Lesson 1 Times Series Basics

1.1 Overview of Time Series Characteristics

- Univariate Time Series

A univariate time series is a sequence of measurements of the same variable collected over time. Most often, the measurement are mode at regular time intervals.

Note

1. The data are not necessary independent
2. The data are not necessarily identically distributed
3. Order matters because there is dependency and changing the order could change the meaning of the data

- Objective of the Analysis

1. Describe the importance features of the time series pattern
2. Explain how the last effects the future
3. Explain how two time series can interact
4. Forecast future values of the series

- Types of Models

There are two basic types of time-domain models

1. Models that relate the present value of a series to past values and past prediction errors (ARIMA).
2. Ordinary regression models that use time indices as x-variables. These can be helpful for an initial description of the data and form the basis of several simple forecasting methods.

- Important Characteristics to Consider First

1. Trend: the measurements tend to increase or decrease over time

2. Seasonality: a regular repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, etc

3. Outliers: outliers are far away from the other data

4. Long-run cycle or period

5. Constant variance

6. Abrupt changes (to either the level of the series or the variance)

- The AR(1) Model

One of the simplest ARIMA type models is a model in which we use a linear model to predict the value at the present time using the value at the previous time. AR(1) Stands for autoregressive model of order 1.

Theoretically, the AR(1) model is written

$$X_t = \delta + \phi_1 X_{t-1} + W_t$$

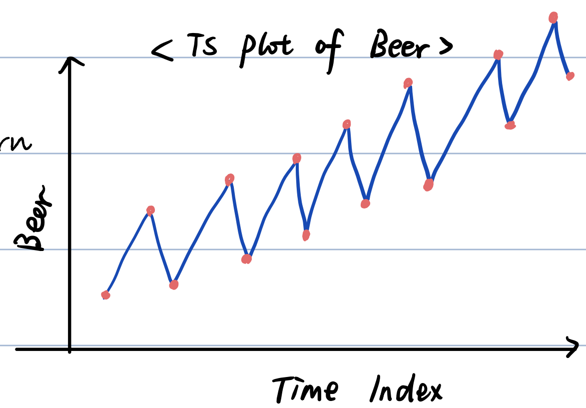
Assumptions:

1. $W_t \overset{iid}{\sim} N(0, \sigma_w^2)$

2. Errors w_t are independent of x

- Classic Regression. Ethos for trend and seasonal effects

ARIMA methods can deal with series that exhibit both trend and seasonality, but we can also use ordinary regression methods.



To use traditional regression methods, we might model the pattern in the beer production data as a combination of the trend over time and quarterly effect variables.

Suppose that the observed series is X_t, for $t = 1,2,3,\ldots,n$

- For a linear trend, use t as a predictor variable in a regression
- For a quadratic trend, we might consider using both t and t
- For a quarterly data, with possible seasonal (quarterly) effects, we can define indicator variables such as Sj=1 if the observation is in quarter j of the year and 0, otherwise. There are 4 such indicators.

Let $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. A model with additive components for linear trend and seasonal (quarterly) effects might be written

$$X_t = \beta_1 t + \alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_3 + \alpha_4 S_4 + \varepsilon_t$$

To add a quadratic trend, which may be the case in the beer plot, the model is

$$X_t = \beta_1 t + \beta_2 t^2 + \alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_3 + \alpha_4 S_4 + \varepsilon_t$$

- Residual Analysis

After fitting a model, we can check the residuals to see if they violate our assumptions

- Sample Autocorrelation Function (ACF)

The sample autocorrelation function for a series gives correlations between the series X_t and lagged values of the series for lags of 1,2,3, and so on. The lagged values can be written as X_{t-1}, X_{t-2}, X_{t-3}, and so on. The ACF gives correlations between X_t and X_{t-1}, X_{t-2}, and X_{t-3}, …

The ACF can be used to identify the possible structure of time series data. The ACF of the residuals for a model is also useful. The ideal for a sample ACF of residuals is that there are not any significant correlations for any lag.

1.2 Sample ACF and Properties of AR(1) Model

This lesson defines the sample autocorrelation function (ACF) in general and devices the pattern of the ACF for an AR(1) model. An AR(1) model is a linear model that predicts the present values of a time series using the immediately prior value in time.

- Stationary Series

For an ACF to make sense, the series must be a weakly stationary series. This means that the autocorrelation for any particular lag is the same regardless of where we are in time.

- Weakly Stationary Series

A series x_t is said to be weakly stationary if it satisfies the following properties:

1. The mean $E(x_t)$ is the same for all t
2. The variance of x_t is the same for all t
3. The covariance and also correlation between x_t and x_{t-h} is the same for all t at each lag h

- Autocorrelation Function (ACF)

Let x_t denote the value of a time series at time t. The ACF of the series gives correlations between x_t and x_{t-h} for h = 1,2,3,... Theoretically, the autocorrelation between x_t and x_{t-h} equals

$$\text{Covariance}(x_t, x_{t-h})/\text{Variance}(x_t)$$

- The First-Order Autoregression Model (AR(1))

In this model, the value of x at time t is a linear function of the value of x at time t-1. The algebraic expression of the model is as follows:

$$X_t = \delta + \phi_1 X_{t-1} + w_t$$

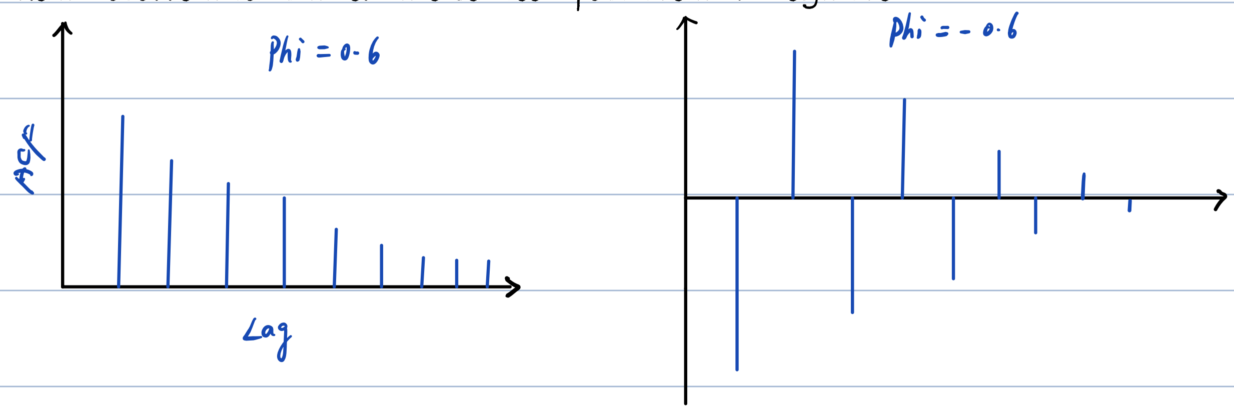
Assumptions

1. $w_t \overset{iid}{\sim} N(0, \sigma_w^2)$, meaning that the errors are independently distributed with a normal distribution has mean 0 and constant variance
2. $|\phi_1| < 1$ (For weakly stationary)

- Patterns of ACF for AR(1) Model

For a positive value of phi1, the ACF exponentially decreases to 0 as the lag h increase.

For a negative value of phi1, the ACF exponentially decreases to 0 as the h increase but the algebraic signs for the autocorrelations alternate between positive and negative.



Lesson 2 MA Models, Partial Autocorrelation, Notational Convention

2.1 Moving Average Models

Time series models known as ARIMA models may include autoregressive terms and/or Moving Average terms. We learned an autoregressive term for the variable x_t is a lagged value of x_t.

- Moving Average

A moving average term is a past error multiplied by a coefficient.

Let $w_t \overset{iid}{\sim} N(0, \sigma_w^2)$ meaning that w_t are identically independently distributed, each with a normal distribution having mean 0 and the same variance.

$$MA(1):\quad X_t = \mu + W_t + \theta_1 W_{t-1}$$

$$MA(2):\quad X_t = \mu + W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2}$$

$$MA(q):\quad X_t = \mu + W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \cdots + \theta_q W_{t-q}$$

- ACF for General MA(q) Models

$$\begin{cases} ACF \neq 0, & lag \leq q \\ ACF = 0, & lag > q \end{cases}$$

- Invertibility of MA Models

An MA model is said to be invertible if it is algebraically equivalent to a converging infinite order AR model. By converging, we mean that the AR coefficients decrease to 0 as we move back in time.

2.2 Partial Autocorrelation Function

A partial correlation is a conditional correlation.

For instance, consider a regression context in which y is the response variable and x_1, x_2, and x_3 are predictor variables. The partial correlation between y and x_3 are related to x_1 and x_2. In time series, y is the variable of x at time t, which is x_t.

We can define the partial correlation in our example as

$$\frac{Covariance(y, x_3 | x_1, x_2)}{\sqrt{Var(y | x_1, x_2)\, Var(x_3 | x_1, x_2)}}$$

It shows that the partial correlation is related to

1. Regression in which we predict y from x_1 and x_2
2. Regression in which we predict x_3 from x_1 and x_2.

Basically, we correlate the "parts" of y and x_3 that are not predicted by x_1 and x_2.

- Partial Autocorrelation

Similarly, for time series, the partial autocorrelation between x_t and x_{t-h} is defined as the conditional correlation between x_t and x_{t-h}, conditional on $x_{t-h+1}, \ldots, x_{t-1}$, the set of observations that come between the time points t and t-h.

1. PACF of MA(1): equals to the 1st order autocorrelation
2. PACF of MA(2): $$\frac{Covariance(X_t, X_{t-2} | X_{t-1})}{\sqrt{Var(X_t | X_{t-1})\, Var(X_{t-2} | X_{t-1})}}$$
3. PACF of MA(3): $$\frac{Covariance(X_t, X_{t-3} | X_{t-1}, X_{t-2})}{\sqrt{Var(X_t | X_{t-1}, X_{t-2})\, Var(X_{t-3} | X_{t-1}, X_{t-2})}}$$

- Some Useful Facts About PACF and ACF Patterns

1. identification of an AR model is with PACF

For an AR model, the theoretical PACF "shuts off" past the order of the model.

2. Identification of an MA model is with ACF

For a MA model, the PACF tapes towards 0 and ACF will have non-zero autocorrelations only at lags involved in the model.

2.3 Notation Convention

Time series models involve lagged terms and may involve differences data to account for trend. There are useful notations use for each.

- Backshift Operator

Using B before either a value of the series x_t or an error term w_t means to move that element back one time. For example, $BX_t = X_{t-1}$, $B^2X_t = X_{t-2}$

- AR Models and the AR Polynomial with Backshift Operator

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t$$

$$X_t = \delta + \phi_1 \cdot BX_t + \phi_2 (B^2 X_t) + W_t$$

$$X_t - \phi_1 \cdot BX_t - \phi_2 \cdot B^2 X_t = \delta + W_t$$

$$(1 - \phi_1 B - \phi_2 B^2) \cdot X_t = \delta + W_t$$

$$\Phi(B) \cdot X_t = \delta + W_t \quad ; \quad \Phi(B) = 1 - \phi_1 B - \phi_2 B^2$$

AR polynomial for AR(2)

- MA Models with Backshift Operators

$$X_t = \mu + W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2}$$

$$X_t = \mu + W_t + \theta_1 BW_t + \theta_2 B^2 W_t$$

$$X_t = \mu + (1 + \theta_1 B + \theta_2 B^2) W_t$$

$$X_t = \Theta(B) W_t + \mu \quad ; \quad \Theta(B) = 1 + \theta_1 B + \theta_2 B^2$$

MA polynomial for MA(2)

- Differencing

Differencing is used to account for non-stationary that occurs in the form of trend and/or seasonality.

$\because \quad X_{t-1} = BX_t$

$\therefore \quad X_t - X_{t-1} = X_t - BX_t$

$$= \underbrace{(1-B)}_{\nabla} X_t$$

$\therefore \quad X_t - X_{t-1} = \nabla X_t$

$X_t - X_{t-2} = \nabla_2 X_t$

$X_t - X_{t-12} = \nabla_{12} X_t$

$(X_t - X_{t-1}) + (X_{t-1} - X_{t-2}) = (X_t - BX_t) + (BX_t - B^2 X_t)$

$$= (1 - B^2) X_t$$

Lesson 3 ARIMA

3.1 Non-seasonal ARIMA Models

ARIMA models possibly include autoregressive terms, moving average terms, and differencing operations.

- Identifying a Possible Model

1. plot a time series to check trend, seasonality, constant variance, outliers
2. ACF and PACF can be used together to determine the order of AR and MA

(1) if the ACF does not tail off but has values that stay close to 1 over many lags, the series is non-stationary and differencing will be needed.

(2) if the series autocorrelations are non-significant, then the series is random. You are done at that point.

(3) if the first differences were necessary and all the differences autocorrelations are non-significant, then the original series is called a random walk and you are done.

- Estimating and Diagnosing a Possible Model

1. most software will use maximum likelihood estimation methods to make the estimates.
2. Look at the significance of the coefficients
3. Look at the ACF of the residuals
4. Look at Box-Pierce tests for possible residual autocorrelation at various lags
5. Look at a plot of residuals versus fits and/or a time series plot of the residuals for non-constant variance.

- Model Selection

Sometimes, more than one model can seem to work for the same dataset. When that is the case, we can use the followings to decide between the models:

1. Possibly choose the model with the fewest parameters
2. Pick the model with generally lowest standard errors for predictions of the future
3. Compare models with regard to statistics such as the MSE, AIC, AICc, and BIC. Lower values of these statistics are desirable.

3.3 Forecasting with ARIMA Models

In an ARIMA model, we express x_t as a function of past values of x and/or past errors. When we forecast a value past the end of the series, we might need values from the observed series on the right side of the equation, or we might need values that are not yet observed.

In general, assuming a sample size of n, the following procedure is as follows:

1. For any w_j with $1 <= j <= n$, use the sample residual for time point j
2. For any w_j with $j > n$, use 0 as the value of w_j
3. For any x_j with $1 <= j <= n$, use the observed value of x_j
4. For any x_j with $j > n$, use the forecasted value of x_j

Lesson 4 Seasonal Models

4.1 Seasonal ARIMA Models

In a seasonal ARIMA model, seasonal AR and MA terms predict x_t using data values and errors at times with lags that are multiples of S, which defines the number of time periods until the pattern repeats again.

1. With monthly data, say S=12, a seasonal first order autoregressive model would use x_{t-12} to predict x_t. For instance if we were selling cooling fans, we might predict this August's sales using last August's sales.
2. A seasonal second order autoregressive model would use x_{t-12} and x_{t-24} to predict x_t. Here we would predict this August's values from the past two Augusts.
3. A seasonal first order MA(1) model with S=12 would use w_{t-12} as a predictor. A seasonal second order MA(2) model would use w_{t-12} and w_{t-24}

- Differencing

Almost by definition, it may be necessary to examine differenced data when we have seasonality. Seasonality usually causes the series to be nonstationary because the average values at some particular times within the seasonal span may be different than the average values at other times.

1. Seasonal differencing

Seasonal differencing is defined as a difference between a value and a value with lag that is a multiple of S.

E.g. $x_t - x_{t-4}$

Seasonal differencing removes seasonal trend and can also get rid of a seasonal random walk type of nonstationary

2. Non-Seasonal Differencing

If trend is present in the data, we may also need non-seasonal differencing. Often a first difference will defend the data, that is, we use $x_t - x_{t-1}$ in the presence of trend

Seasonal and non-seasonal differencing can be used together. With seasonal data, it is likely that short run non-seasonal components will still contribute to the model. In the cooling fans example, sales in the previous month or two, along with sales from the same month a year ago, may help predict this month's sales.

- Seasonal ARIMA Model

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model.

$$\underbrace{ARIMA(p,d,q)}_{\text{e.g. previous month may affect current month} \to \text{non-seasonal}} \times \overbrace{(P,D,Q)S}^{\text{previous year may affect current year} \to \text{seasonal}}$$

Where p=non-seasonal AR order

d=non-seasonal differencing

q=non-seasonal MA order

P=seasonal AR order

D=seasonal differencing order

Q=seasonal MA order

S=time span of repeating seasonal pattern

- Identifying a Seasonal Model

Step 1: do a time series plot of the data to examine any trend and seasonality

Step 2: do any necessary differencing if there is any trend or seasonality (linear trend can do first order differencing, curved trend can do transformation)

Step 3: Examine the ACF and PACF of the differenced data (if differencing is necessary).

We are using this information to determine possible models, Some basic guidance:

A. Non-seasonal terms: examine the early lags (1,2,3...) to judge non-seasonal terms

B. Seasonal terms: examine the patterns across lags that are multiple of S (e.g. 12, 24, 48...)

Step 4: Estimate the model that might be reasonable on the basis of step 3

Step 5: Estimate the model(s) that might be reasonable

Step 6: Examine the residuals to see if the model seems good (can use AIC/BIC) for model comparison)

Lesson 5 Smoothing and Decomposition Methods

5.1 Decomposition Models

Decomposition procedures are used in time series to describe the trend and seasonal factors in a time series.

One of the main objectives for a decomposition is to estimate seasonal effects that can be used to create an present seasonally adjusted values. A seasonally adjusted value removes the seasonal effect from a value so that trends can be seen more clearly.

- Basic Structure

The following two structures ae considered for a basic decomposition models:

1. Additive: x_t=trend+seasonal+random (used when the seasonal variation is relatively constant)
2. Multiplicative: x_t=trend•seasonal•random (used when the seasonal variation is increase over time)

- Basic Steps in Decomposition

Step 1: Estimate trend using moving average or linear regression

Step 2: De-trend the series (for additive model, substrate the trend, for multiplicative model, dividing)

Step 3: Estimate the seasonal factor using the detrend series.

For monthly data, estimate the effect for each month of the year. For quarterly data, estimate the effect for each quarter. For example, to get the seasonal effect for January, we could average the de-trended values for all Januaries in the series.

Step 4: Determine the random component

For additive model, random=series-trend-seasonal

For multiplicative model, random=series/(trend•seasonal)

The random component could be analyzed for such as mean or variance, or even for whether the component is actually random, or be modeled with an ARIMA model.

A few programs iterate through the steps 1 to 3. For example, after step 3, we could use the seasonal factors to de-seasonalize the series and then return to step 1 to estimate the trend based on the de-seasonalized item.

5.2 Smoothing Time Series

Smoothing is usually done to help us better see patterns, like trends. Generally smooth out the irregular roughness to see a clear signal. For seasonal data, we might smooth out the seasonality so that we can identify the trend. Smoothing does not provide us with a model, but it can be a good first step in describing various components of the series.

- Moving Average

The traditional use of the term, moving average, is that, at each point in time, we determine averages of observed values that surround a particular time.

For example, at time t, a "centered moving average of length 3" with equal weights would be the average of values at time t-1, t, and t+1.

To take away seasonality from a series so we can better see trend, we would use a moving average with a length=seasonal span. Thus in the smoothed series, each smoothed value has been averaged across all seasons.

For example, for quarterly data, we could define a smoothed value of time t as $(x_t+x_{t-1}+x_{t-2}+x_{t-3})/4$, the average of this time and the previous 3 quarters.

To smooth away seasonality in quarterly data, in order to identify trend, the usual convention is to use the moving average smoothed at time t is

$$\frac{1}{8}X_{t-2} + \frac{1}{4}X_{t-1} + \frac{1}{4}X_t + \frac{1}{4}X_{t+1} + \frac{1}{8}X_{t+2}$$

That is, we apply weight 1/8 to values on the right and left side, and weight 1/4 to all the values in the middle.

- Non-Seasonal Noise

For non-seasonal series, you ar not bound to smooth over any particular span. For smoothing you should experiment with moving averages of different spans. Those spans of time could be relatively short. The objective is to knock off the rough edges to see what trend or pattern might be there.

- Other Smoothing Methods

LOWESS (local weighted regression) is widely used. Single exponential smoothing is another option too.

Lesson 6 The Periodogram

6.1 The Periodogram

Any time series can be expressed as a combination of cosine and sine waves with differing periods (how long it takes to complete a full cycle) and amplitudes (maximum/minimum value during the cycle). This fact can be utilized to examine the periodic behavior in a time series.

- Periodogram

A Periodogram is used to identify the dominant periods of a time series. This can be helpful for identifying the dominant cyclical behavior in a series, particularly when the cycle are not related to the commonly encountered monthly or quarterly seasonality.

- Properties of a Cosine Function

For discrete time, these definitions are useful for cosine wave:

1. period

(T) is the number of time periods required to complete a single cycle of the cosine function

2. Frequency

w=1/T. It is the fraction of the complete cycle that is completed in a single time period.

Imagine fitting a single cosine wave to a time series observed in discrete time. Suppose that we write this cosine wave as

$$X_t = A \cdot \cos(2\pi \omega t + \phi)$$

- A is the amplitude. It determines the maximum absolute height of the curve
- w is the frequency. It controls how rapidly the curve oscillates
- Phi is the phase. It determines the starting point, in angle degrees, for the cosine wave

One goal of an analysis is to identify the important frequencies or periods in the observed series. A starting tool for doing this is the Periodogram. The Periodogram graphs a measure of the relative importance of possible frequency values that might explain the oscillation pattern of the observed data. Suppose that we have observed data at n distinct time points, and for convenience we assume that n is even. Our goal is to identify important frequencies in the data. To pursue the investigation, we consider the set of possible frequencies $w_j = j/n$ for $j=1,2,3,\ldots,n/2$. These are called the harmonic frequencies. We will represent the time series as

$$X_t = \sum_{j=1}^{\frac{n}{2}} \left[\beta_1\left(\frac{j}{n}\right)\cos(2\pi w_j t) + \beta_2\left(\frac{j}{n}\right)\sin(2\pi w_j t)\right]$$

This is a sum of sine and cosine functions at the harmonic frequencies. A mathematical device called the Fast Fourier Transform is used. After the parameters has been estimated, we define

$$P\left(\frac{j}{n}\right) = \hat{\beta}_1^2\left(\frac{j}{n}\right) + \hat{\beta}_2^2\left(\frac{j}{n}\right)$$

This is the value of the sum of squared "regression" coefficients at the frequency j/n.

- Interpretation and Use

A relatively large value of P(j/n) indicates relatively more importance for the frequency j/n in explaining the oscillation in the observed series. The dominant frequencies might be used to fit cosine waves to the data, or might be used simply to describe the important periodicities in the series.

Lesson 8 Regression with ARIMA errors, Cross Correlation Functions, and Relationships between Two Time Series

8.1 Linear Regression Models with Autoregressive Errors

When we do regression using time series variables, it is common for the errors to have a time series structure. This violates the usual assumption of the independent errors made in ordinary least squares regression. The consequence is that the estimates of coefficients and their standard errors will be wrong if the time series structure of the errors is ignored.

It is possible to make adjustments when the errors have a general ARIMA structure.

- The Regression Model with AR Errors

Suppose that y_t and x_t are time series variables. A simple linear regression model with autoregressive errors can be written as

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

with $\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \cdots + \omega_t$, and $\omega_t \sim iid\ N(0, \sigma^2)$

If we let $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots$, then we can write the AR model for the errors as $\phi(B)\varepsilon_t = \omega_t$

So the model can be written as

$$y_t = \beta_0 + \beta_1 x_t + \phi^{-1}(B)\omega_t$$

where w_t is the usual white noise series.

Notes: this generalizes to the multiple linear regression structure as well. We can have more than one x-variable (time series) on the right side of the equation. Each x-variable is adjusted in the manner described below.

- Examining Weather This Model May be Necessary

1. start by doing an ordinary regression. Store the residuals

2. Analyze the time series structure of the residuals to determine if they have an AR structure

3. If the residuals from the ordinary regression appear to have an AR structure, estimate this model and diagnose whether the model is appropriate

- Carrying Out the Procedure

The basic steps are

1. Use OLS regression to estimate the model

(note: we are modeling a potential trend over time with $\beta_0 + \beta_1 t$ and may need to de-trend x if x is strongly correlated with time t.

2. Examine the ARIMA structure (if any) of the sample residuals from the model in step 1

3. If the residuals do have an ARIMA structure, use MLE to simultaneously estimate the regression model using ARIMA estimation for the residuals.

4. Examine the ARIMA structure (if any) of the sample residuals from the model in step 3. If white noise is present, then the model is complete. If not, continue to adjust the ARIMA model for the errors until the residuals are white noise.

8.2 Cross Correlation Functions and Lagged Regressions

In the relationship between two time series (y_t and x_t), the series y_t may be related to past lags of the x-series. The sample cross correlation function (CCF) is helpful for identifying lags of the x-variable that might be useful predictors of y_t

The sample CCF is defined as the set of sample correlations between x and y_t for h=0, +/-1, +/-2, +/-3, and so on. A negative value for h is a correlation between the x-variable at a time before t and the y-variable that might be useful predictors of y_t. For instance, consider h=-2, the CCF value would give the correlation between x_{t-2} and y_t.

- When one or more x_{t+h}, with h negative, are predictors of y_t, it is said that x leads y
- When one or more x_{t+h}, with h positive, are predictors of y_t, it is said that x lags y

In some problems, the goal may be to identify which variable is leading and which is lagging. In many situations, we will examine the x-variables to be a leading variables of y-variable because we will want to use values of x-variable to predict future values of y. Thus, we usually look at what is happening at the negative values of h on the CCF plot.

- Transfer Function Models

In a full transfer function model, we model y_t as potentially a function of past lags of y_t and current and past lags of the x-variables. We also usually model the time series structure of the x-variables as well.

- Detect CCF

We can use the CCF in time series packages to find out which lags of x-variable are correlated with y-variable. Usually, we would start with lags having a high correlations.

- Scatterplots

Time series packages have functions to produces scatterplots of y_t versus x_{t+h} for negative h from 0 back to a lag that you specify.

- Regression Models

There are a lot of models that we could try based on the CCF and lagged scatterplots. We could experiment with several models and compare the results and residuals to choose the one making more sense.

- Complications

The CCF pattern is affected by the underlying time series structures of the two variables and the trend that each series has. It is always helpful to de-trend and take into account the univariate ARIMA structure of the x-variable before graphing the CCF.

Lesson 9 Prewhitening and Intervention Analysis

9.1 Pre-whitening as an Aid to Interpreting the CCF

The basic problem that we are considering is the construction of a lagged regression in which we predict a y-variable at the present time using lags of an x-variable (including lag 0) and lags of the y-variable. One difficulty is that the CCF is affected by the time series structure of the x-variable and any "in common" trends the x and y series may have over time.

One strategy for dealing with this difficulty is called "pre-whitening". The steps are:

1. Determine a time series model for the x-variable and store the residuals from this model
2. Filter the y-variable series using the x-variable model. In other words, we find the difference between observed y-values and "estimated" y-values based on the x-variable model from step 1.
3. Examine the CCF between the residuals from step 1 and filtered y-values from step 2. This CCF can be used to identify the possible terms for a lagged regression.

This strategy stems fro the fact that, when the input series (say w_t) is white noise, the patterns of the CCF between w_t and z_t, (z_t is a linear combination of w_t), are easily identifiable and derived. Step 1 above creates a "white noise" series as the input. Conceptually, step 2 above arises from a starting point that y-series = linear combination ? of x-series. If we "transform" the x-series to white noise (residuals from its ARIMA model), ? then we should apply the transformation to both sides of the equation to preserve an equality ? of sorts.

It is not absolutely crucial that we find the model for x exactly. We just want to get close to the white noise input situation.

Pre-whitening is just used to help us identify which lags of x may predict y. After identifying possible model from the CCF, we work with the original variables to estimate the lagged regression.

Alternative strategies to pre-whitening include:

1. looking at the CCF for the original variables-this sometimes works

2. Detrending the series using either first differences or linear regressions with time as a predictor

Note: it is hard for me to understand the step 2 of pre-whitening. I will circle back when necessary in the future.

9.2 Intervention Analysis

Suppose that at time t=T(where T will be known), there has been an intervention to a time series. By intervention, we mean a change to a procedure, or law, or policy, etc. That is intended to change the values of the series x_t. We want to estimate how much the intervention has changed the series. For example, suppose that a region has instituted a new maximum speed limit on its highways and wants to learn how much the new limit has affected accident rates.

Intervention analysis in time series refers to the analysis of how the mean level of gets after an intervention, when it is assumed that the same ARIMA structure for the series x_t holds both before and after the intervention.

- Overall Intervention Model

Suppose that the ARIMA model for x_t (the observed series) with no intervention is

$$x_t - u = \frac{\Theta(B)}{\phi(B)} w_t$$

with the usual assumptions about the error series w_t.

$\Theta(B)$ is the usual MA polynomial and $\phi(B)$ is the usual AR polynomial.

Let z=the amount of change at time t that is attributable to the intervention. $z_t = 0$ before time T (time of the intervention). The value of z_t may or may not be 0 after time T.

Then, the overall model, including the intervention effect, may ne written as

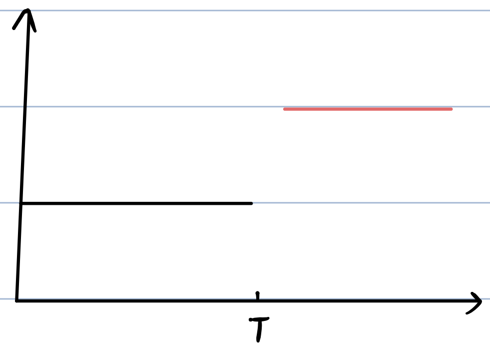
$$X_t - u = \frac{\Theta(B)}{\phi(B)} W_t + Z_t$$

- Possible Patterns for Intervention Effect (patterns for z_t)

There are several possible patterns for how an intervention may affect the values of a series for t>=T (the intervention point). Four possible patterns are as follows:

- Pattern 1

  Permanent constant change to the mean level: An amount has been added or substrates to each value after time T.

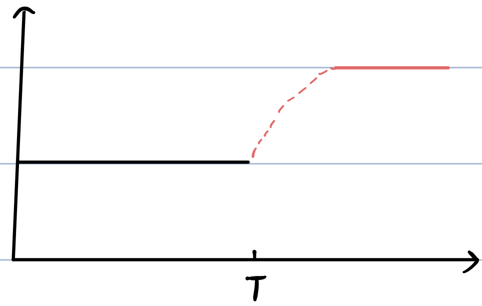


- Pattern 2

  Brief constant change to the mean level: there may be a temporary change for one or more periods, after which there is no effect of the intervention.

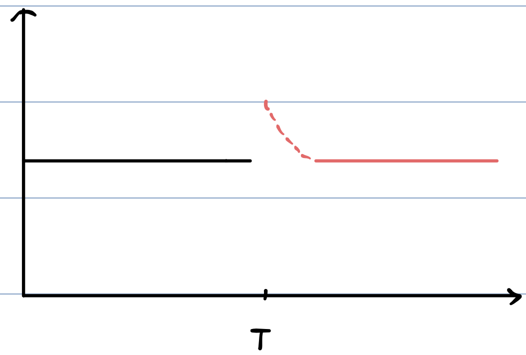
- Pattern 3

  Gradual increase or decrease to new mean level: there may be a gradually increasing amount that is added or substrates which eventually levels off at a new level.



○ Pattern 4

Initial change followed by gradual return to the no change: There may be immediate change to the values of the series, but the amount added or substrates to each value after time T approaches 0 over time.



- Models for the Pattern

z_t = the amount of change at time t that is attributable to the intervention.

Suppose that I_t is an indicator variable such that $I_t=1$ when $t>=T$ and $I_t=0$ when $t<T$.

○ Pattern 1

$$Z_t = \delta_0 \cdot I_t$$

$$= \begin{cases} \delta_0, & t \geq T \\ 0, & t < T \end{cases}$$

$$Z_t = \delta_0 (1-B^d) \cdot I_t$$

$$= \begin{cases} \delta_0, & T \leq t \leq T+d \\ 0, & o.w. \end{cases}$$

- Pattern 3

$$Z_t = \omega_1 Z_{t-1} + \delta_0 I_t \quad ; \quad |\omega_t| < 1$$

$$= \begin{cases} 0 , & t < T \\ \omega_1 Z_{t-1} + \delta_0 = \dfrac{\delta_0 (1 - \omega_t^{\,t-T+1})}{1 - \omega_1} \end{cases}$$

- Pattern 4

$$Z_t = \omega_1 Z_{t-1} + \delta_0 P_t \quad ; \quad |\omega_t| < 1$$

$$= \begin{cases} 0 , & t < T \ (Z_{t-1} = 0 , \ P_t = 0) \\ \delta_0 , & t = T \ (Z_{t-1} = 0 , \ P_t = 1) \\ \omega_1 Z_{t-1} , & t > T \ (Z_{t-1} \neq 0 , \ P_t = 0) \end{cases}$$

The coefficients, ω_1, δ_0, will be estimated using the data.

- Estimating the Intervention Effect

Two parts of the overall model have to be estimated - the basic ARIMA model for the series and the intervention effect. Several approached have been proposed. One approach has the following steps:

1. Use the data before the intervention point to determine the ARIMA model for the series
2. Use that ARIMA model to forecast values for the period after the intervention
3. Calculate the differences between actual values after the intervention and the forecasted values
4. Examine the differences in step 3 to determine a model for the intervention effect

What we do after step 4 depends on a viable software. If the right program is available we can use all of the data ton estimate the overall model that combines the ARIMA for the series and the intervention model. Otherwise, we might use only the differences from step 4 above to make estimates of the magnitude and nature of the intervention.

Lesson 10 Longitudinal Analysis/Repeated Measured

The term repeated measures refers to experimental designs (or observational studies) in which each experimental unit or subject is measured at several points in time. The term longitudinal data is also used for this type of data.

- Typical Design

Experimental units are randomly allocated to one of g treatments. A short time series is observed for each observation. An example in which there are 3 treatment groups with 3 units per treatment, and each unit is measured at four times is as follows:

Treatment	Unit	T_1	T_2	T_3	T_4
1	1				
1	2				
1	3				
2	4				
2	5				
2	6				
3	7				
3	8				
3	9				

In the example design, we have observed nine short time series, one for each experiment unit. The focus of the analysis would be to determine treatment differences, both in mean level and patterns across time. We may also want to characterize the overall time pattern.

- A commonly used Model

One approach to analyzing these data is the ANOVA model

$$Y = \text{treatment} + \text{Animal}(\text{treatment}) + \text{time} + \text{treatment} \cdot \text{time} + \text{error}$$

  - Animal is a random factor assumed to have mean 0 and an unknown constant variance. "Random" means that the animals are considered to be a random sample from a larger population of animals
  - The notation Animal(Treatment) specifies that animals were nested in treatment meaning that different animals were in the different treatment groups
  - The Treatment factor measures whether the mean response differs for different treatments when we average over all animals and all times
  - The Time factor measures whether the mean response differs over time when we average over all animals and all treatments
  - The Time•Treatment interaction which is sensitive to whether the pattern across time depends upon the specific treatment used
  - The errors are assumed to be independently normally distributed with mean 0 and constant variance

- Consequence of the Assumptions for This Model - Compound Symmetry

The assumptions for this model lead to another assumption- that, in theory, the correlation is equal for all time gaps between observations. For instance, the correlation between data at times 1 and 2 is the same as the correlation between data at times 1 and 3, and is also the same as the correlation between times 1 and 4. This is called the compound symmetry assumption.

More flexible models for the correlation structure of the observations across time are also available.

Lesson 11 Vector Autoregressive Models/ARCH Models

11.1 ARCH/GARCH Models

An ARCH (autoregressive conditionally heteroscedastic) is a model for the variance of a time series. ARCH models are used to describe a changing, possibly volatile variance. Although an ARCH model could possible be used to describe a gradually increasing variance over time, most often it is used in situations in which there may be short periods of increased variation. (Gradually increasing variance connected to a gradually increasing mean level might be better handled by transforming the variable.)

ARCH models were created in the context of econometric and finance problems having to do with the amount that investments or stocks increase (or decrease) per time period, so there is a tendency to describe them as models for that type of variable. For that reason, STAT510 suggests that the variable of interest in these problems might either be $y_t = (x_t - x_{t-1})/x_{t-1}$, the proportion gained or lost since the last time, or $\log(x_t/x_{t-1}) = \log(x_t) - \log(x_{t-1})$, the logarithm of the ratio of this time's value to last time's value. It is not necessary that one of these be the primary variable of interest. An ARCH model could be used for any series that has periods of increased or decrease variable. This might, for example, be a property of residuals after an ARIMA model has been fit to the data.

Note: ARCH is used to model the heteroscedastic variance)

- The ARCH(1) Variance Model

Suppose that we are modeling the variance of a series y_t. The ARCH(1) model for the variance of model y_t is that conditional on y_{t-1}, the variance at time t is

$$\mathrm{var}(y_t \mid y_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2$$

We impose the constrains $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$ to avoid negative variance.

Note!

The variance at time t, σ_t^2, is connected to the value of the series at time t-1, y_{t-1}. A relatively large value of y_{t-1}^2 gives a relatively large value of the σ_t^2.

If we assume that the series has mean = 0 (this can always be done by centering), the ARCH model could be written as

$$y_t = \sigma_t \varepsilon_t \qquad \text{with} \qquad \sigma_t = \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2}$$
$$\varepsilon_t \overset{iid}{\sim} D(0, 1)$$

For inference (and maximum likelihood estimation), we could also assume that the error terms are normally distributed.

- Generalization

An ARCH(m) process is one for which the variance at time t is conditional on observations at the previous m times, and the relationship is

$$Var(y_t \mid y_{t-1}, \cdots, y_{t-m}) = \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_m y_{t-m}^2$$

with certain constrains imposed on the coefficients, by y_t series squared will theoretically be AR(m).

A GARCH (generalized autoregressive conditionally hesteroscedastic) model uses values of the past squared observations and past variances to model the variance at time t. As an example, a GARCH(1,1) is

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

In the GARCH notation, the first subscript refers to the order of the y^2 terms on the right side, and the second subscript refers to the order of the σ^2 terms.

- Identifying an ARCH/GARCH Model in Practice

The best identification tool may be a time series plot of the series. It is usually easy to spot periods of increased variation sprinkled through the series. It can be fruitful to look at the ACF and PACF of both y_t and y_t^2. For instance, if y_t appears to be white noise and y_t^2 appears to be AR(1), then an ARCH(1) model for the variance is suggested. If the PACF of the y_t^2 suggests AR(m), then ARCH(m) may work. GARCH models may be suggested by an ARMA type look to the ACF and PACF of y_t^2. In practice, you might have to experiment with various ARCH and GARCH structures after spotting the need in the same time series plot of the series.

11.2 Vector Autoregressive Models , VAR(p) models

VAR models are used for multi variate time series. The structure is that each variable is a linear function of past lags of itself and last lags of the other variables.

As an example, suppose that we measure three different time series variables, denoted by $x_{t,1}$, $x_{t,2}$, and $x_{t,3}$. The vector Autoregressive model of order 1, denoted as VAR(1), is as follows:

$$x_{t,1} = \alpha_1 + \phi_{11} x_{t-1,1} + \phi_{12} x_{t-1,2} + \phi_{13} x_{t-1,3} + w_{t,1}$$

$$x_{t,2} = \alpha_2 + \phi_{21} x_{t-1,1} + \phi_{22} x_{t-1,2} + \phi_{23} x_{t-1,3} + w_{t,2}$$

$$x_{t,3} = \alpha_3 + \phi_{31} x_{t-1,1} + \phi_{32} x_{t-1,2} + \phi_{33} x_{t-1,3} + w_{t,3}$$

Each variable is a linear function of the lag 1 values for all variables in the set.

In a VAR(2) model, the lag 2 values for all variables are added to the right side of the equations. In the case of three x-variables, there would be six predictors on the right side of each equation, three lag 1 terms and three lag 2 terms.

In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable.

VAR models are specific case of more general VARMA models. VARMA models for multi variate time series include the VAR structure above along with moving average terms for each variable. More generally, these are special cases of ARMAX models that allow for the addition of other predictors that are outside the multivariate set of principal interest.

Lesson 12 Spectral Analysis

Skip this lesson for now will review in the further.

Lesson 13 Fractional Differencing and Threshold Models

13.1 Long Memory Models and Fractional Differences

Long memory models may possibly be used when the ACF of the series tapers slowly to 0. The usual solution in this situation is to explore the first differences of the series.

Often, data for which a first difference is successful will typically have a first lag autocorrelation quite close to 1. In this case, we would fit an ARMA model with a coefficient equal to 1 for the first lag of AR term. This creates a first order autocorrelation for the original series close to 1.

In some instances, however, we may see a persistent pattern of non-zero correlations that begins with a first lag correlation that is not close to 1. In these cases, models that incorporate "fractional differencing" may be useful. A simple model that utilizes fractional differencing is

$$(1-B)^d X_t = w_t$$

where d is a value such that $|d|<0.5$ and w_t is the usual white noise term.

In a fractionally differences model, the difference coefficient d is a parameter to be estimated. It can be done in statistical package. Again, an indication that this model might be useful is a slowly tapering sample ACF without particularly high autocorrelations.

- Generalizations

The model can be expanded to include AR and MA terms as well as the fractional difference. These models are called ARFIMA models. To identify an ARFIMA model, we first use the simple fractional difference model $(1-B)^d X_t = w_t$ and then explore the ACF and PACF of the residuals from this model. This is analogous to exploring the ACF and PACF of the first differences when we carry out the usual steps for non-stationary data.

- Interpretation Difficulty

The main difficulty is that a fractional difference is difficult to interpret. Basically, it is a mathematical device that is used to expand a model into a high order AR with autocorrelations that match the persistent "long memory" pattern of the ACF of the series.

13.2 Threshold Models

Threshold models are used in several different areas of statistics, not just time series. The general idea is that a process may behave differently when the values of a variable exceed a certain threshold. That is, a different model may apply when values are greater than a threshold than when they are below he threshold. For example, in drug toxicology applications, I may be that all doses below a threshold amount are safe whereas there is increasing toxicity as the dose is increased above the threshold amount.

Threshold models are a special case of regime switching models (RSM). In RMS modeling, different models apply to different intervals of values of some key variable(s).

Threshold autoregressive models (TAR) are used for unitarians time series. In a TAR model, AR models are estimated separately in two or more intervals of values as defined by the dependent variable. These AR models may or may not be of the same order. For convenience , it is often assumed that they are of the same order.

One feature for that that a TAR model may work is that the rates of increase/decrease may diffe when the values are above some level than then the values are below that level.

The estimation of the threshold level is more or less subjective. Many analysts explore several different threshold levels in an attempt to provide a good fit to the data as measured by MSE values and general features of the residuals. The orders of the AR models can also be trial and error expedition, especially when the inherent model for the data may not be an AR. Generally, analysts start with what they think may be a higher order than necessary and then reduce the order as necessary.