Getting Started with Time Series Models

Danny Modlin

Senior Analytic Training Consultant

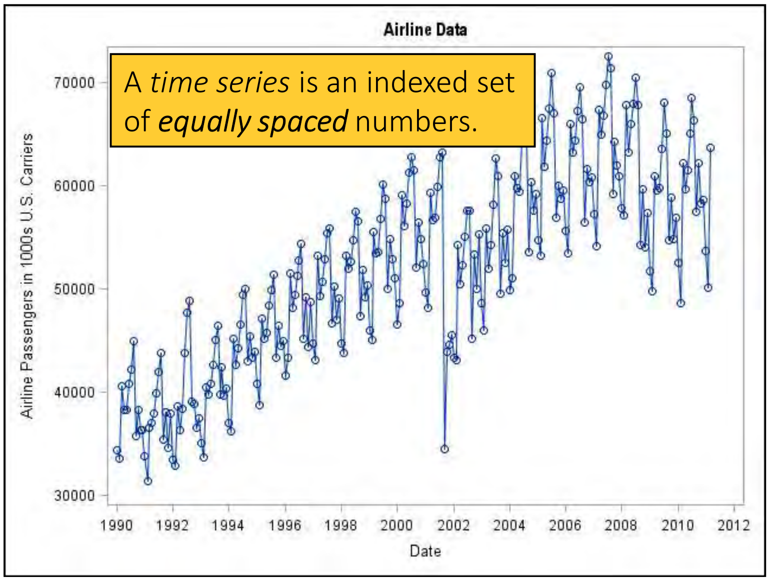
SAS – Cary Offices

Objectives

- Define a time series and the main ideas in time series data creation.
- Define and explore the systematic components in a time series.
- Introduce accuracy statistics to measure forecast accuracy.
- Briefly describe three types of time series models covered in this course.

2

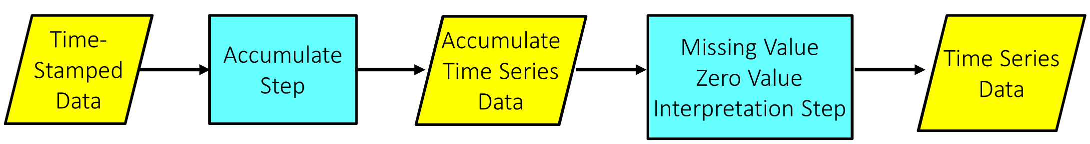
A Time Series



3

Transactional Data Preparation Steps

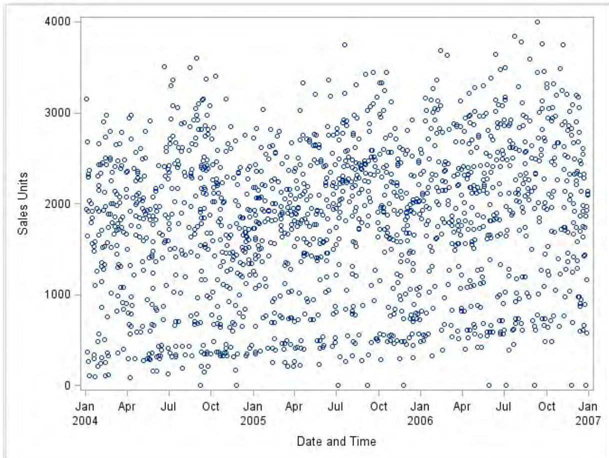
Before you can perform time series analysis and forecasting, the observation (raw) series must be accumulated and interpreted to form a time series. The following diagram summarizes the process of converting timestamped data to time series data:



4

Transactional Data: Example

Historical Timestamped Data

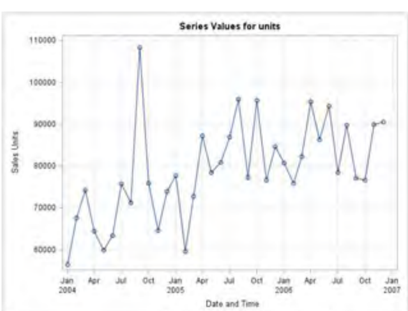


No obvious patterns exist in the transactional data.

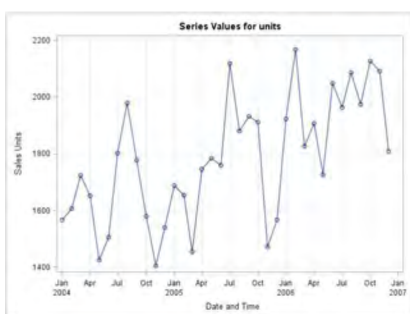
5

Transactional Data Accumulation

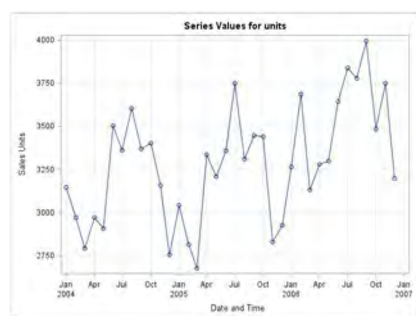
For this timestamped data, accumulating on a monthly *average* or *maximum* basis might be better than accumulating on a *total* basis.



ACCUMULATE=TOTAL



ACCUMULATE=AVERAGE

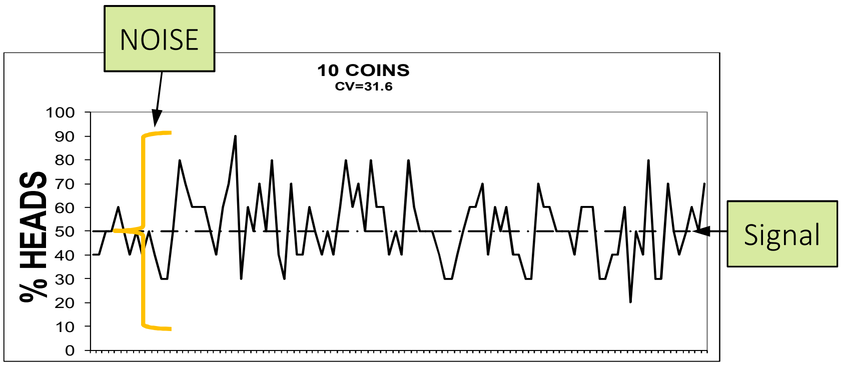


ACCUMULATE=MAXIMUM

6

Variation in Time Series Data: Two Main Parts

- signal
- noise



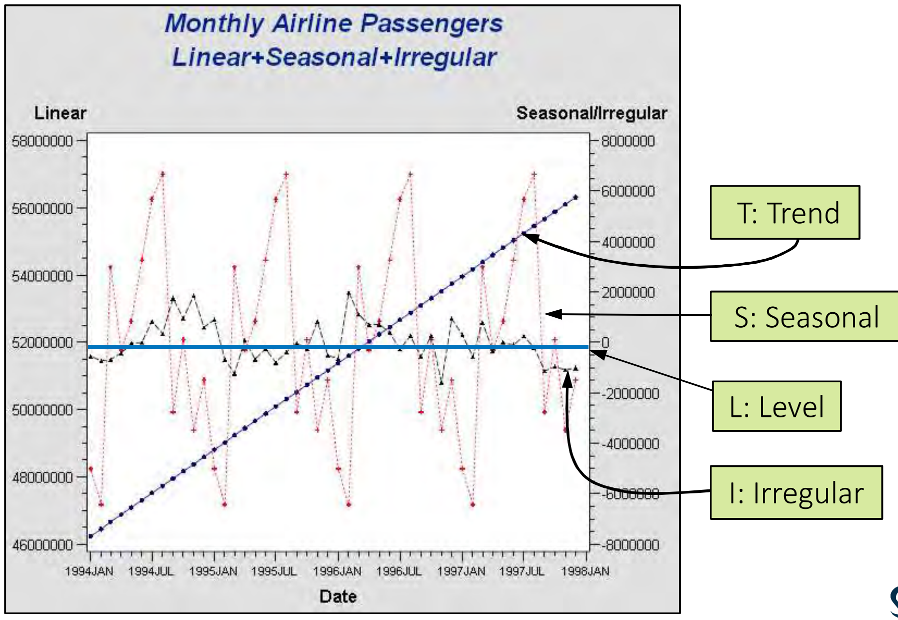
7

Signal Components

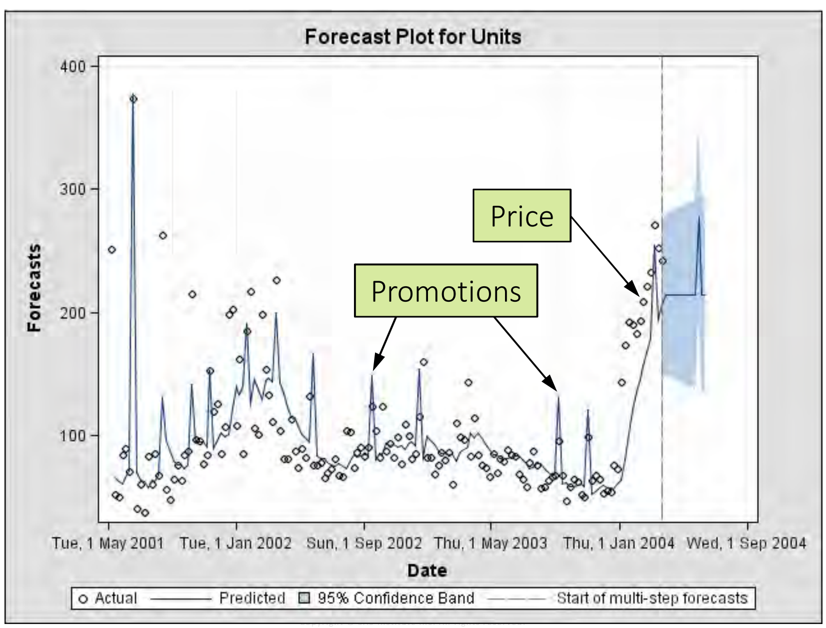
- Level
- Trend
- Seasonality
- Cycle
- Exogenous (also known as *explanatory variable effects*)
- Irregular

8

Component Decomposition



Signal Components: Exogenous Effects (Inputs)



Correlation of Y with Past Y: *Autocorrelation*

Autocorrelation (Order 1):

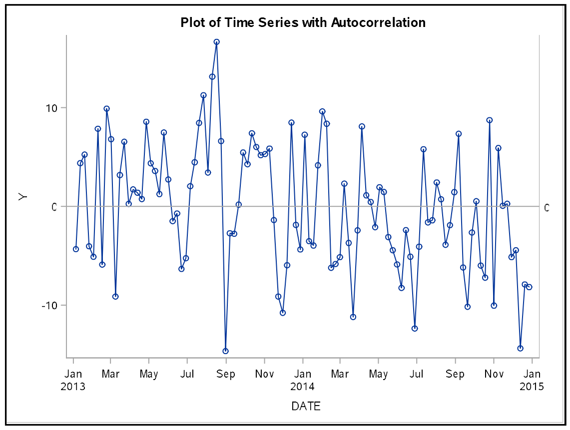
$$Y_t \text{ is correlated with } Y_{t-1}$$

Time Series at Time t:

$$Y_t = Y(t)$$

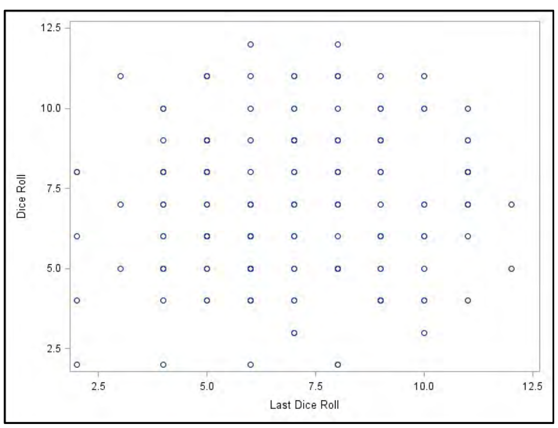
First Lag:

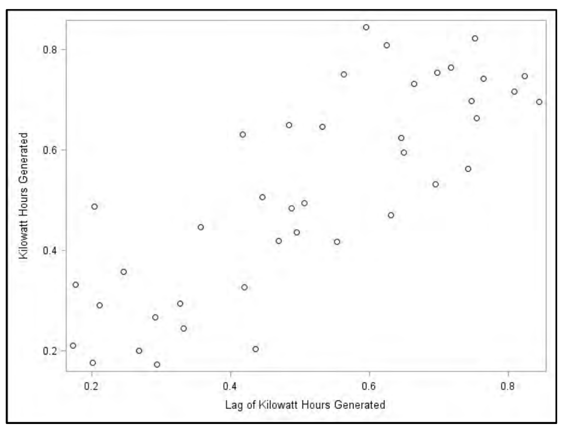
$$Y_{t-1} = Y(t-1)$$



11

Autocorrelation Scatter Plots

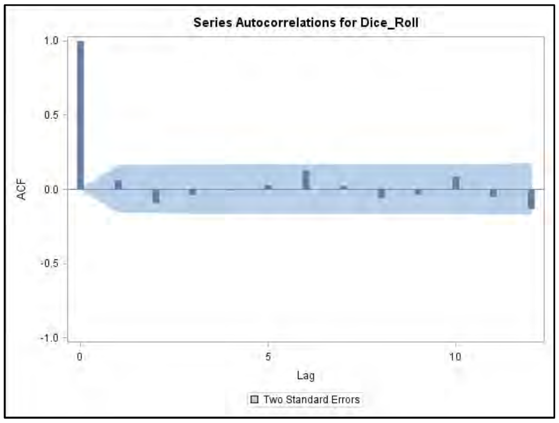


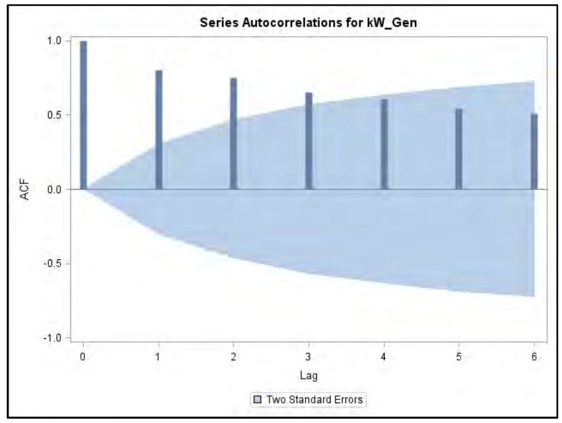


- Autocorrelation is the correlation of present values versus lagged values.
- Autocorrelation between the present value and the first lagged value is called *first order* autocorrelation.

12

Autocorrelation Plots





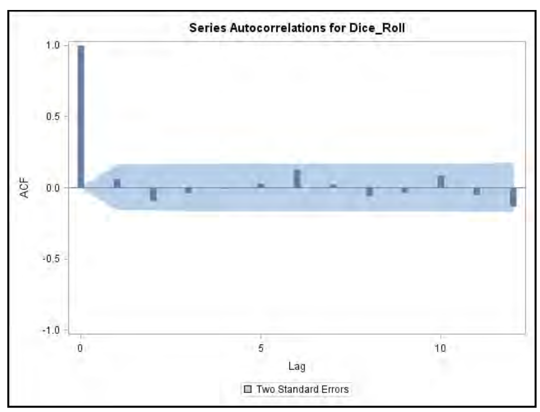
- The autocorrelation plot enables you to see the autocorrelation at multiple lags.
- The blue range indicates 95% confidence intervals for each lag.

13

A White Noise Series

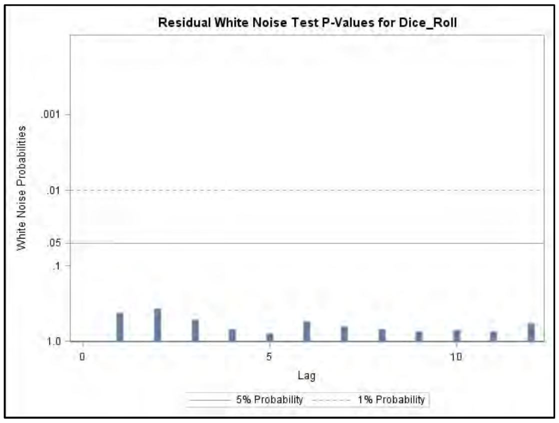
A series that:

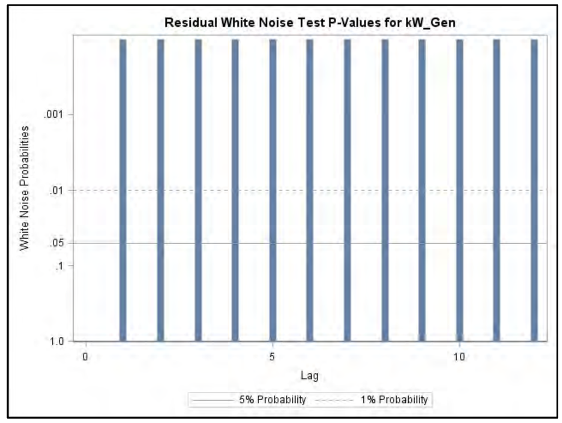
- varies randomly around its mean
- has no systematic variation
- is comprised of only random variation
- has constant variance



14

The Ljung-Box Chi-Square Test for White Noise





“White means white.”

The Three Families of Time Series Models

- exponential smoothing (ESM)
- autoregressive moving average [with exogenous variables] (ARMA[X])
- unobserved components (UCM)

Exponential Smoothing Premise

- Weighted averages of past values can produce good forecasts of the future.
- The weights should emphasize the most recent data.
- Forecasting should require only a few parameters.
- Forecast equations should be simple and easy to implement.

ARMAX Models

ARMAX: AutoRegressive Moving Average with eXogenous variables

- AR: Autoregressive ⇒ Time series is a function of its own past.
- MA: Moving Average ⇒ Time series is a function of past shocks (deviations, innovations, errors, and so on).
- X: Exogenous ⇒ Time series is influenced by external factors.

Unobserved Components Models (UCMs)

- Also known as *structural time series models*
- Decompose time series into components:
  - trend
  - season
  - cycle
  - irregular
  - regressors
- General form:

$$Y_t = \text{Trend} + \text{Season} + \text{Cycle} + \text{Regressors}$$

19

UCMs

- Each component captures some important feature of the series dynamics.
- Components in the model have their own models.
- Each component has its own source of error.
- The coefficients for trend, season, and cycle are dynamic.
- The coefficients are testable.
- Each component has its own forecasts.

20

Comparison of Families

	Usability	Complexity*	Robustness	Able to Acc Dynamic Reg
Best	ESM	ESM	ESM	ARIMAX
----	UCM	ARIMAX	ARIMAX	UCM
Worst	ARIMAX	UCM	UCM	ESM

*Best complexity is the least complex model

Forecasting

If someone asks you whether you can forecast something, your answer should always be "Yes."

If someone asks you whether you can forecast something *accurately*, you cannot answer until you establish what accuracy means and until you perform preliminary diagnostics of the series.

Liability

- You need to assume that the underlying future behavior remains consistent with past behavior.
- However, you have no control over future events that might affect future behavior, such as catastrophes, economic downturns, war, the integrity of key players, the survival of key players, and so on.

Fit and Holdout Samples

	Quarter	*t*	
Ultimate Goal: Forecast the next four quarters.	4Q2018	Y_{t+4}	
	3Q2018	Y_{t+3}	
	2Q2018	Y_{t+2}	
	1Q2018	Y_{t+1}	
How well can you forecast these four most recent observed quarters?	4Q2017	Y_t	Holdout Sample
	3Q2017	Y_{t-1}	
	2Q2017	Y_{t-2}	
	1Q2017	Y_{t-3}	
Forecasting observed values with the remaining observed series	4Q2016	Y_{t-4}	Fit Sample
	…	…	

Honest Assessment: Simulating a Prospective Study

1) Divide the time series data into two segments.
   1) The *fit sample* is used to derive a forecast model.
   2) The *holdout sample* is used to evaluate forecast accuracy.
2) Derive a set of candidate models.
   - Exponential Smoothing Models
   - ARMA and ARMAX Models
   - Unobserved Components Models
3) Calculate the chosen model accuracy statistic for each model by using the fit sample to forecast the holdout sample.
4) Choose the model with the best accuracy statistic.
5) Using the best model, generate forecasts for *n* future periods

25

Summary of Data Used for Forecast Model Building

Fit Sample

- **Used to estimate model parameters for accuracy evaluation**
- **Used to forecast values in holdout sample**

Holdout Sample

- **Used to evaluate model accuracy**
- **Simulates retrospective study**

Full = Fit + Holdout data is used to fit a deployment model.

26

Necessary Conditions for Good Forecasts

- The identified signal continues into the future.
- Forecasting model complexity should be adequate to capture signal components.
- Forecasting models should not be overly complex.
- The best forecasting model is the one that captures and extrapolates the most signal, and that also ignores the noise.

27

Choosing a Winning Set of Forecasts

Good forecasts should

- be highly correlated with the actual series values
- exhibit small forecast errors
- capture the prominent features of the original time series.

In addition, assessment of forecast quality should be based on the business, engineering, or scientific problem that is being addressed.

28

Ready to learn more?

Visit the Learn Area in the Quad for the latest training and books.

SAS Training Course of Interest

Post-Conference training

- Time Series Forecasting in SAS University Edition: ARIMA Models

Public training course

- Time Series Modeling Essentials

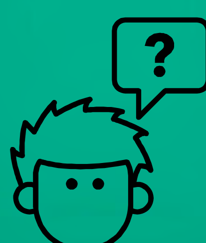
Get 20% off this book

Use code: SASGF

- SAS for Forecasting Time Series

29

Questions?



30