the modules on autoregressive and moving average processes.

For non-insurance products with high consumer loyalty, we have two scenarios:

- For short-duration products, like cigarettes and beer, high consumer loyalty raises the ϕ_1 parameter in a time series of monthly sales. We can't distinguish repeat purchases from autoregressive effects.
- For long-duration products, like autos and mobile phones, consumer loyalty is lower and the lag between purchases varies widely. If autos last three to ten years, a 50% consumer loyalty to a given model slightly raises the ϕ_j parameter for lags of three to ten years. But stochasticity in sales overwhelms these effects.

Property-casualty insurance has high renewal rates and 6 or 12 month terms. The ARIMA process for auto policy counts may have a ϕ_6 coefficient of 85% and a ϕ_1 coefficient of 5%.

The difference in the type of seasonality is seen in multi-year averages and correlograms spanning several years.

- If a high 12 month autocorrelation stems from weather patterns or holiday sales, the multi-year average shows the pattern more strongly.
- If the 12 month autocorrelation reflects consumer loyalty, the effect wears off over time.

If the insurer has an 85% renewal rate (retention rate), the renewal effect after ten years is $0.85^{20} = 3.88\%$. Even with a 95% renewal rate, the effect after ten years is $0.95^{20} = 35.85\%$. But if policy counts vary much from month to month by random fluctuations, the ϕ_6 autoregressive parameter may be 80% to 90%.

The correlogram shows the seasonality clearly. The correlogram shows the short-term 6 month sample autocorrelation over the past ten years. The average of 20 half-years eliminates the random fluctuations. The sample autocorrelations of lags 5 and 7 are close to zero, and the sample autocorrelation of lag 6 is 85%.

The difference in the type of seasonality is seen in the correlogram:

- For daily temperature, the 120 month sample autocorrelation is about the same as the 12 month sample autocorrelation.
- For policy counts, the 120 month sample autocorrelation is near zero.

Scenario #3: Combination

Many seasonal time series are a combination of these two patterns.

Illustration: A sporting goods store sells both summer and winter sports equipment: surf-boards and bicycles in the summer and ski equipment in the winter.

The relative volume of winter vs summer goods varies over the years.

- A cold and rainy summer may depress sales of surf-boards.
- A cold winter with heavy snow may raise sales of ski equipment.

An ARIMA process of monthly sales is distorted by the seasonality.

- If we know the relative volume of summer vs winter goods, we use the relative volume to adjust for seasonality in sales.
- Our best estimate of the relative volume is last year's relative volume. But last year's monthly sales volume is affected by random fluctuations that do not repeat.

Illustration: Suppose the store gets much of its revenue from sales of bicycles in May, June, and July. If the weather last year was exceptionally rainy in May but clear in June, sales may have been low in May and high in June.

The optimal adjustment for seasonality may be a combination of a long-term average by month, and last year's sales by month.

Illustration: We examine a time series of unemployment rates for workers age 18 to 25 by month in Boston. Boston is a college town, with many students who look for jobs in May and June, so unemployment rates are high. In November and December, retail stores hire young workers for high holiday sales, so unemployment rates are low. The effects vary from year to year, depending on minimum wage laws and the types of employers looking for staff each year.

- Stores selling children's toys or video equipment have high demand for sales clerks in November and December.
- Firms offering part-time programming work hire college students in summers.

The stochasticity of unemployment rates affects the optimal seasonality adjustment.

- If the unemployment rates are not stochastic, we use last year's monthly relativities to adjust for seasonality.
- If the unemployment rates are highly stochastic, we use a multi-year average to adjust for seasonality.

For many time series, the proper method to adjust for seasonality is unclear. Your student project can adjust a time series with both methods and compare the residuals.

Illustration: Begin with the monthly CPI with no seasonal adjustment. Use two models:

1. Seasonally adjust the data using average monthly relativities. Take logarithms and first differences, and fit an ARIMA process.
2. Take logarithms and first differences, and fit an ARIMA process with a 12 month seasonal term.

Choose *corresponding* processes to compare the two models.

- If the first model is an AR(2) process, the second model may be an AR(2) process with a ϕ_{12} parameter as well.
- If the optimal model differs for the two methods, choose one model. If the first model is an AR(2) process, and the second model is an AR(1) process with a ϕ_{12} parameter, use the AR(2) process for both methods.

Use Bartlett's test and the Box-Pierce Q statistic to select the adjustment for seasonality. One adjustment may lead to an ARIMA process with a lower Box-Pierce Q statistic than the other adjustment.