GENERAL ASSEMBLY

MULTIVARIATE TIME SERIES

Chris Connell
Data Scientist

LEARNING OBJECTIVES

- Understanding of multivariate time series
- Be able to fit parameters

REVIEW

TIME SERIES

TIME SERIES MODELING

- In previous classes, we focused on exploring time series data, common statistics for time series analysis, and running univariate modeling.

- In this class, we will advance those techniques to show how to predict or forecast forward from time series data using multiply variables.

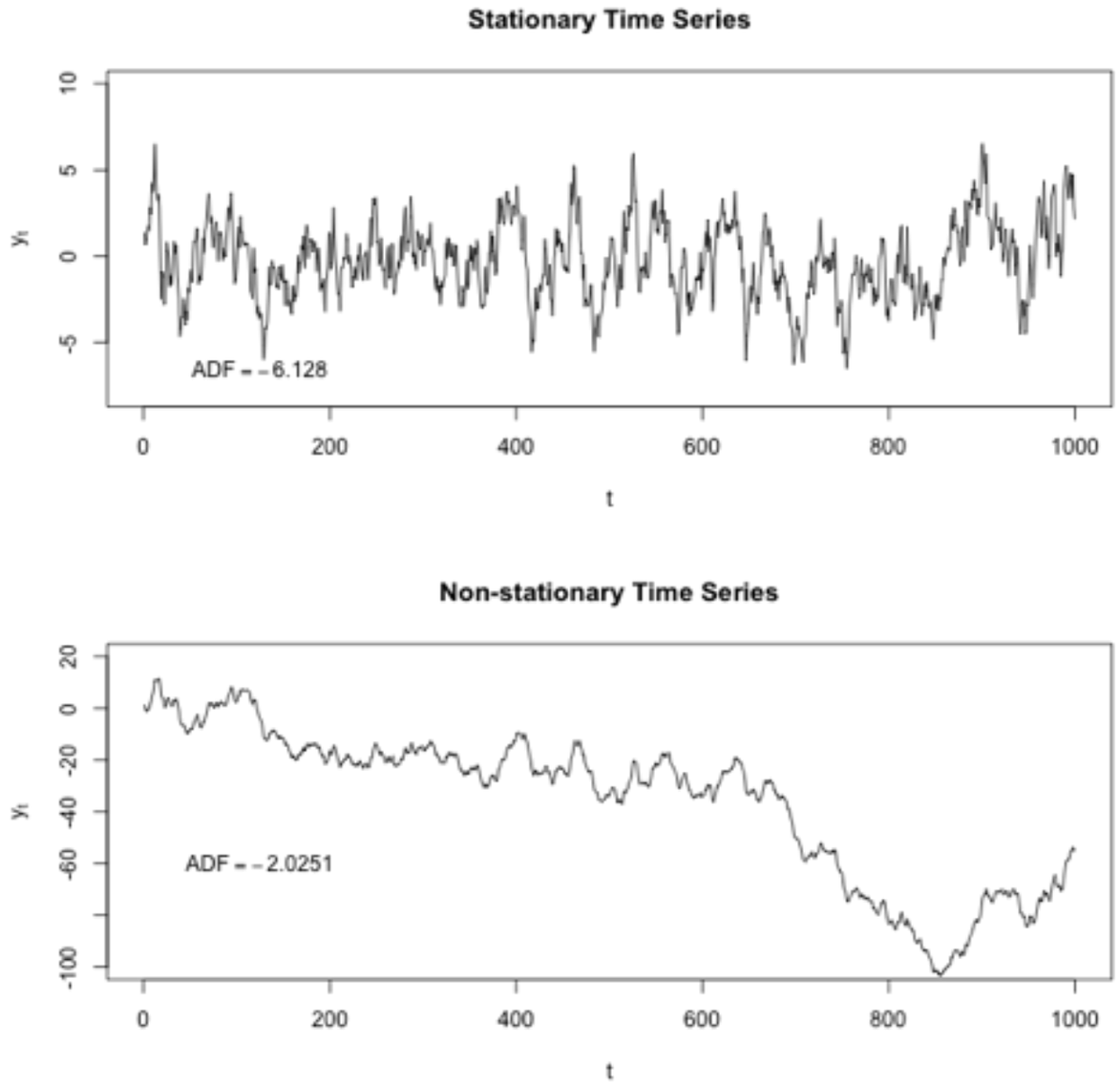
- With a sequence of values (a time series), we will use the techniques in this class to predict a future value.

WHAT ARE TIME SERIES MODELS?

- The training dataset would likely contain data from *before* AND *after* a test dataset.

- This would not be possible in real life (you can't use future, unseen data points when building your model). Therefore, it's not a valid test of how our model would perform in practice.

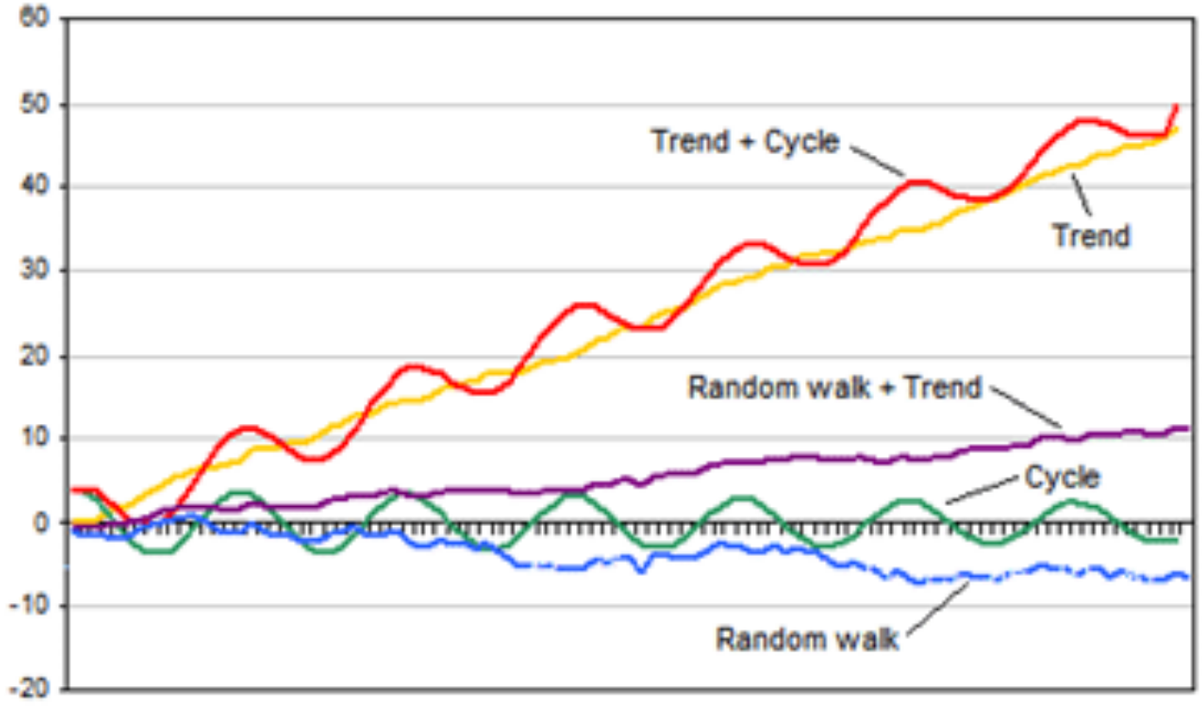
PROPERTIES FOR TIME-SERIES PREDICTION



PROPERTIES FOR TIME-SERIES PREDICTION

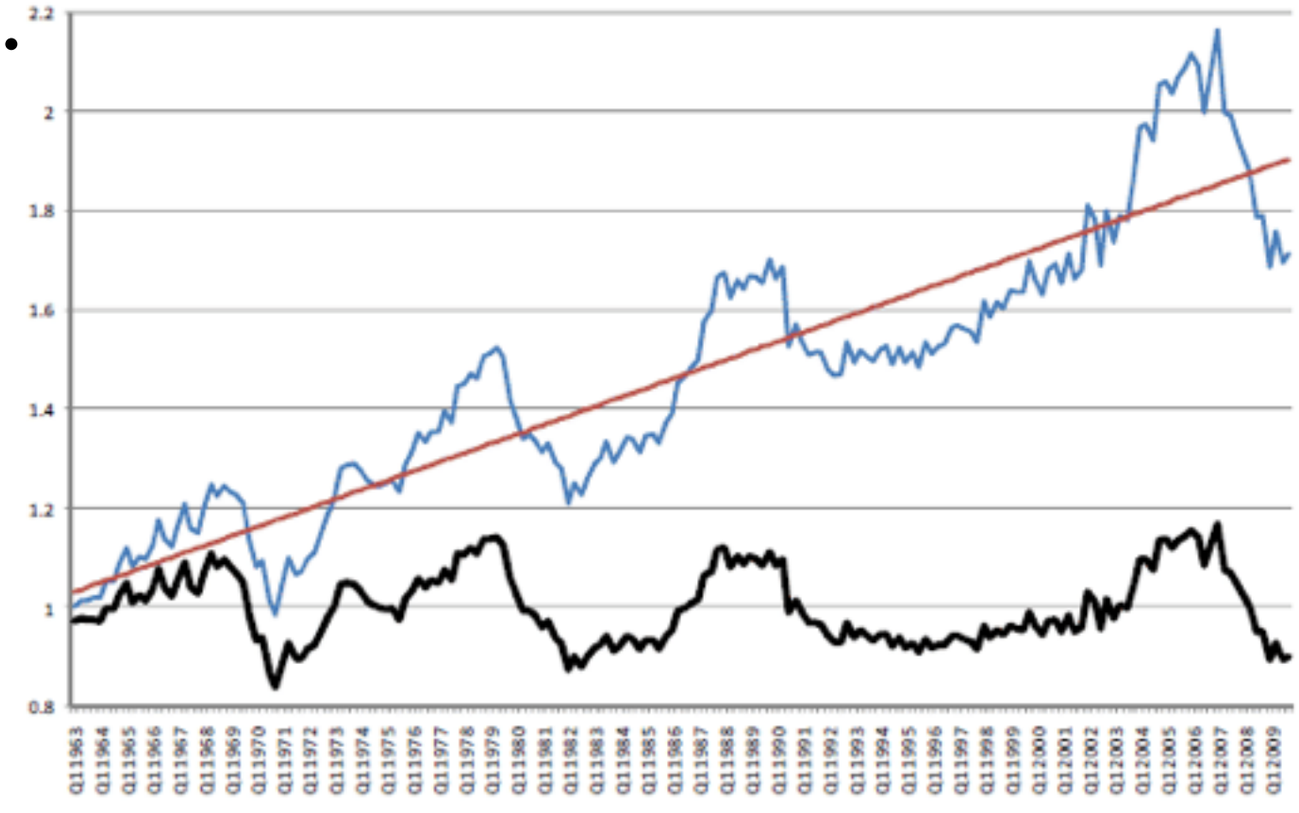
- Below are simulated examples of non-stationary time series and why they might occur.

Table 1 Non-stationary behavior



PROPERTIES FOR TIME-SERIES PREDICTION

- Below is an example where we look at US housing prices over time. Clearly, there is an upward trend, making the time series non-stationary (ie: the mean house price is increasing).



AR MODELS

- Autoregressive (AR) models are those that use data from previous time points to predict the next.
- This is very similar to previous regression models, except as input, we take the previous outcome.
- If we are attempting to predict weekly sales, we use the sales from a previous week as input.
- Typically, AR models are notes AR(p) where *p* indicates the number of previous time points to incorporate, with AR(1) being the most common.

MA MODELS

- **Moving average (MA) models**, as opposed to AR models, do not take the previous outputs (or values) as inputs. They take the previous error terms.
- We will attempt to predict the next value based on the overall average and how off our previous predictions were.

MA MODELS

- This model is useful for handling specific or abrupt changes in a system.
- AR models slowly incorporate changes in the system by combining previous values; MA models use prior errors to quickly incorporate changes.
- This is useful for modeling a sudden occurrence - something going out of stock or a sudden rise in popularity affecting sales.

IDENTIFYING BASE COEFICIENTS

Model	ACF	PACF
MA(q): moving average of order q	Cuts off after lag q	Dies down
AR(p): autoregressive of order p	Dies down	Cuts off after lag p
ARMA(p,q): mixed autoregressive-moving average of order (p,q)	Dies down	Dies down
AR(p) or MA(q)	Cuts off after lag q	Cuts off after lag p
No order AR or MA (White Noise or Random process)	No spike	No spike

DEMO

TESTING FOR STATIONARITY

ADF

- An Augmented Dickey–Fuller test (ADF) is a test for a unit root or stationarity in a time series sample

- The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

- If the test is significant, we reject the null hypothesis that the series is non-stationary

ACTIVITY: KNOWLEDGE CHECK

EXERCISE

ANSWER THE FOLLOWING QUESTIONS

Check for stationarity in the variables from either the Walmart or Rossman data set.

DELIVERABLE

Answers to the above questions

INTRODUCTION

VECTOR AUTOREGRESSION

VAR

- The vector autoregression (VAR) model is one of the most successful, flexible,and easy to use models for the analysis of multivariate time series.

- The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models.

- It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models.

VAR

- The basic VAR(p) model may be too restrictive to represent sufficiently the main characteristics of the data. In particular, other deterministic terms such as a linear time trend or seasonal dummy variables may be required to represent the data properly. Additionally, stochastic exogenous variables may be required as well.

VAR

- The vector autoregression (VAR) model is one of the most successful, flexible,and easy to use models for the analysis of multivariate time series.

- The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models.

- It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models.

DEMO

VAR & LAGGED ORDER SELECTION

INTRODUCTION

STRUCTURAL ANALYSIS

STRUCTURAL ANALYSIS

- In structural analysis, certain assumptions about the causal structure of the data under investigation are imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables on the variables in the model are summarized. These causal impacts are usually summarized with impulse response functions, forecast error variance decompositions, and granger causality.

IMPULSE REPONSE FUNCTIONS

- In economics, and especially in contemporary macroeconomic modeling, impulse response functions are used to describe how the economy reacts over time to exogenous impulses, which economists usually call shocks, and are often modeled in the context of a vector autoregression.

- Impulses that are often treated as exogenous from a macroeconomic point of view include changes in government spending, tax rates, and other fiscal policy parameters.

- Impulse response functions describe the reaction of endogenous macroeconomic variables such as output, consumption, investment, and employment at the time of the shock and over subsequent points in time.

GRANGER CAUSALITY

- One of the main uses of VAR models is forecasting. The structure of the VAR model provides information about a variable's or a group of variables' forecasting ability for other variables. The following intuitive notion of a variable's forecasting ability is due to Granger (1969).

- If a variable, or group of variables, y1 is found to be helpful for predicting another variable, or group of variables, y2, then y1 is said to Granger-cause y2; otherwise it is said to fail to Granger-cause y2.

- The notion of Granger causality does not imply true causality. It only implies forecasting ability.

FORECAST ERROR VARIANCE DECOMPOSITION

- Forecast error variance decomposition (FEVD) is used to aid in the interpretation of a vector autoregression (VAR) model once it has been fitted

- The variance decomposition indicates the amount of information each variable contributes to the other variables in the autoregression.

- It determines how much of the forecast error variance of each of the variables can be explained by exogenous shocks to the other variables.

DEMO

IRF, FEVD, AND GRANGER CAUSALITY

ACTIVITY: KNOWLEDGE CHECK

ANSWER THE FOLLOWING QUESTIONS

What is the difference between AR() and MA()?

DELIVERABLE

Answers to the above question on your desk

DEMO

FORECASTING

INDEPENDENT PRACTICE

FINAL PROJECT WORK

CONCLUSION

TOPIC REVIEW

CONCLUSION

- Time-series models use previous values to predict future values, also known as forecasting.
- VAR is a multivariate version of the AR process.
- Impulse response functions describe the reaction of shocks on variables over time.
- Granger causality does not imply true causality. It only implies forecasting ability.

CONCLUSION

- The Forecast error variance decomposition (FEVD) indicates the amount of information each variable contributes to the other variables in the autoregression