Statistical Forecasting Workshop

PhredTek, Inc.
logistics consulting services

Workshop Contents
Day 1

- Introduction to Business Forecasting
- Introduction to Time Series, Simple Averages, Moving Averages and Exponential Smoothing
- Regression Models for Forecasting
- Forecasting Accuracy
- Putting it all Together – The Forecasting Process

Workshop Contents Day 2

- Basic statistics review
- Autocorrelation and Partial Autocorrelation
- The family of ARIMA models
  - Autoregressive
  - Moving average
  - Differencing
  - Seasonal Autoregressive
  - Seasonal moving average
  - Seasonal differencing
  - Transfer functions
- The general ARIMA model
- The modeling process

Workshop Contents
Day 2 (continued)

- Example of the modeling process– General Mills biscuits
  - Univariate model
  - Regression model
  - ARIMA transfer function model
  - Modeling outliers
- Another example – daily sales of a beer brand and package at one grocery store

PhredTek, Inc.
logistics consulting services

Workshop Objectives

- This workshop is designed to provide you with an understanding and conceptual basis for some basic time series forecasting models.
- We will demonstrate these models using MS Excel.
- You will be able to make forecasts of business / market variables with greater accuracy than you may be experiencing now.

1. Introduction to Business Forecasting

Why Forecast?

- ***Make best use of resources by***
  - Developing future plans
  - Planning cash flow to prevent shortfalls
  - Achieving competitive delivery times
  - Supporting business financial objectives
  - Reducing uncertainty

PhredTek, Inc.
logistics consulting services

What Is Riding On The Forecast?

- Allocation of resources:
  - Investments
  - Capital
  - Inventory
  - Capacity
  - Operations budgets
  - Marketing budget
  - Manpower and hiring

PhredTek, Inc.
logistics consulting services

Who Needs The Forecast?

- The people who allocate resources
  - Marketing
  - Sales
  - Finance
  - Production
  - Supply chain

What Level Do We Forecast?

- Geographic level
  - Shipments to distributors (minimum)
  - Shipments to retailers (better)
  - Sales to Consumer (best)

Forecast as close to the consumer as your data will allow

PhredTek, Inc.
logistics consulting services

What Level Do We Forecast?

- Time Frame
  - Annual
  - Quarterly
  - Monthly – solar or lunar
  - Weekly
  - Daily
  - Hourly

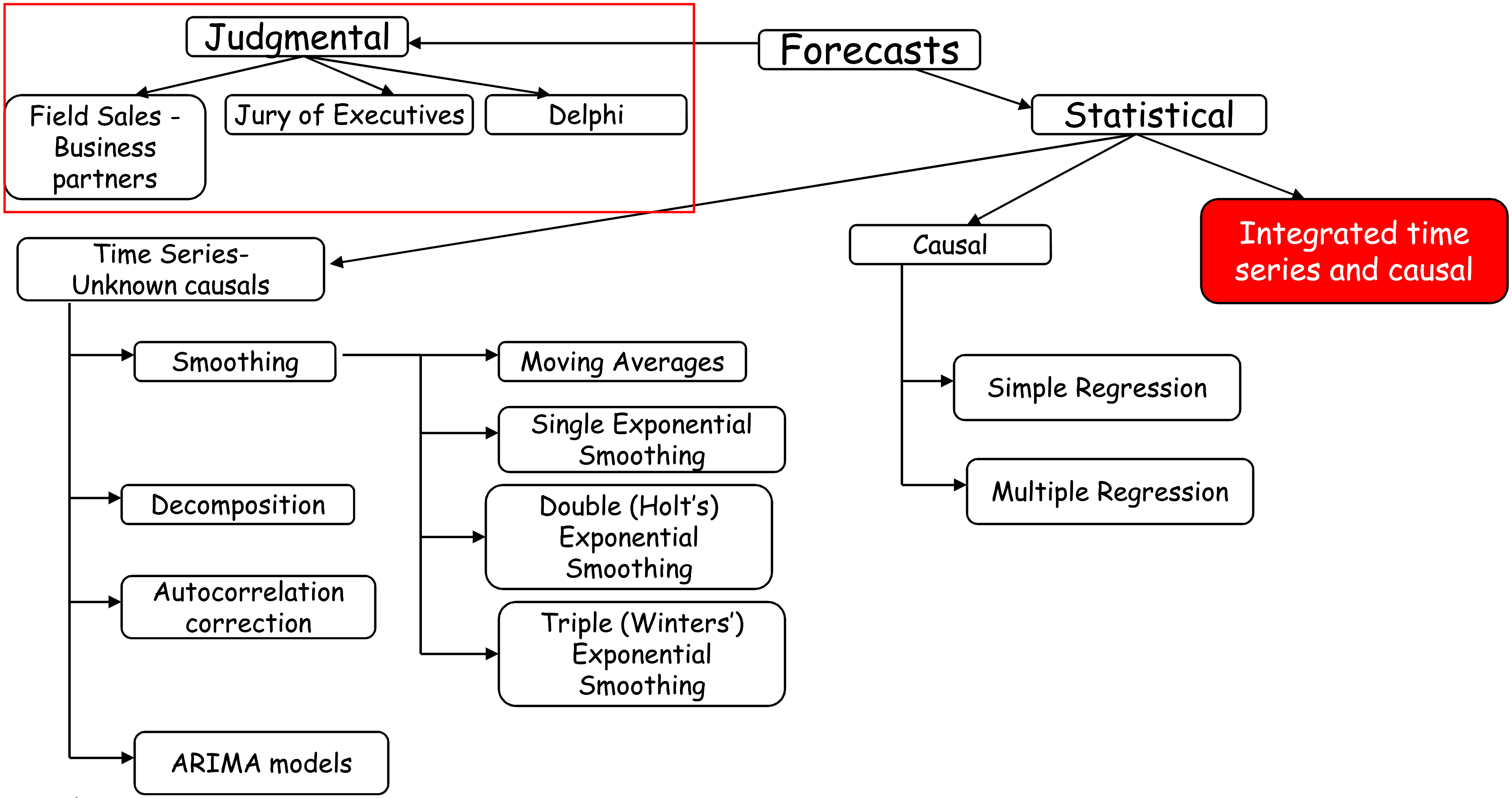
It is generally best to forecast the most detailed time frame your data will allow. Aggregate forecasts to the time frame in which resources are allocated.

What Do We Know About Forecasts?

- Forecasts are always wrong
- Error estimates for forecasts are imperative.
- Good forecasts require a firm knowledge of the process being forecast.
- It is essential that there be a business process in place to use the forecast. Otherwise it will never be used.
- Forecasts are more accurate for larger groups of items
- Forecasts are more accurate for longer time periods. E.g. annual forecasts are more accurate than monthly, which are more accurate than weekly, and so on.
- Forecasts are generally more accurate for fewer future time periods

PhredTek, Inc.
logistics consulting services

Overview of Forecasting Methods



PhredTek, Inc.

Judgmental Forecasts
Sales Force and Business Partners

- These people often understand nuances of the marketplace that effect sales.
- Are motivated to produce good forecasts because it effects their workload.
- Can be biased for personal gain.
- An effective alternative is to provide them with a statistical based forecast and ask them to adjust it.

Judgmental Forecasts
Jury of Executive Opinion

- An executive generally knows more about the business than a forecaster.
- Executives are strongly motivated to produce good forecasts because their performance is dependent on them.
- However, executives can also be biased for personal gain.
- An alternative is to provide them with a statistical forecast and ask them to adjust it.

PhredTek, Inc.
logistics consulting services

Judgmental Forecasts
Delphi Method

Six Steps:

1. Participating panel members are selected.
2. Questionnaires asking for opinions about the variables to be forecast are distributed to panel members.
3. Results from panel members are collected, tabulated, and summarized.
4. Summary results are distributed to the panel members for their review and consideration.
5. Panel members revise their individual estimates, taking account of the information received from the other, unknown panel members.
6. Steps 3 through 5 are repeated until no significant changes result.

PhredTek, Inc.
logistics consulting services

Time Series Models

- Naïve Forecast
  - Tomorrow will be the same as today
- Moving Averaqe
  - Unweighted Linear Combination of Past Actual Values
- Exponential Smoothing
  - Weighted Linear Combination of Past Actual Values
- Decomposition
  - Break time series into trend, seasonality, and randomness.

PhredTek, Inc.
logistics consulting services

Causal/Explanatory Models

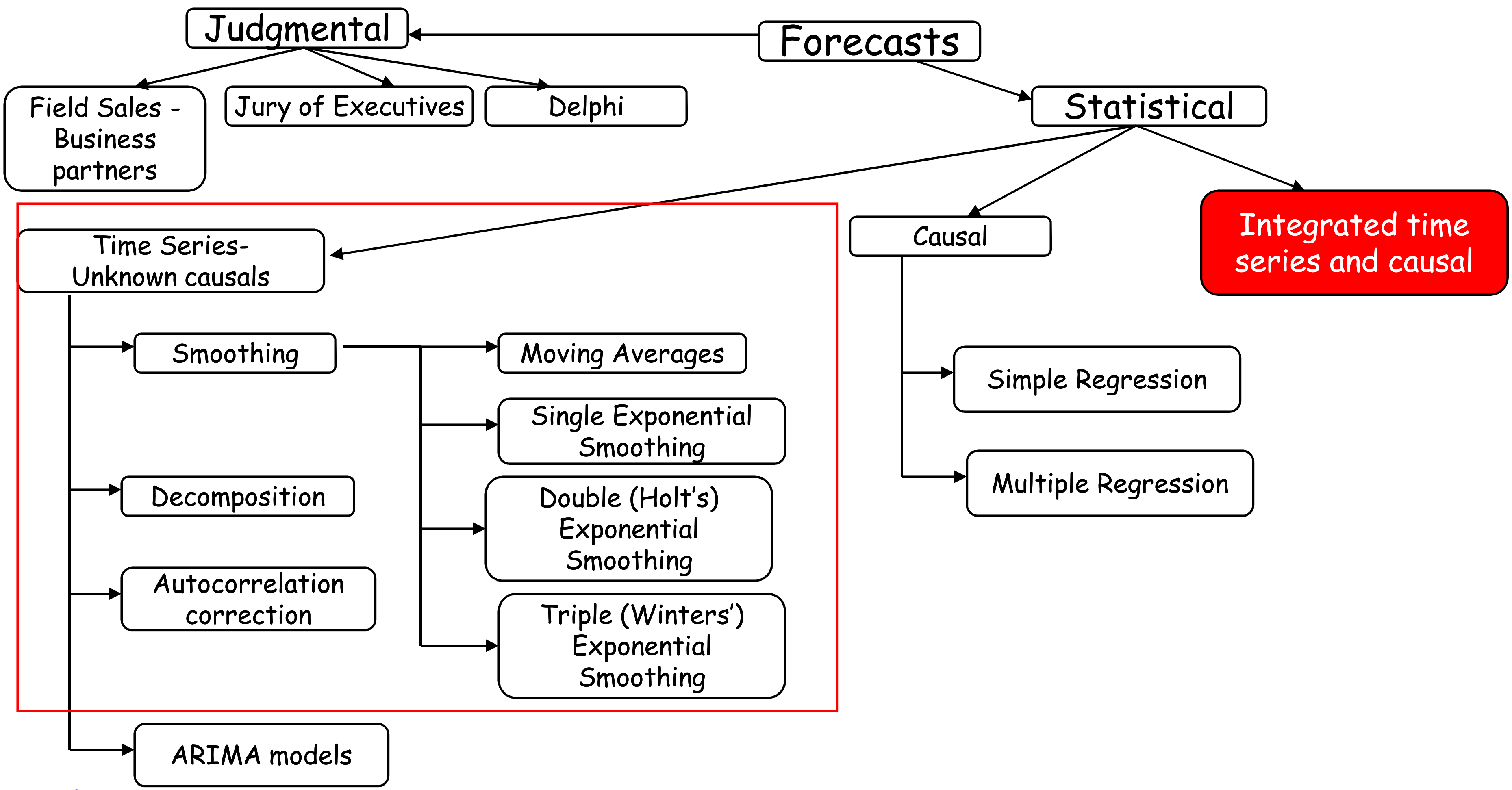
- Simple Regression:
  – Variations in dependent variable is explained by one independent variable.

- Multiple Regression:
  – Variations in dependent variable is explained by multiple independent variables.

PhredTek, Inc.
logistics consulting services

2. Time Series Methods

Overview of Forecasting Methods



PhredTek, Inc.
logistics consulting services

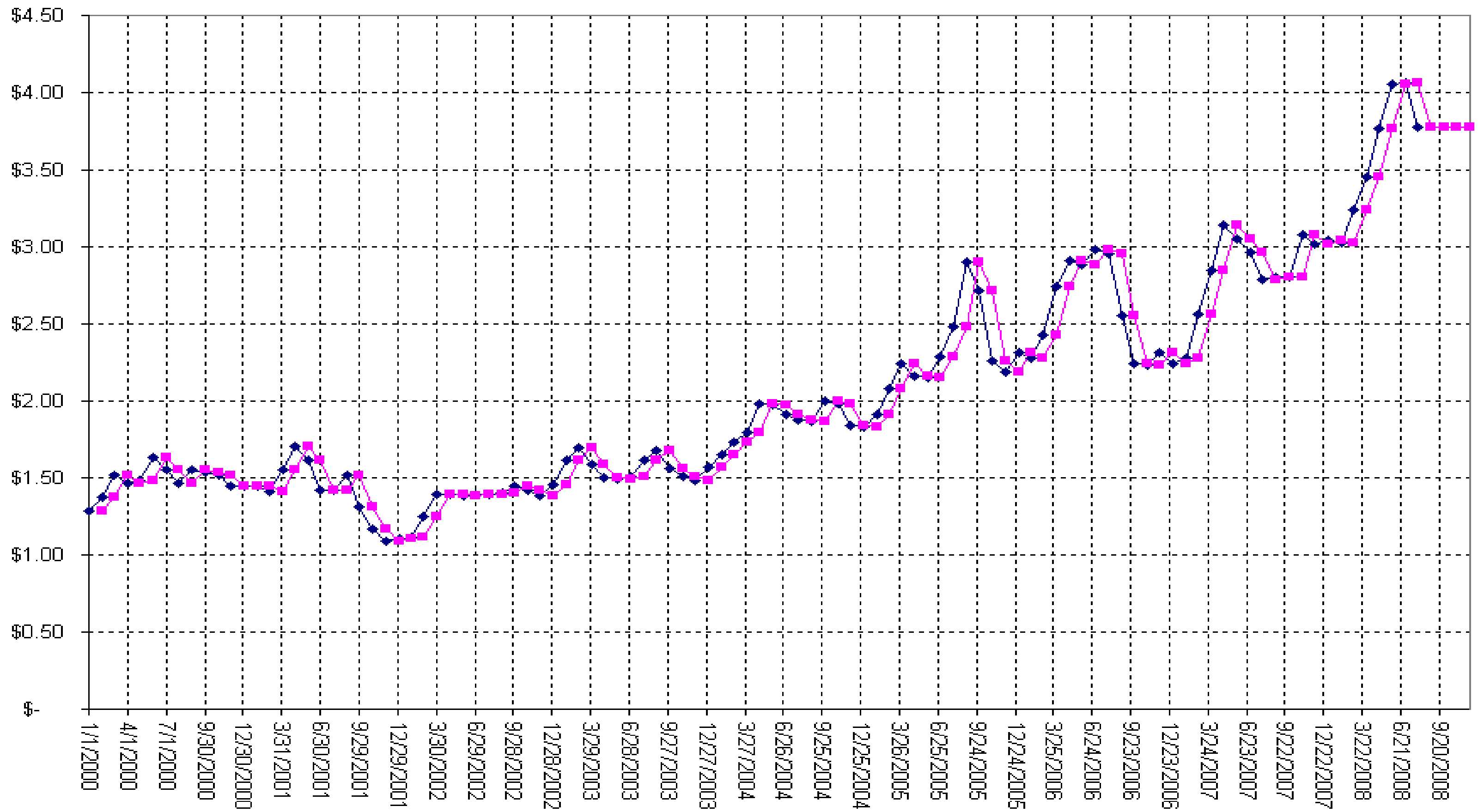
Naïve Model

- Special Case of Single Exponential Smoothing
- Forecast Value is equal to the Previously Observed Value
  - Stable Environment
  - Slow Rate of Change (if any)

PhredTek, Inc.
logistics consulting services

Naïve Model

Naïve Forecast of Monthly Gasoline Prices



Why Use The Naïve Model?

- It's safe. It will never forecast a value that hasn't happened before.
- It is useful for comparing the quality of other forecasting models. If forecast error of another method is higher than the naïve model, it's not very good.

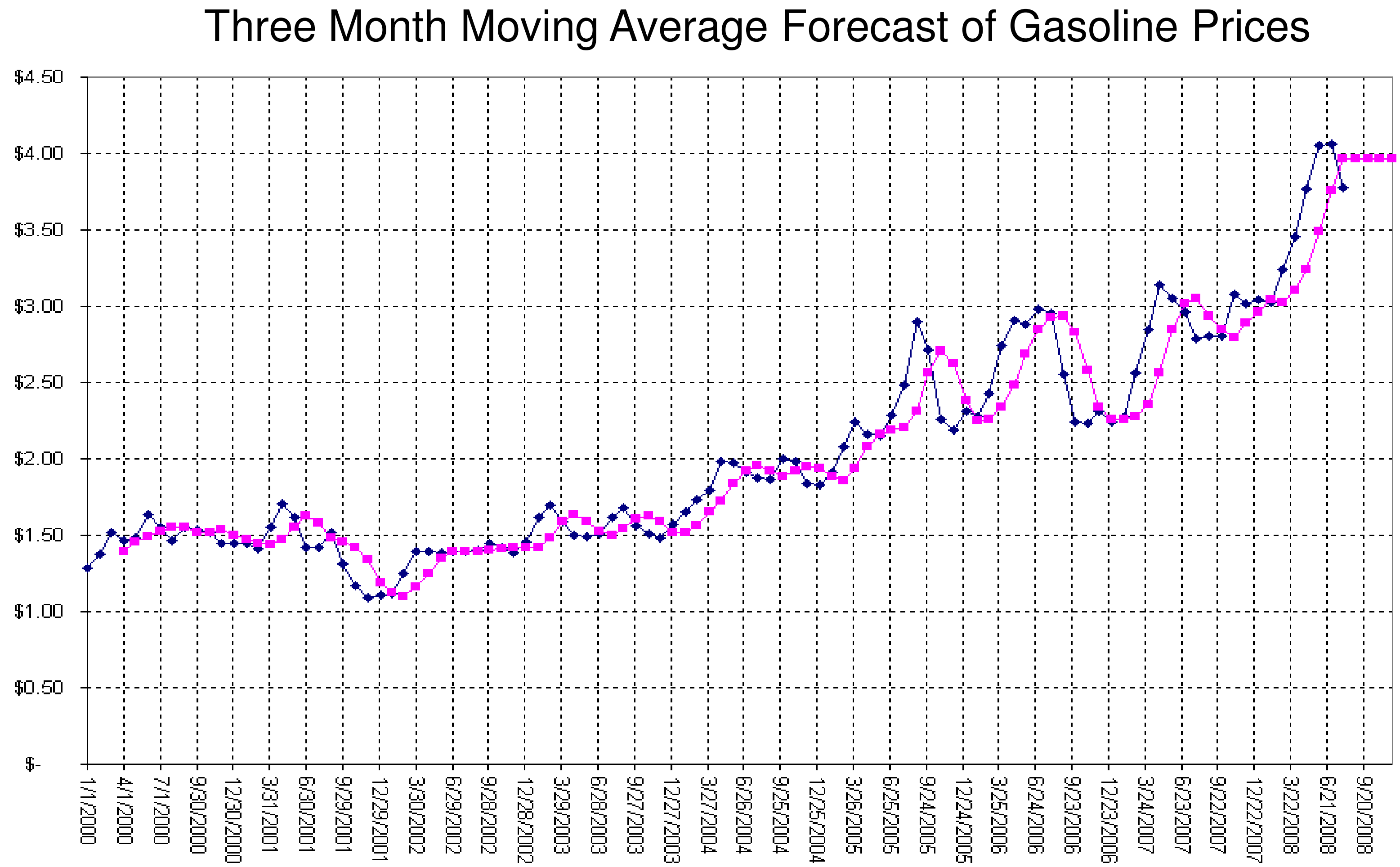
PhredTek, Inc.
logistics consulting services

Moving Average Model

- Easy to Calculate
  - Select Number of Periods
  - Apply to Actual
- Assimilates Actual Experience
- Absorbs Recent Change
- Smoothes Forecast in Face of Random Variation
- Safe – never forecasts outside historical values.

PhredTek, Inc.
logistics consulting services

Moving Average Model



Phredlek, Inc.
logistics consulting services

Time Series Methods – Exponential Smoothing

- Single Exponential Smoothing
- Double - Holt's Exponential Smoothing
- Winters' Exponential Smoothing

Exponential Smoothing

- Widely Used
- Easy to Calculate
- Limited Data Required
- Assumes Random Variation Around a Stable Level
- Expandable to Trend Model and to Seasonal Model

Smoothing Parameters

- Level (Randomness) – Simple Model
  - Assumes variation around a level
- Trend – Holt's Model
  - Assumes linear trend in data
- Seasonality – Winter's Model
  - Assumes recurring pattern in periodicity due to seasonal factors

PhredTek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

- Single Exponential Smoothing
  - $F_{t+1}=\alpha A_t+(1-\alpha)F_t$
  - Where
    - F_{t+1} = forecasted value for next period
    - α = the smoothing constant ($0\leq\alpha\leq1$)
    - A_t = actual value of time series now (in period t)
    - F_t = forecasted value for time period t

- Moving Averages give equal weight to past values, Smoothing gives more weight to recent observations.

PhredTek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

- Weights for alpha = .1

Time	Calculation	Weight for A_t
t		0.100
t-1	0.9 * 0.1	0.090
t-2	0.9 * 0.9 * 0.1	0.081
t-3	0.9 * 0.9 * 0.9 * 0.1	0.073
.		
.	α=.1	
.		
		1.000

- Moving Averages give equal weight to past values, Smoothing gives more weight to recent observations.

PhredTek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

- Weights for alpha = .9

Time	Calculation	Weight for A_t
t		0.9000
t-1	0.1 * 0.9	0.0900
t-2	0.1 * 0.1 * 0.9	0.0090
t-3	0.1 * 0.1 * 0.1 * 0.9	0.0009
.		
.	**α=.9**	
.		
		1.0000

- Moving Averages give equal weight to past values, Smoothing gives more weight to recent observations.

PhredTek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

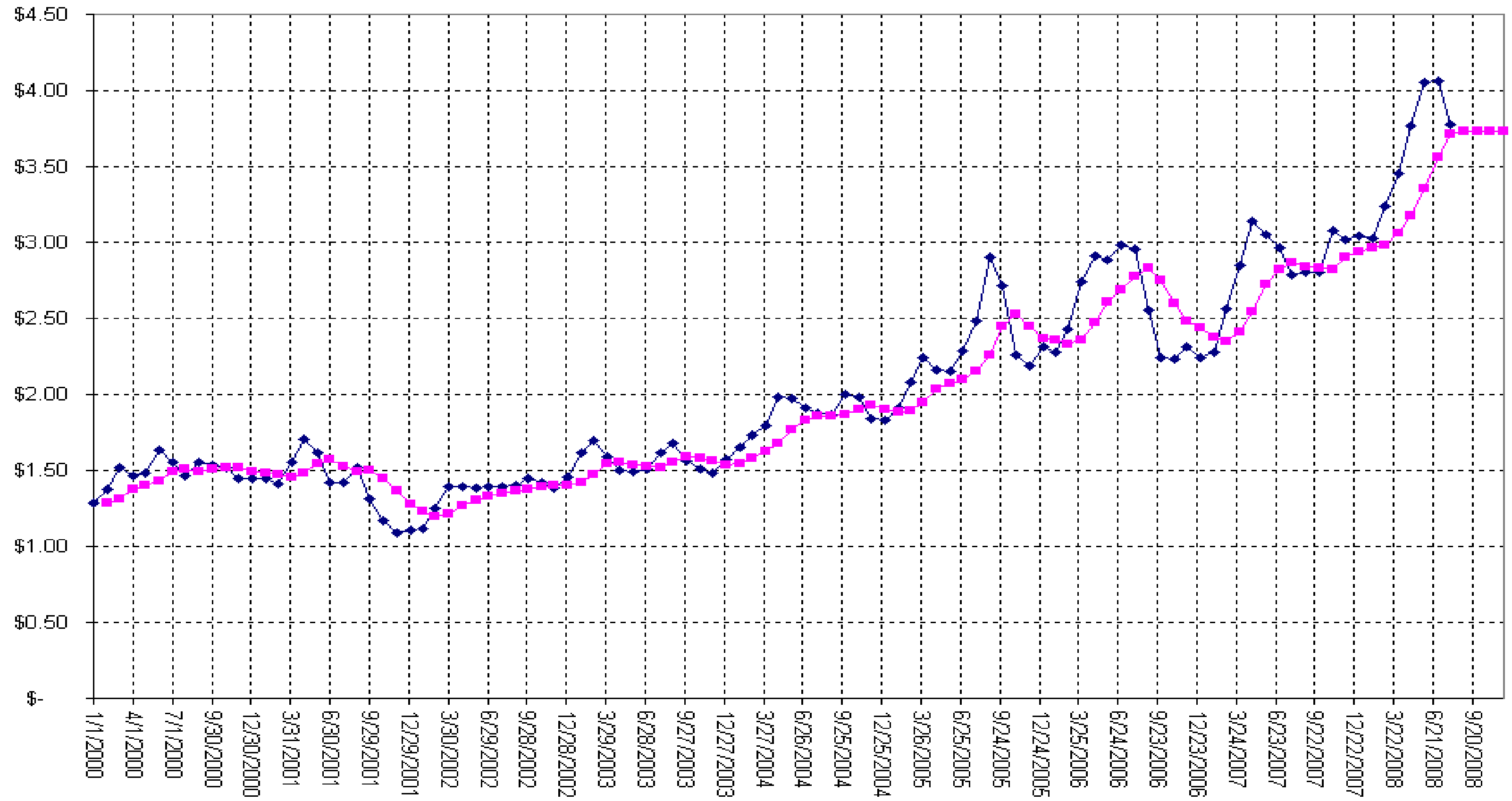
- The Single Exponential Smoothing Rule of Thumb
- The closer to 1 the value of alpha, the more strongly the forecast depends upon recent values.

In actual practice, **alpha values from 0.05 to 0.30 work very well** in most Single smoothing models. If a value of greater than 0.30 gives the best fit this usually indicates that another forecasting technique would work even better.

PhredTek, Inc.
logistics consulting services

Exponential Smoothing Model

Exponential Smoothing Forecast of Gasoline Prices – alpha = .3



Phredlek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

- Holt's Exponential Smoothing
  - $F_{t+1}=\alpha A_t+(1-\alpha)(F_t+T_t)$
  - $T_{t+1}=\beta(F_{t+1}-F_t)+(1-\beta)T_t$
  - $F_{t+m}=F_{t+1}+mT_{t+1}$
- Where
  - F_{t+1} = forecasted value for next period
  - α = the smoothing constant ($0\leq\alpha\leq1$)
  - A_t = actual value of time series now (in period t)
  - F_t = forecasted value for time period t
  - Tt+1= trend value for next period
  - Tt = actual value of trend now (in period t)
  - β = the trend smoothing constant
  - m = number of periods into the future to forecast from the last actual level and trend values

Phredtek, Inc.
logistics consulting services

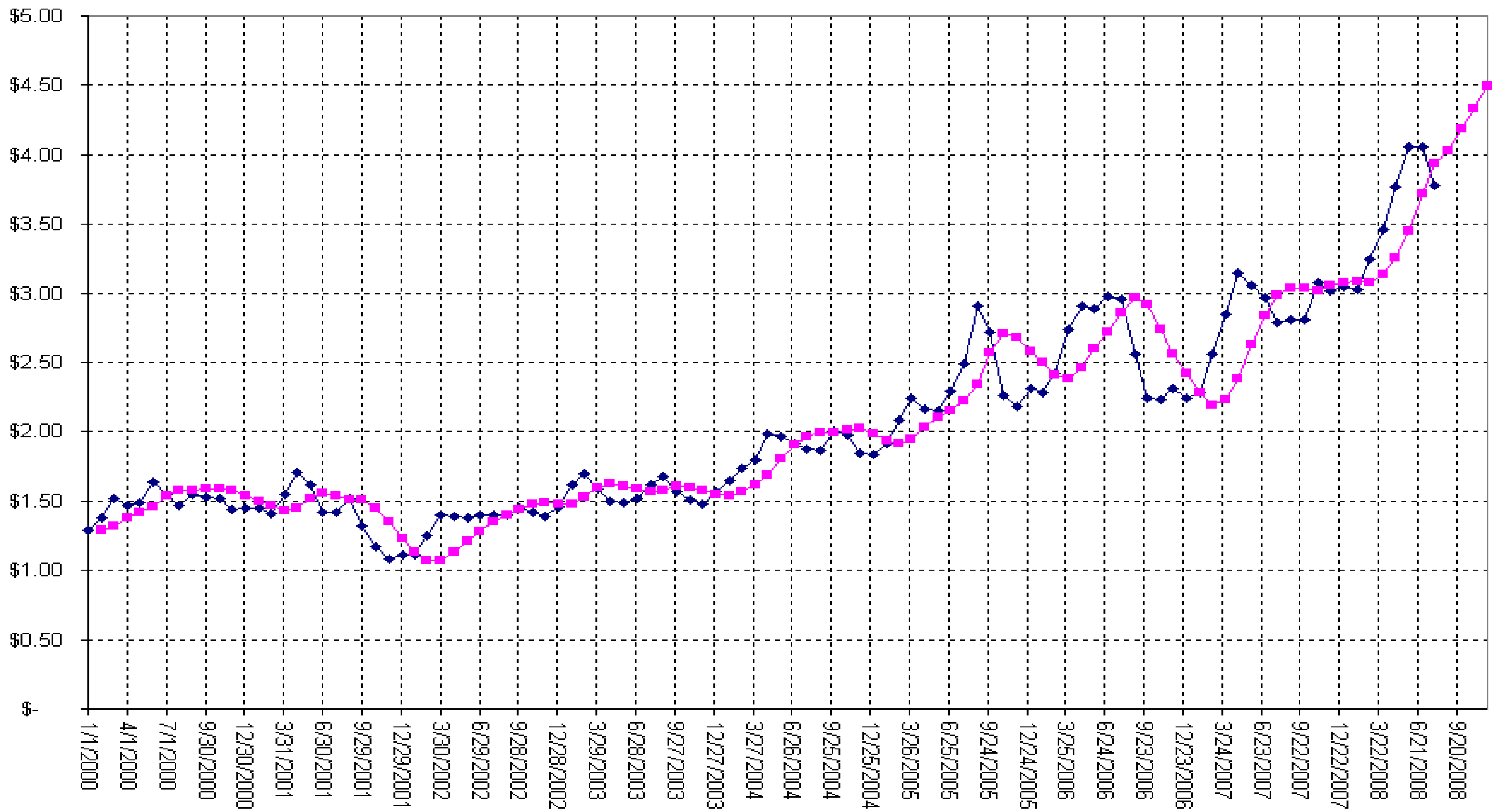
Time Series Methods
Exponential Smoothing

- Holt's Exponential Smoothing
  - Used for data exhibiting a trend over time (±)
  - Data display non-seasonal pattern
  - Involves two smoothing factors (constants), a single smoothing factor and trend smoothing factor

PhredTek, Inc.
logistics consulting services

Holt’s Double Exponential Smoothing Model

Forecast of Gasoline Prices – alpha = .3 – beta = .4



Phredlek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

- Winters' Exponential Smoothing
  - Adjusts for both trend and seasonality
  - Even more complex calculations but also simple to apply using software
  - Involves the use of three smoothing parameters, a single smoothing parameter, a trend smoothing parameter, and a seasonality smoothing parameter

PhredTek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

- Winters' Exponential Smoothing
  - $F_t = \alpha(A_t/S_{t-p}) + (1-\alpha)(F_{t-1} + T_{t-1})$
  - $S_t = \beta(A_t/F_t) + (1-\beta)S_{t-p}$
  - $T_t = \gamma(F_t - F_{t-1}) + (1-\gamma)T_{t-1}$
  - $WF_{t+m} = (F_t + mT_t)S_{t+m-p}$

PhredTek, Inc.
logistics consulting services

Time Series Methods Exponential Smoothing

Pros

- Requires a limited amount of data
- Relatively Simple compared to other forecasting methods
- Expandable to Trend Model and to Seasonal Model

Cons

- Cannot include outside causal factors
- Rarely corrects for the actual autocorrelation of the series.

PhredTek, Inc. logistics consulting services

Time Series Decomposition

- This is classical approach to economic / time series forecasting.
- It assumes that an economic time series can be decomposed into four components:
  - Trend
  - Seasonal variation
  - Cyclical variation
  - Random variation
- Originated at the US Bureau of the Census in the 1950s

Time Series Decomposition

- For example:
- The value of a company's sales could be viewed as:
- Y=T*S*C*I (multiplicative)
  - where:
    - Y=sales
    - T=trend
    - S=seasonal variation
    - C=cyclical variation
    - I=irregular component

PhredTek, Inc.
logistics consulting services

Classical Decomposition Model

- A different approach to forecasting seasonal data series
  - Calculate the seasonals for the series
  - De-seasonalize the raw data
  - Apply the forecasting method
  - Re-sesonalize the series

PhredTek, Inc.
logistics consulting services

Classical Decomposition Model Practical Application

- US Bureau of the Census developed X11 procedure during the 1950s.
- This is the most widely used method.
- USBOC developed Fortran code to implement. – Still available if you search hard.
- SAS has a PROC X11.
- Only works well for very stable processes that don't change much over time.

PhredTek, Inc.
logistics consulting services

Summary of Time Series Forecasting

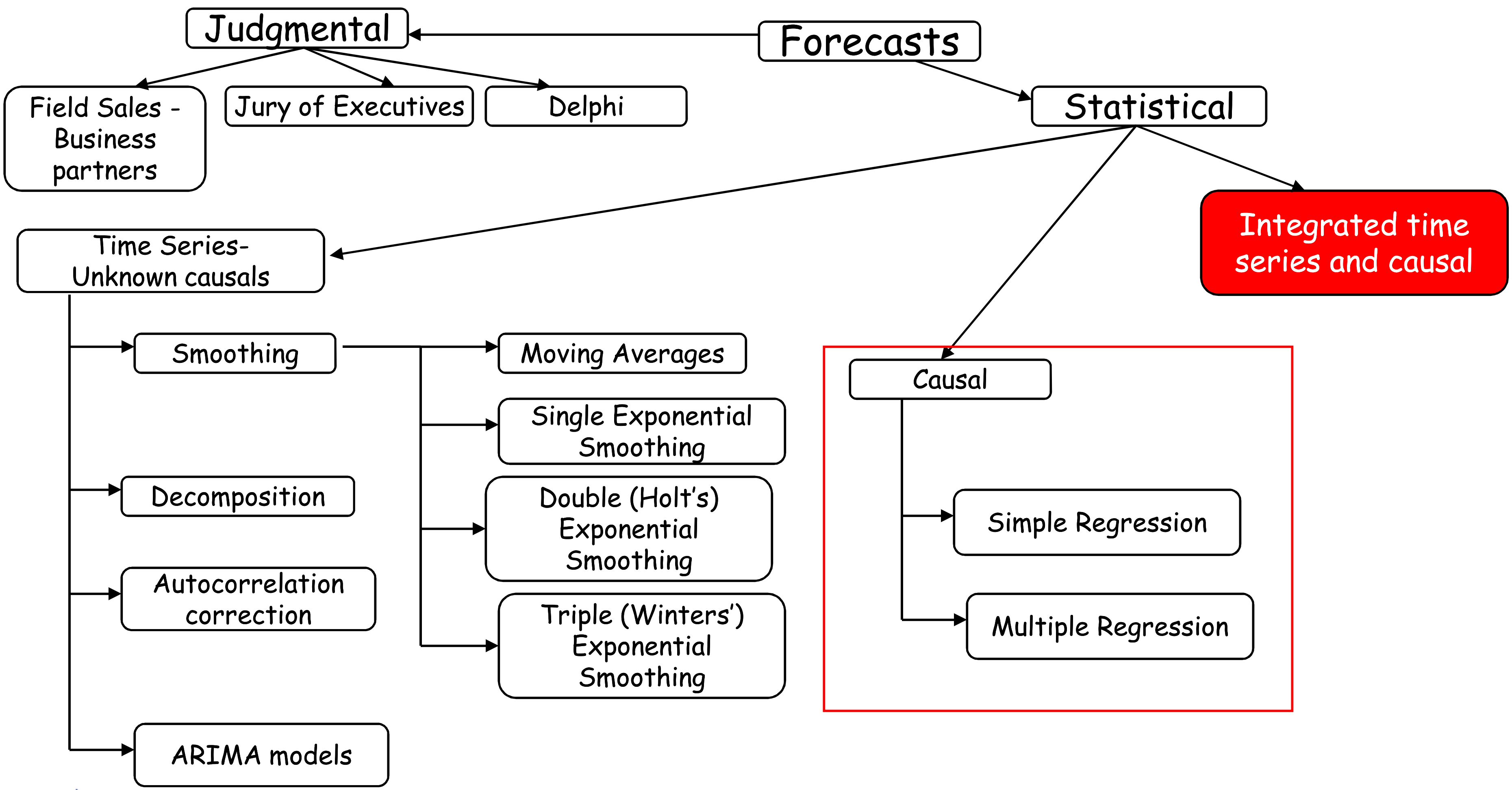
- Should be used when a limited amount of data is available on external factors (factors other than actual history), e.g., price changes, economic activity, etc.
- Useful when trend rates, seasonal patterns, and level changes are present
- Otherwise, Causal/Explanatory techniques may be more appropriate.

PhredTek, Inc.
logistics consulting services

Excel Exercise

3. Cause-and-Effect Models
Regression Forecasting

Overview of Forecasting Methods



PhredTek, Inc.
logistics consulting services

Regression Models

- A plethora of software packages are available - but there is a danger in using canned packages unless you are familiar with the underlying concepts.
- Today's software packages are easy to use, but **<u>learn the underlying concepts.</u>**

PhredTek, Inc.
logistics consulting services

Simple Regression

$$Y = a + bX + e$$

- Y = Dependent Variable
- X = Independent Variable
- a = Intercept of the line
- b = Slope of the line
- e = Residual or error

The Intercept and Slope

- The intercept (or "constant term") indicates where the regression line intercepts the vertical axis.
- The slope indicates how Y changes as X changes (e.g., if the slope is positive, as X increases, Y also increases -- if the slope is negative, as X increases, Y decreases).

PhredTek, Inc.
logistics consulting services

Simple Regression Forecasting

- Four steps in regression modeling:

1. Specification
2. Estimation
3. Validation
4. Forecasting

PhredTek, Inc.
logistics consulting services

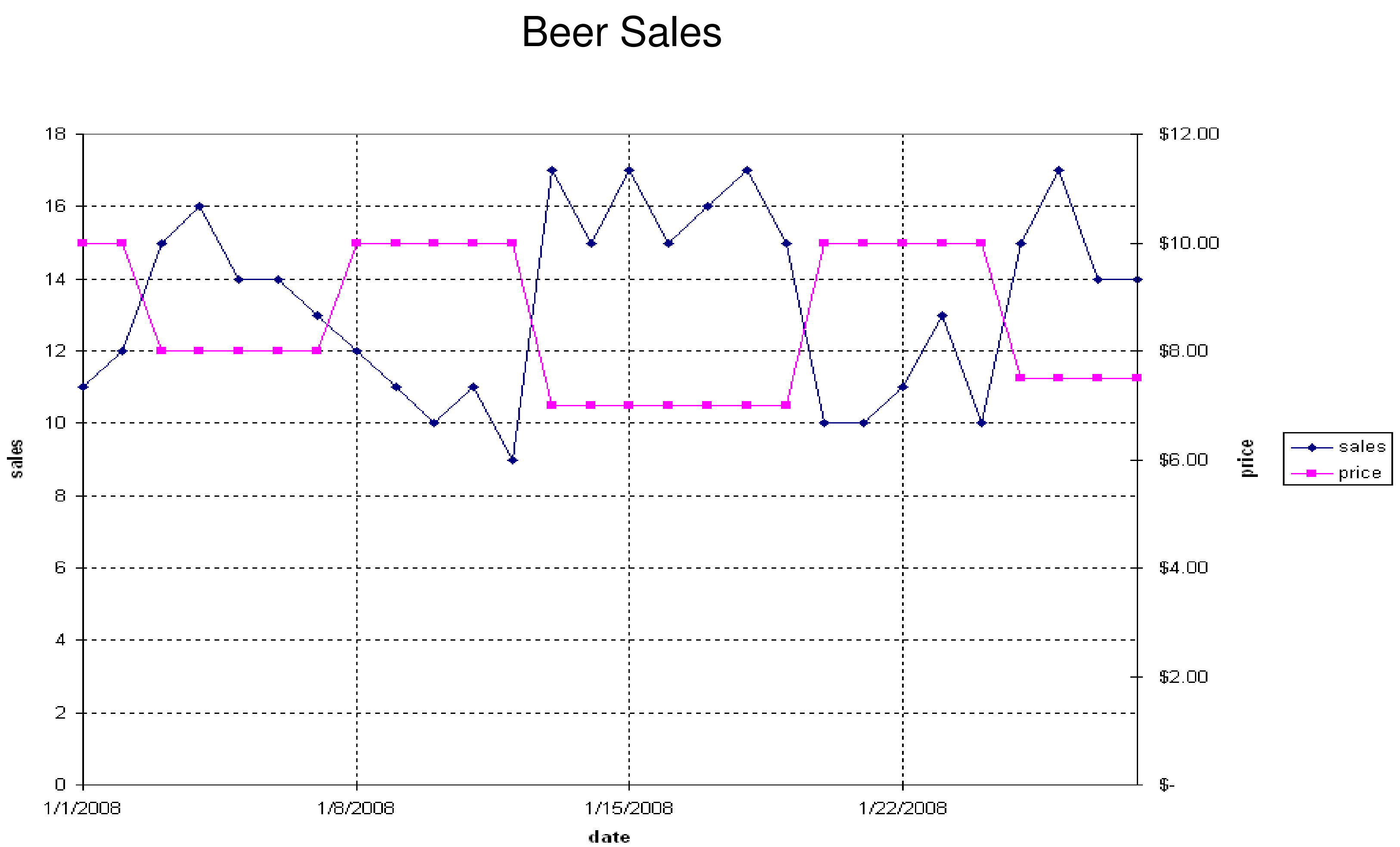
Simple Regression Forecasting
1) Specification

- Determine Variables:
  - Dependent variable = Y (e.g. sales)
  - Independent variable = X (e.g. price or trend)
- Make sure there is a business reason why X effects Y.

Simple Regression Forecasting 2) Estimation

- Set up data in spreadsheet or other software package.
- Software finds the parameters (intercept and slope) which minimizes the sum of squares of the residual – it's that simple – no magic.
- Software also produces summary statistics for validation.

Simple Regression Example



Phredlek, Inc.
logistics consulting services

Simple Regression Example

Specification

Sales = a + b * price

Estimation

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.905603							
R Square	0.820116							
Adjusted R	0.813197							
Standard E	1.09223							
Observatio	28							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regressior	1	141.4114	141.4114	118.5377	3.51E-11			
Residual	26	31.01713	1.192966					
Total	27	172.4286						
	Coefficients	*standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	28.0134	1.361888	20.56954	1.31E-17	25.214	30.8128	25.214	30.8128
price	-1.71705	0.157709	-10.8875	3.51E-11	-2.04123	-1.39288	-2.04123	-1.39288

Result

Sales = 28.0 – 1.72 * price

PhredTek, Inc.
logistics consulting services

Simple Regression Forecasting
3) Validation

- T test
- R square
- Standard Error of the residuals
- Sign of the coefficients
- F test
- Autocorrelation

PhredTek, Inc.
logistics consulting services

Simple Regression Forecasting
3) Validation - T Test

- Tests whether or not the slope is really different from 0.
- T value is the ratio of the slope to its standard deviation – that is, it is the number of standard deviations away from 0.
- P value is the probability of getting that slope if the true slope is 0.
- Generally accepted "good" P value is .05 or less.

Simple Regression Forecasting 3) Validation - R Square

- Measures the fraction of the variability in the dependent variable that is explained by the independent variable.
- Ranges between 0 and 1
  - 1 means all the variability is explained.
  - 0 means none of the variability is explained.
- Adjusted R square – an attempt by statisticians to purify the raw R square that I have never found useful.
- R squares > .9 are very good for forecasting.
- However, R squares as low as .5 can still be useful.

PhredTek, Inc.
logistics consulting services

Simple Regression Forecasting 3) Validation – Standard Error of the Residuals

- Gives a good estimate of future forecast error.
- Can be used to determine. confidence limits on the forecast.
- Can be used to set safety stock.

PhredTek, Inc.
logistics consulting services

Simple Regression Forecasting 3) Validation - F test

- Tests whether both the slope and the intercept are simultaneously greater than 0.
- P value of .05 or less again is generally accepted as significant.
- Passing the F test alone does not say the regression is useful.
- I have never found this test useful.

PhredTek, Inc.
logistics consulting services

Simple Regression Forecasting 3) Validation - Autocorrelation

- Autocorrelation results in a pattern in the residuals.
- Can be determined two ways:
  - Visually from a graph (easy)
  - A Durbin-Watson statistic < 1.5 or > 2.5
- Sometimes autocorrelation can be corrected by adding variables (multiple regression)
- Most of the time, autocorrelation cannot be corrected without advanced techniques – we will discuss this later.

Simple Regression Forecasting
4) Doing the Forecast

- Tabulate the values of the independent variable for all time periods to be forecast.
- Apply the regression formula to the values.

The forecast for our example

date	forecast	price
1/29/2008	10.84	$ 10.00
1/30/2008	10.84	$ 10.00
1/31/2008	10.84	$ 10.00
2/1/2008	17.71	$ 6.00
2/2/2008	17.71	$ 6.00
2/3/2008	17.71	$ 6.00
2/4/2008	17.71	$ 6.00

PhredTek, Inc.
logistics consulting services

The Danger of Outliers

Beer sales simulated for 1000 days

- Coefficients about the same
- R square and t Stats are up.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.949561							
R Square	0.901667							
Adjusted R	0.901568							
Standard E	0.985718							
Observatio	1000							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regressior	1	8891.64	8891.64	9151.174	0			
Residual	998	969.696	0.971639					
Total	999	9861.336						
	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	29.97759	0.177794	168.6082	0	29.6287	30.32648	29.6287	30.32648
price	-1.99176	0.020821	-95.6618	0	-2.03261	-1.9509	-2.03261	-1.9509

Last night, Tom, the new data entry clerk, keyed one point of 10000 sales at a price of $10000. What happens?

PhredTek, Inc.
logistics consulting services

The Danger of Outliers

Results with the outlier included

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.999894							
R Square	0.999789							
Adjusted R	0.999789							
Standard E	4.58751							
Observatio	1001							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regressior	1	99624718	99624718	4733834	0			
Residual	999	21024.21	21.04525					
Total	1000	99645743						
	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*lpper 95.0%*
Intercept	4.831199	0.145243	33.26283	6.6E-164	4.546182	5.116216	4.546182	5.116216
price	0.99945	0.000459	2175.738	0	0.998548	1.000351	0.998548	1.000351

One bad outlier out of 1000 good data points can make the regression meaningless

PhredTek, Inc.
logistics consulting services

Multiple Regression Forecasting

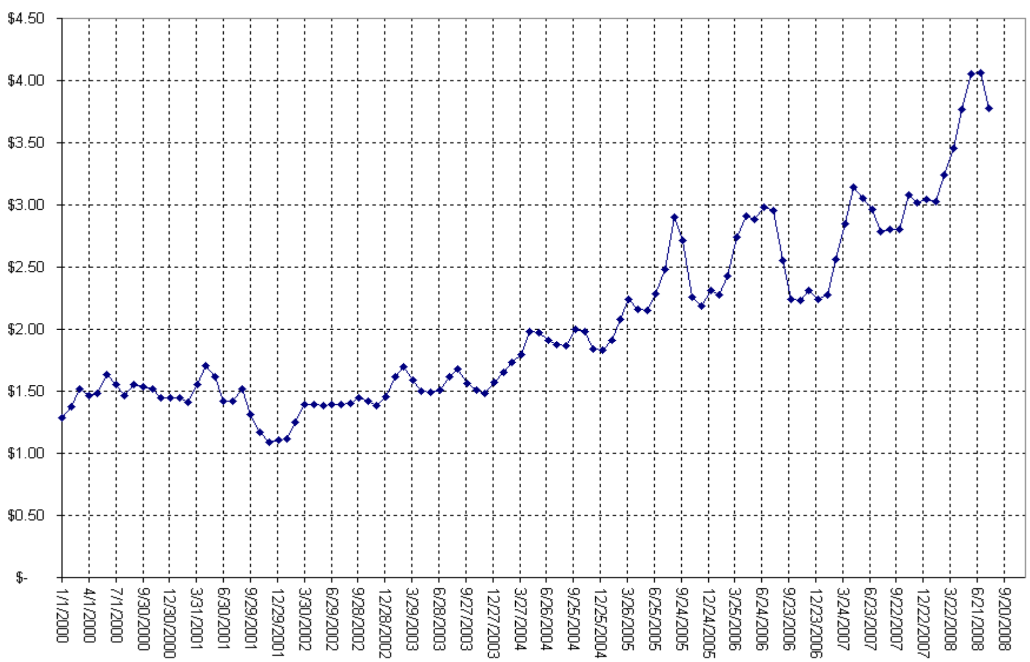
- Same as simple, but more than one variable.
- Y = Bt + e where:
  - Y is a column of numbers (vector) which are the values of the dependent variable.
  - B is a matrix where the first column is all ones and each successive column contains the valued of the independent variables.
  - t is a column of numbers which are the coefficients to be estimated.
  - e is a column of numbers which are the error.
- Y and B are known. t and e are to be estimated.

PhredTek, Inc.
logistics consulting services

Multiple Regression Forecasting

- Steps are the same as simple regression.
- Having a business reason for why each independent variable effects the dependent variable is more important.
- Most all software packages that do simple regression also do multiple regression.

Multiple Regression Forecasting
Handy Dummy Variables

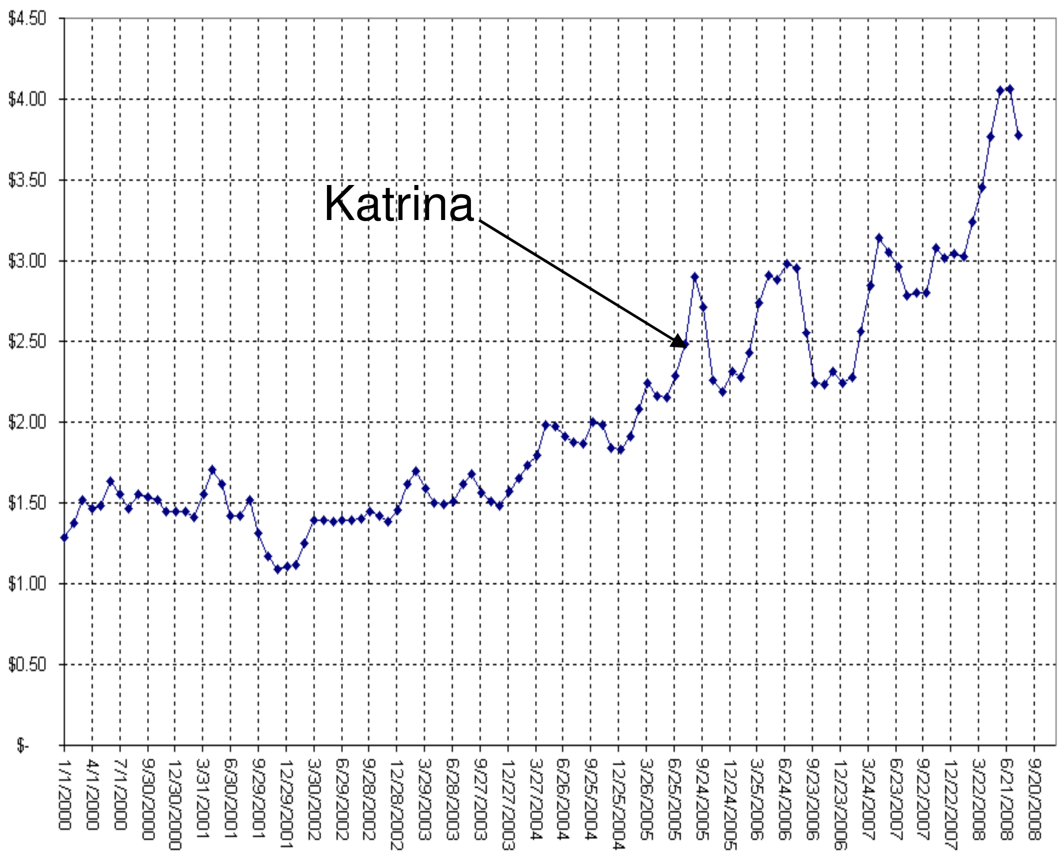


Trend

date	price	trend
10/1/2002	$ 1.45	0
11/1/2002	$ 1.42	0
12/1/2002	$ 1.39	0
1/1/2003	$ 1.46	0
2/1/2003	$ 1.61	1
3/1/2003	$ 1.69	2
4/1/2003	$ 1.59	3
5/1/2003	$ 1.50	4
6/1/2003	$ 1.49	5
7/1/2003	$ 1.51	6
8/1/2003	$ 1.62	7
9/1/2003	$ 1.68	8
10/1/2003	$ 1.56	9
11/1/2003	$ 1.51	10
12/1/2003	$ 1.48	11

PhredTek, Inc.
logistics consulting services

Multiple Regression Forecasting Handy Dummy Variables



Events Like Katrina

date	price	Kat	Kat-1	Kat-2
7/1/2005	$2.29	0	0	0
8/1/2005	$2.49	1	0	0
9/1/2005	$2.90	0	1	0
10/1/2005	$2.72	0	0	1
11/1/2005	$2.26	0	0	0
12/1/2005	$2.19	0	0	0

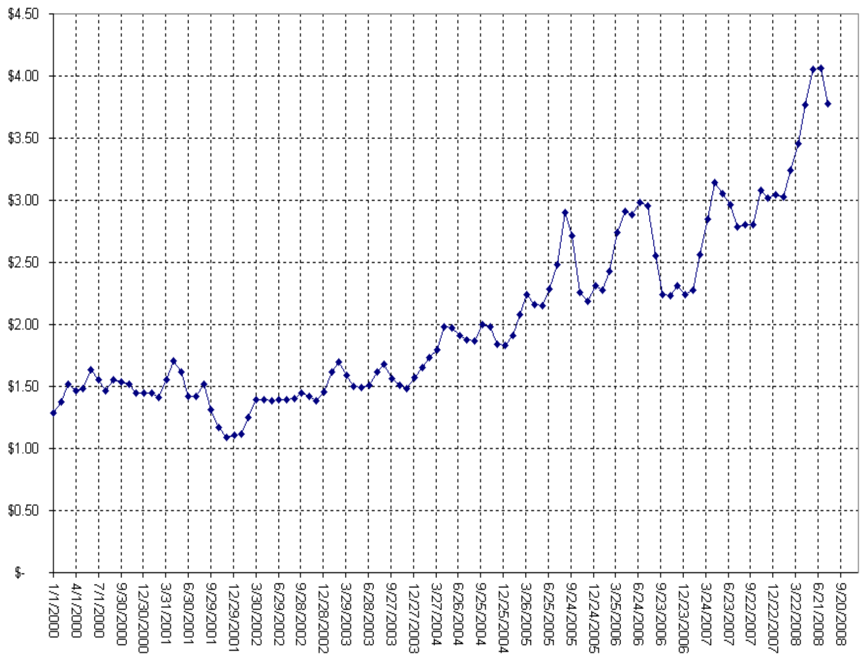
PhredTek, Inc.
logistics consulting services

Multiple Regression Forecasting New Problem

- Two or more independent variables can be related to one another – called multicollinearity.
- Example – the population of a town and the number of churches in it.
- Keeping both in the model causes bad coefficient estimates for both.
- Use business based reasoning to determine which is the true causal variable.
- If the correlation is exactly one, a good regression package will drop one of them.

PhredTek, Inc.
logistics consulting services

Multiple Regression Forecasting Handy Dummy Variables



Seasonal Dummies

date	price	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
1/1/2000	$ 1.29	1	0	0	0	0	0	0	0	0	0	0
2/1/2000	$ 1.38	0	1	0	0	0	0	0	0	0	0	0
3/1/2000	$ 1.52	0	0	1	0	0	0	0	0	0	0	0
4/1/2000	$ 1.47	0	0	0	1	0	0	0	0	0	0	0
5/1/2000	$ 1.49	0	0	0	0	1	0	0	0	0	0	0
6/1/2000	$ 1.63	0	0	0	0	0	1	0	0	0	0	0
7/1/2000	$ 1.55	0	0	0	0	0	0	1	0	0	0	0
8/1/2000	$ 1.47	0	0	0	0	0	0	0	1	0	0	0
9/1/2000	$ 1.55	0	0	0	0	0	0	0	0	1	0	0
10/1/2000	$ 1.53	0	0	0	0	0	0	0	0	0	1	0
11/1/2000	$ 1.52	0	0	0	0	0	0	0	0	0	0	1
12/1/2000	$ 1.44	0	0	0	0	0	0	0	0	0	0	0

PhredTek, Inc.
logistics consulting services

Multiple Regression Forecasting Gas Price Forecast Using Dummies

Estimated Model

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.966761
R Square	0.934626
Adjusted R	0.924342
Standard E	0.194816
Observatio	104

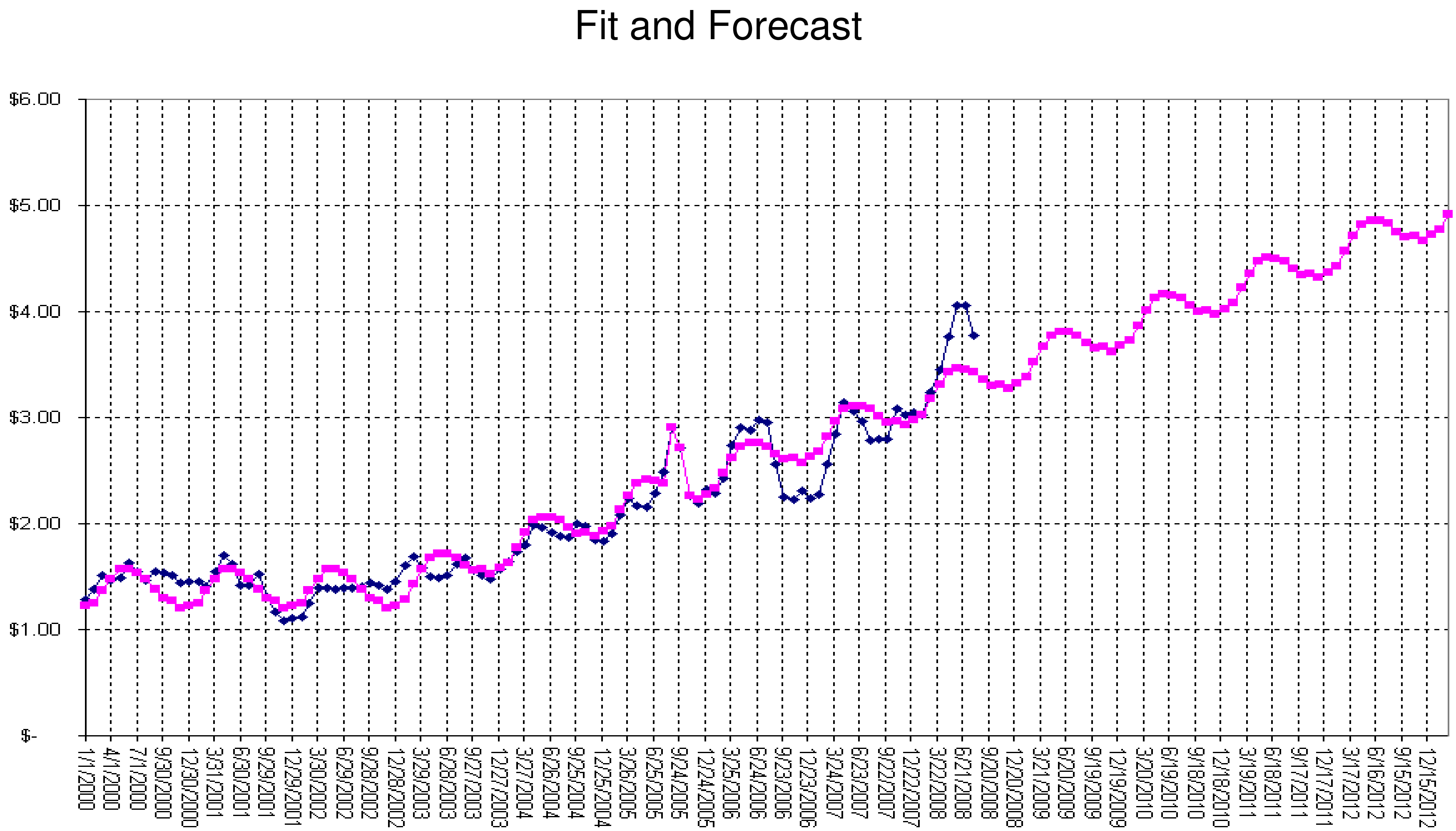
ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressior	14	48.29185	3.449418	90.88544	2.2E-46
Residual	89	3.377859	0.037953		
Total	103	51.66971			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	1.207092	0.071366	16.91407	1.52E-29	1.065289	1.348895	1.065289	1.348895
trend	0.02911	0.000854	34.089	7.26E-53	0.027413	0.030807	0.027413	0.030807
Katrina -1	0.59249	0.208597	2.840363	0.005584	0.178013	1.006967	0.178013	1.006967
Katrina -2	0.461872	0.208617	2.213964	0.029387	0.047353	0.87639	0.047353	0.87639
Jan	0.022151	0.094677	0.233962	0.815552	-0.16597	0.210273	-0.16597	0.210273
Feb	0.046633	0.094669	0.492588	0.623517	-0.14147	0.234739	-0.14147	0.234739
Mar	0.162449	0.094665	1.716037	0.089634	-0.02565	0.350546	-0.02565	0.350546
Apr	0.273486	0.094664	2.889025	0.004853	0.085391	0.461581	0.085391	0.461581
May	0.359635	0.094666	3.79898	0.000265	0.171535	0.547735	0.171535	0.547735
Jun	0.363006	0.094672	3.834356	0.000235	0.174895	0.551117	0.174895	0.551117
Jul	0.326599	0.094681	3.449464	0.000861	0.13847	0.514729	0.13847	0.514729
Aug	0.27297	0.094694	2.882664	0.004943	0.084816	0.461125	0.084816	0.461125
Sep	0.171895	0.100874	1.704063	0.091858	-0.02854	0.372329	-0.02854	0.372329
Oct	0.087404	0.10086	0.866584	0.3885	-0.113	0.28781	-0.113	0.28781
Nov	0.069819	0.09741	0.716754	0.475403	-0.12373	0.26337	-0.12373	0.26337

PhredTek logistics consulting

Multiple Regression Forecasting
Gas Price Forecast Using Dummies



Regression Danger

- Do not include any variable unless you have a business reason to believe it affects sales.
- Example:

month	sales	trend	trend^2	trend^3	t
Jan-00	2	1	1	1	
Feb-00	6	2	4	8	
Mar-00	2	3	9	27	
Apr-00	6	4	16	64	
May-00		5	25	125	
Jun-00		6	36	216	
Jul-00		7	49	343	
Aug-00		8	64	512	
Sep-00		9	81	729	
Oct-00		10	100	1000	

Regression Danger

- Do not include any variable unless you have a business reason to believe it affects sales.
- Regression Results:

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	65535
Standard Error	0
Observations	4

ANOVA

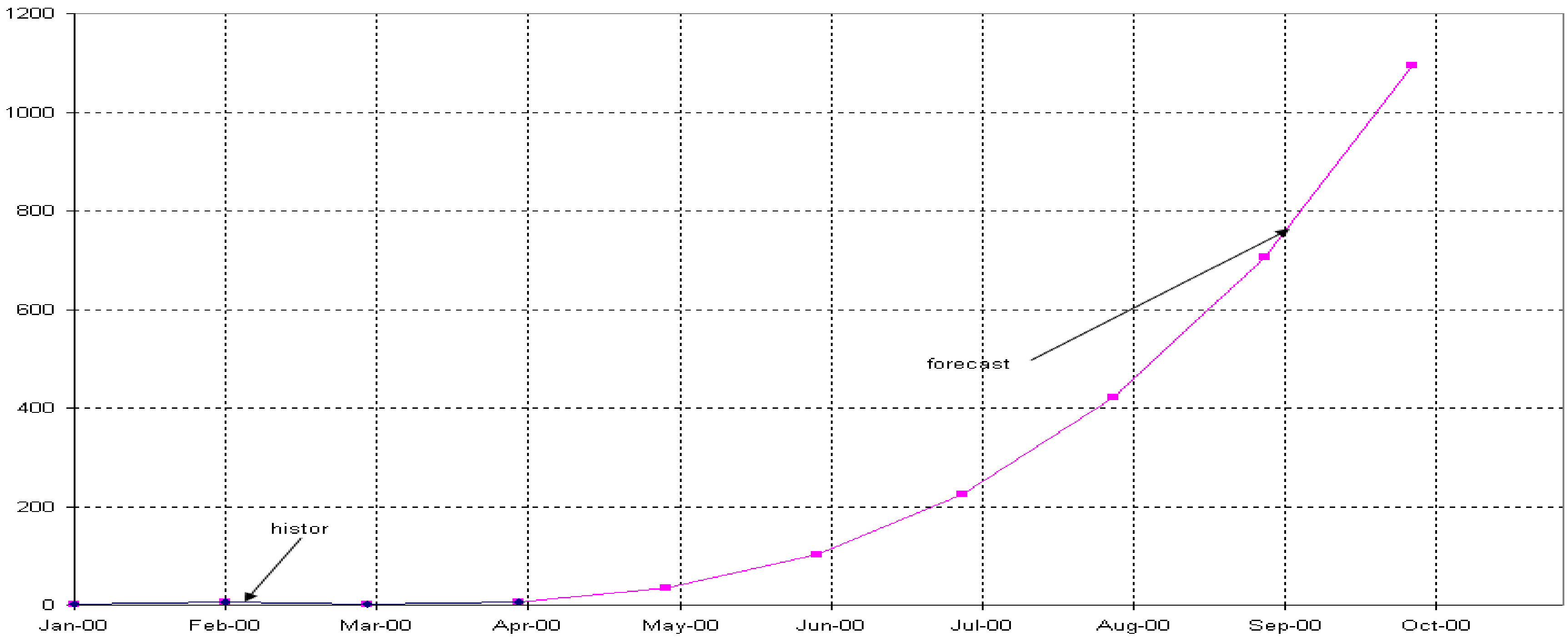
	df	*SS*	*MS*	*F*	*ignificance F*
Regression	3	16	5.333333	#NUM!	#NUM!
Residual	0	0	65535		
Total	3	16			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-26	0	65535	#NUM!	-26	-26	-26	-26
trend	45.33333	0	65535	#NUM!	45.33333	45.33333	45.33333	45.33333
trend^2	-20	0	65535	#NUM!	-20	-20	-20	-20
trend^3	2.666667	0	65535	#NUM!	2.666667	2.666667	2.666667	2.666667

PhredTek, In

Regression Danger

- Do not include any variable unless you have a business reason to believe it affects sales.
- Fit and forecast:



PhredTek, Inc.
logistics consulting services

Regression Forecasting

- Regression assumptions:
  - The relationship between independent and dependent variables is linear
  - Errors must be independent – that is, they have no autocorrelation.
  - All errors have the same variance (said to be homoscedastic
  - All errors are normally distributed.

PhredTek, Inc.
logistics consulting services

Excel Exercise

4. Forecasting Accuracy

Three Basic Error Measures

- $\text{Error}_t = \text{Forecast}_t - \text{Actual}_t = F_t - A_t$
- $\text{Absolute error}_t = |F_t - A_t|$
- $\text{Squared error} = (F_t - A_t)^2$

These are the error measures for a single time period. They are the basis for developing summary error measures

PhredTek, Inc.
logistics consulting services

Summary Error Measures Single Series

- Average or mean error = $\bar{e} = \frac{1}{n}\sum_{t=1}^{n}(F_t - A_t)$
- Net percent error = $\frac{\bar{e}}{\frac{1}{n}\sum_{t=1}^{n}A_t}$
- Mean Absolute Error (MAE) = $\frac{1}{n}\sum_{t=1}^{n}\left|F_t - A_t\right|$

PhredTek, Inc.
logistics consulting services

Summary Error Measures
Single Series

- ***Mean Absolute Percent Error (MAPE)*** = $\dfrac{MAE}{\frac{1}{n}\sum_{t=1}^{n} A_t}$

- Error Variance = $\dfrac{\sum_{t=1}^{n}(e_t - \bar{e})^2}{n-1}$

- Error Standard deviation = $\sqrt{\dfrac{\sum_{t=1}^{n}(e_t - \bar{e})^2}{n-1}}$

- Sum of Squares of Error = $\sum_{t=1}^{n}(F_t - A_t)^2$

Summary Error Measures Multiple Series

- Let $F_{i,t} = \text{forecast for series i at time t}$
- And $A_{i,t} = \text{actual for series i at time t}$
- And $\bar{e}_i = \text{average error for series i}$
- And $\bar{A}_i = \frac{1}{n}\sum_{t=1}^{n} A_{i,t} = \text{average of } \mathrm{A_i}$
- And $MAE_i = \text{mean absolute error of series i}$
- And that there are m series

PhredTek, Inc.
logistics consulting services

Summary Error Measures Multiple Series

- Unweighted mean percent error

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\overline{e}_i}{\overline{A}_i}$$

- Unweighted mean absolute percent error

$$\frac{1}{m}\sum_{i=1}^{m}\frac{MAE_i}{\overline{A}_i}$$

PhredTek, Inc.
logistics consulting services

Summary Error Measures Multiple Series

- Weighted mean percent error

$$\frac{\sum_{i=1}^{m} \bar{e}_i \bar{A}_i}{\sum_{i=1}^{m} \bar{A}_i} = \frac{\sum_{i=1}^{m}\sum_{t=1}^{n}(F_{i,t} - A_{i,t})}{\sum_{i=1}^{m}\sum_{t=1}^{n} A_{i,t}}$$

- Weighted mean absolute percent error

$$\frac{\sum_{i=1}^{m} MAE_i \bar{A}_i}{\sum_{i=1}^{m} \bar{A}_i} = \frac{\sum_{i=1}^{m}\sum_{t=1}^{n}\left|F_{i,t} - A_{i,t}\right|}{\sum_{i=1}^{m}\sum_{t=1}^{n} A_{i,t}}$$

PhredTek, Inc.
logistics consulting services

Uses of Summary Error Measures Single Series

- Average or mean error – describes the bias of the forecast. It should go to zero over time or the forecast needs improvement.
- Net percent error – describes the bias relative to average sales – most useful for communication to others
- Mean absolute error – describes the pain the forecast will inflict because over forecasts are generally just as bad or worse than under forecasts.
- Mean absolute percent error (MAPE) – most useful statistic for communicating forecast error to others.

PhredTek, Inc.
logistics consulting services

Uses of Summary Error Measures Single Series

- Error variance and standard deviation are used for determining confidence limits of forecasts and for setting safety stock.
- Sum of squared error is the most common statistic that is minimized to fit a forecast model. It is also useful for comparing one model fit to another.

PhredTek, Inc.
logistics consulting services

Uses of Summary Error Measures Multiple Series

- Mean percent error and mean absolute percent error have the same use for multiple series as they to for a single series.
- Use weighted when the units of measure for all series are the same.
- Use unweighted when the units of measure are different.

PhredTek, Inc.
logistics consulting services

Save All Your Forecasts

- Data storage is inexpensive.
- Database should contain:
  - Forecast origin – last date in history used to generate the forecast
  - Forecast date – the date being forecast
  - Forecast value
  - Actual sales for the forecast date once it is available.
- For example if you do a 26 week forecast, you will have 26 forecasts for each week.
- You can now evaluate any forecast error and:
  - Clearly demonstrate improvement over time
  - Identify those forecasts in need of improvement

Save All Your Forecasts

Forecasts As They Happen

Forecast Origin (last month of data used in the forecast)

date	price	3/2000	4/2000	5/2000	6/2000	7/2000	8/2000	9/2000	10/2000	11/2000	12/2000	1/2001	2/2001	3/2001	4/2001	5/2001	6/2001	7/2001	8/2001	9/2001	10/2001	11/2001	12/2001
1/1/2000	$ 1.29																						
2/1/2000	$ 1.38																						
3/1/2000	$ 1.52																						
4/1/2000	$ 1.47	$1.39																					
5/1/2000	$ 1.49	$1.39	$1.45																				
6/1/2000	$ 1.63	$1.39	$1.45	$1.49																			
7/1/2000	$ 1.55	$1.39	$1.45	$1.49	$1.53																		
8/1/2000	$ 1.47		$1.45	$1.49	$1.53	$1.56																	
9/1/2000	$ 1.55			$1.49	$1.53	$1.56	$1.55																
10/1/2000	$ 1.53				$1.53	$1.56	$1.55	$1.52															
11/1/2000	$ 1.52					$1.56	$1.55	$1.52	$ 1.52														
12/1/2000	$ 1.44						$1.55	$1.52	$ 1.52	$ 1.53													
1/1/2001	$ 1.45							$1.52	$ 1.52	$ 1.53	$ 1.50												
2/1/2001	$ 1.45								$ 1.52	$ 1.53	$ 1.50	$1.47											
3/1/2001	$ 1.41									$ 1.53	$ 1.50	$1.47	$1.45										
4/1/2001	$ 1.55										$ 1.50	$1.47	$1.45	$1.44									
5/1/2001	$ 1.70											$1.47	$1.45	$1.44	$1.47								
6/1/2001	$ 1.62												$1.45	$1.44	$1.47	$1.55							
7/1/2001	$ 1.42													$1.44	$1.47	$1.55	$1.62						
8/1/2001	$ 1.42														$1.47	$1.55	$1.62	$1.58					
9/1/2001	$ 1.52															$1.55	$1.62	$1.58	$1.49				
10/1/2001	$ 1.32																$1.62	$1.58	$1.49	$1.45			
11/1/2001	$ 1.17																	$1.58	$1.49	$1.45	$ 1.42		
12/1/2001	$ 1.09																		$1.49	$1.45	$ 1.42	$ 1.34	
1/1/2002	$ 1.11																			$1.45	$ 1.42	$ 1.34	$ 1.19
2/1/2002	$ 1.11																				$ 1.42	$ 1.34	$ 1.19
3/1/2002	$ 1.25																					$ 1.34	$ 1.19
4/1/2002	$ 1.40																						$ 1.19

PhredTek, Inc.
logistics consulting services

Save All Your Forecasts

How To Store Them and Compute Error

date	price	1 mo out fcst	2 mo out fcst	3 mo out fcst	4 mo out fcst	1 mo out error	2 mo out error	3 mo out error	4 mo out error	1 mo out abs error	2 mo out abs error	3 mo out abs error	4 mo out abs error
1/1/2000	$ 1.29												
2/1/2000	$ 1.38												
3/1/2000	$ 1.52												
4/1/2000	$ 1.47	$1.39				-0.071				0.071			
5/1/2000	$ 1.49	$1.45	$1.39			-0.032	-0.091			0.032	0.091		
6/1/2000	$ 1.63	$1.49	$1.45	$1.39		-0.144	-0.180	-0.239		0.144	0.180	0.239	
7/1/2000	$ 1.55	$1.53	$1.49	$1.45	$1.39	-0.023	-0.062	-0.098	-0.157	0.023	0.062	0.098	0.157
8/1/2000	$ 1.47	$1.56	$1.53	$1.49	$1.45	0.091	0.063	0.024	-0.012	0.091	0.063	0.024	0.012
9/1/2000	$ 1.55	$1.55	$1.56	$1.53	$1.49	0.000	0.006	-0.022	-0.061	0.000	0.006	0.022	0.061
10/1/2000	$ 1.53	$1.52	$1.55	$1.56	$1.53	-0.010	0.018	0.024	-0.004	0.010	0.018	0.024	0.004
11/1/2000	$ 1.52	$1.52	$1.52	$1.55	$1.56	-0.001	0.005	0.033	0.039	0.001	0.005	0.033	0.039
12/1/2000	$ 1.44	$1.53	$1.52	$1.52	$1.55	0.090	0.073	0.079	0.107	0.090	0.073	0.079	0.107
1/1/2001	$ 1.45	$1.50	$1.53	$1.52	$1.52	0.050	0.086	0.069	0.075	0.050	0.086	0.069	0.075
2/1/2001	$ 1.45	$1.47	$1.50	$1.53	$1.52	0.019	0.047	0.083	0.066	0.019	0.047	0.083	0.066
3/1/2001	$ 1.41	$1.45	$1.47	$1.50	$1.53	0.038	0.060	0.088	0.124	0.038	0.060	0.088	0.124
4/1/2001	$ 1.55	$1.44	$1.45	$1.47	$1.50	-0.117	-0.105	-0.083	-0.055	0.117	0.105	0.083	0.055
5/1/2001	$ 1.70	$1.47	$1.44	$1.45	$1.47	-0.232	-0.267	-0.255	-0.233	0.232	0.267	0.255	0.233
6/1/2001	$ 1.62	$1.55	$1.47	$1.44	$1.45	-0.062	-0.146	-0.181	-0.169	0.062	0.146	0.181	0.169
7/1/2001	$ 1.42	$1.62	$1.55	$1.47	$1.44	0.202	0.133	0.049	0.014	0.202	0.133	0.049	0.014
8/1/2001	$ 1.42	$1.58	$1.62	$1.55	$1.47	0.159	0.202	0.133	0.049	0.159	0.202	0.133	0.049
9/1/2001	$ 1.52	$1.49	$1.58	$1.62	$1.55	-0.036	0.058	0.101	0.032	0.036	0.058	0.101	0.032
10/1/2001	$ 1.32	$1.45	$1.49	$1.58	$1.62	0.140	0.171	0.265	0.308	0.140	0.171	0.265	0.308
11/1/2001	$ 1.17	$1.42	$1.45	$1.49	$1.58	0.248	0.284	0.315	0.409	0.248	0.284	0.315	0.409
12/1/2001	$ 1.09	$1.34	$1.42	$1.45	$1.49	0.250	0.333	0.369	0.400	0.250	0.333	0.369	0.400
avg pct						1.8%	2.4%	2.7%	3.6%	6.6%	8.2%	9.0%	8.8%

PhredTek, Inc.
logistics consulting services

Improving Forecast Accuracy

- Regularly review forecast error
- Rank from largest to smallest mean absolute error.
- Graph the errors and look for patterns
- When you see a pattern, think about what business driver caused it.
- When you find a new business driver, factor it into the forecast.

PhredTek, Inc.
logistics consulting services

Improving Forecast Accuracy

- Assign responsibility and accountability for forecasts (across all stakeholder groups)
- Set realistic goals error reduction.
- Tie performance reviews to goals
- Provide good forecasting tools

PhredTek, Inc.
logistics consulting services

Goodness of Fit vs. Forecasting Accuracy

- Two methods of choosing forecasting method:
  - Forecast Accuracy - Withhold the last several historic values of actual sales, build a forecast based on the rest. Choose the method which best forecasts the withheld values.
  - Goodness of Fit - Choose the method which best fits all the historical values. That is, the method which minimizes the sum of squares of the fit minus the actual.

Forecasting Accuracy vs. Goodness of Fit

- Two methods of choosing forecasting method:
  - Forecast Accuracy – Withhold the last several historic values of actual sales, build a forecast based on the rest. Choose the method which best forecasts the withheld values using.
  - Goodness of Fit - Choose the method which best characterizes the historical values using pattern recognition and business knowledge of the process. Perform accepted model identification and model revision using statistical modeling procedures.

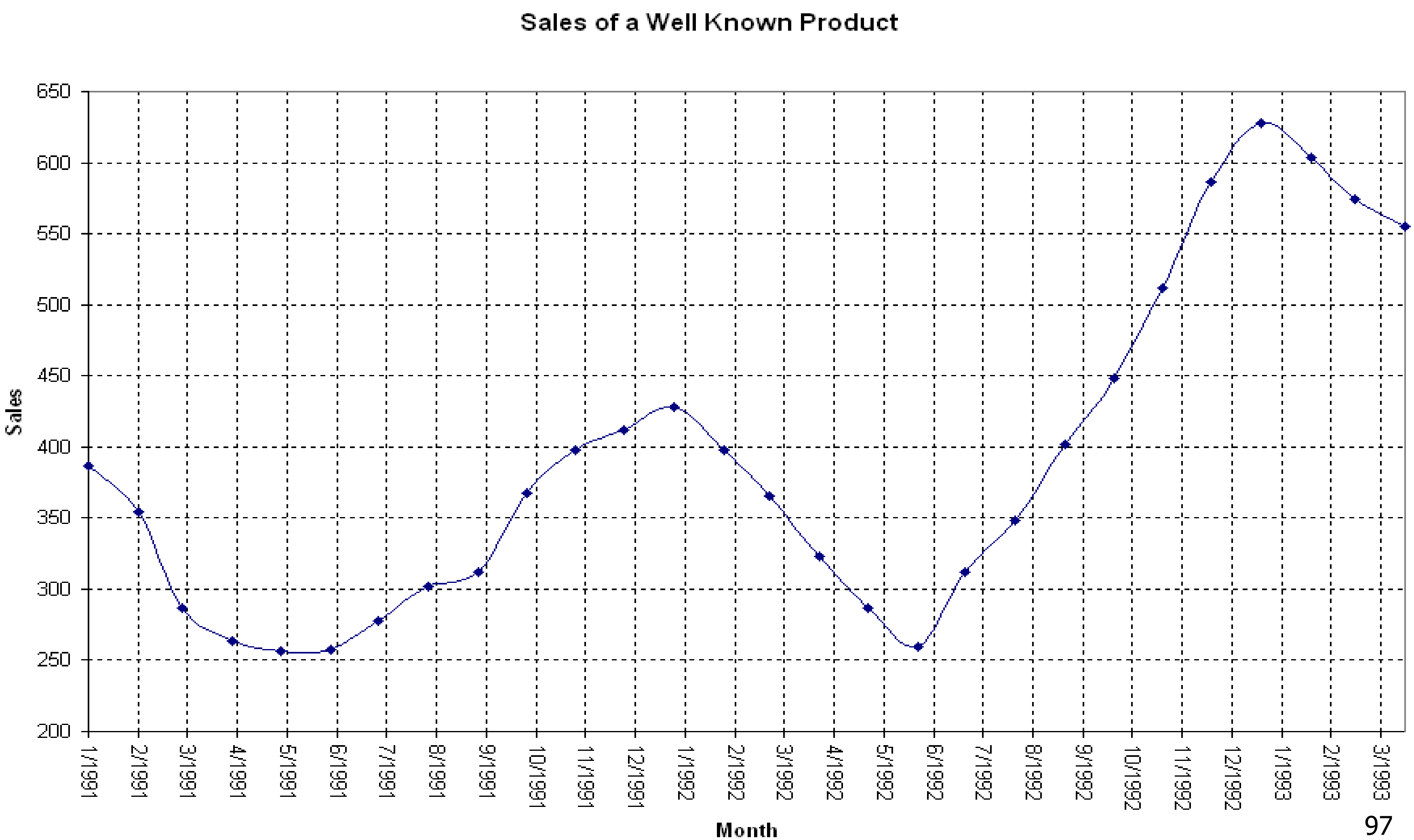
Fitting versus Modeling

- Two methods of choosing forecasting method:
  - Forecast Accuracy – Older Classical method based on using a list of models and simply picking the best of the list. No guarantees about statistical methodologies simply a sequence of trial models. Does not incorporate any knowledge of the business process.
  - Goodness of Fit – Choose variables that are known to affect sales. Iterate to the best model set of parameters by performing formal statistical tests.

Modeling Wins

- Fitting to minimize forecast error method is like the tail wagging the dog.
  - The answer is very different depending on how many observations are withheld.
  - The method chosen can change with each forecast run – leads to unstable forecasts and generally higher real forecast error.
  - No statistical tests are conducted for Necessity or Sufficiency
  - They are not based on solid business based cause and effect relationships.
- Modelling
  - Is stable because only statistically significant coefficients are used and the error process of the model is white noise (random)
  - Makes the management decision process more stable as model forms/parameters are less volatile
  - Makes the forecast more credible as equations have rational structure and can be easily explained to non-quantitative associates
  - Allows systematic forecast improvement.

Example



Sales of a Well Known Product
Sales
650
600
550
500
450
400
350
300
250
200
1/1991
2/1991
3/1991
4/1991
5/1991
6/1991
7/1991
8/1991
9/1991
10/1991
11/1991
12/1991
1/1992
2/1992
3/1992
4/1992
5/1992
6/1992
7/1992
8/1992
9/1992
10/1992
11/1992
12/1992
1/1993
2/1993
3/1993
Month

Mean Absolute Percent Error (MAPE)

Method	Retain 1	Retain 2	Retain 3
Average	39.0%	28.0%	30.0%
Autoregressive 1	11.0%	15.0%	25.0%
Autoregressive 1 and seasonal AR	5.0%	4.1%	9.6%
Exponential smoothing	5.4%	6.8%	15.0%
Enhanced exponential smoothing	13.0%	16.0%	27.0%
Random Walk	4.0%	5.7%	14.0%
Random Walk with shift	5.6%	7.9%	17.0%
Time Trend	21.0%	19.0%	24.0%
Time Trend with AR1	4.0%	14.0%	29.0%
Fourier	47.0%	50.0%	61.0%
Time Trend with seasonal factors	17.0%	19.0%	29.0%
Damped linear trend	11.0%	13.0%	22.0%
Holt Winters	0.6%	1.3%	11.0%
Airline model	0.5%	0.9%	10.0%

Number of months at the end of the series that were not used to develop the model.

Model with best MAPE is highlighted in blue. Note the radical change between withholding 2 and 3 months. One can expect this kind of change to continue with additional months withheld.

The historical values are subordinate to the withheld values. The "best model" depends on the number withheld thus it is not objective.

PhredTek, Inc.
logistics consulting services

Using Forecast Error to Determine Safety Stock

- Compute the standard deviation of the forecast error
  - It's best if you have actual historical forecast error.
  - If not, the standard deviation of the fit will suffice
- Determine the acceptable chances of a stock out. This is dependent on two factors
  - Inventory holding cost. The higher the holding cost, the higher chances of stock outs you are willing to accept.
  - Stock out cost. Generally, this is lost sales. The higher the stock out costs, the lower you want the chances of stocking out.
  - Inventory holding costs are easy to compute, but stock out costs often require an expensive market research study to determine. Therefore, an intuitive guess based on your business knowledge is generally sufficient.
- Look the stock out probability up in a normal table to determine the number of standard deviations necessary to achieve it.
- Multiply that number by the standard deviation of forecast error

PhredTek, Inc.
logistics consulting services

Using Forecast Error to Determine Safety Stock Example

Standard deviation of forecast error is 49.7 units

I want my stock out chances to be 5%. That means that I want a 95% confidence level.

Looking this up in the normal table says I need 1.64 standard deviations of safety stock

Multiply this by 49.7 means I need 82 units of safety stock.

Normal Table

confidence	stds
0.80	0.84
0.81	0.88
0.82	0.92
0.83	0.95
0.84	0.99
0.85	1.04
0.86	1.08
0.87	1.13
0.88	1.17
0.89	1.23
0.90	1.28
0.91	1.34
0.92	1.41
0.93	1.48
0.94	1.55
0.95	1.64
0.96	1.75
0.97	1.88
0.98	2.05
0.99	2.33

PhredTek, Inc.
logistics consulting services

5. Putting it all Together
The Forecast Process

PhredTek, Inc.
logistics consulting services

Building a Forecast Process

- Components of a forecast process
  - People
  - Systems
  - Scheduled Actions
  - Scheduled Communications

A forecast without a process to implement it will never be used.

PhredTek, Inc.
logistics consulting services

Building a Forecast Process

- Necessary scheduled actions
  - Collect data
  - Generate forecast
  - Save the forecast
  - Tabulate forecast error
- Necessary scheduled communications
  - Communicate forecast to stakeholders
  - Communicate error to stakeholders

Building a Forecast Process

- Identify stakeholders – People who allocate resources.
  - Sales
  - Marketing
  - Supply Chain Planning
  - Operations
  - Purchasing
  - Corporate Planning
  - Engineering
  - Distributors
  - Retailers

Building a Forecast Process

- Determine the level of detail in geography and time that each stakeholder requires
- Focus on one to three stakeholders first
- Stakeholders more likely to cooperate
  - Supply Chain Planning
  - Marketing
  - Corporate Planning

PhredTek, Inc.
logistics consulting services

Building a Forecast Process

- Build Necessary components
  - Data collection
  - Forecast generation
  - Forecast data base
  - Forecast reports and distribution schedules
  - Error reports

The degree of automation of these components is directly correlated with your chances of success

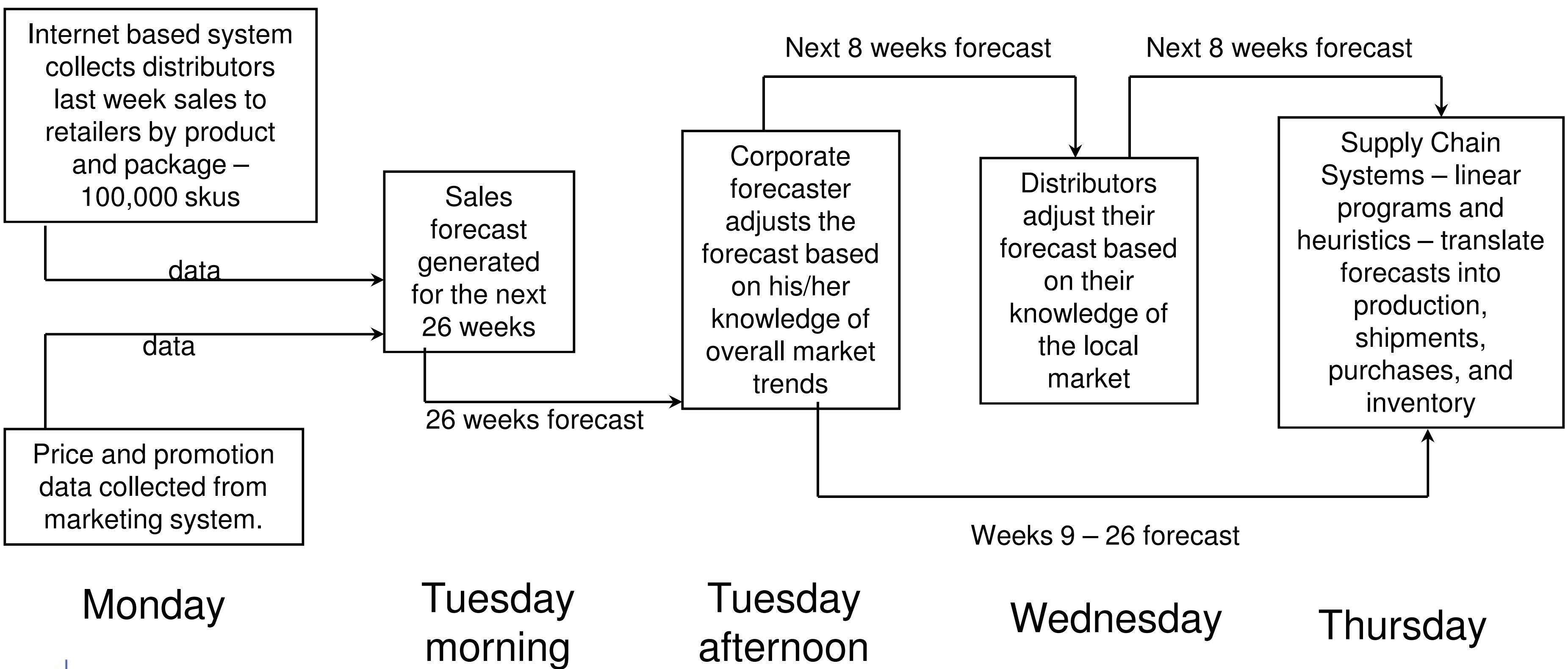
Building a Forecast Process

- Sell the process
  - To stakeholders
  - To management
- Quantify benefits where possible
  - Inventory reduction
  - Reduction in lost sales

If your CEO is not firmly committed to making data base decisions, your chances of success are limited.

Example Forecast Process

- Company X allocates production, transportation, and purchasing resources weekly



PhredTek, Inc.
logistics consulting services

Software Selection

- Software is required for a practical forecasting process.
- Clearly define the problem that the software should address
  - Geographical detail
  - Time detail
  - Number of time series
- Make sure the forecasting process is well defined ahead of time so you can clearly identify where the software fits in.
- Involve key stakeholders
  - Showing them a prototype forecast helps.
- Look at multiple packages.

THE JOURNAL OF BUSINESS FORECASTING

Volume 26, Issue 2 Summer 2007

PhredTek, Inc.
logistics consulting services

Hints From the Journal of Business Forecasting

- Acquire and use software that
  - **Builds both univariate and causal time series models.**
  - **Uses rigorous, well documented statistical techniques.**
  - Provides a transparent audit trail detailing how the model was developed. This can then be used by your experts (independent consultant / local university professor) to assess the thoroughness of the approach.
  - **Provides Robust Estimation incorporating pulses, seasonal pulses, step and trend changes.**
  - Automatically adjusts for changes in variance or parameters over time.

Hints From the Journal of Business Forecasting(and me)

- Acquire and use software that
  - **Recommends a model for use in the forecast and permits model re-use**
  - Has built-in baseline models (e.g. expo smooth, simple trend etc)which can be used to assess forecasting improvements via modelling
  - Provides an easy to understand written explanation of the model for your quantitatively challenged team members
  - Allows you to form your own home-brew model
  - **Can get it's data from multiple sources such as excel, data bases, and sequential text files.**
  - **Can be used in both production and exploratory modes**
  - ALWAYS TEST THE SOFTWARE against either textbook examples or your own data

Day 2
ARIMA and Transfer Function Models

Basic Statistics

PhredTek, Inc.
logistics consulting services

Basic Statistics Review

$$mean = \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

$$\text{var}iance = \sigma^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \bar{x})(x_t - \bar{x}) = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \bar{x})^2$$

$$\text{cov}ariance = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})$$

$$correlation = \frac{\text{cov}ariance}{\sqrt{\text{var}iance(x)}\sqrt{\text{var}iance(y)}}$$

PhredTek, Inc.
logistics consulting services

Autocorrelation and Partial Autocorrelation

Key Extensions to Time Series Statistics

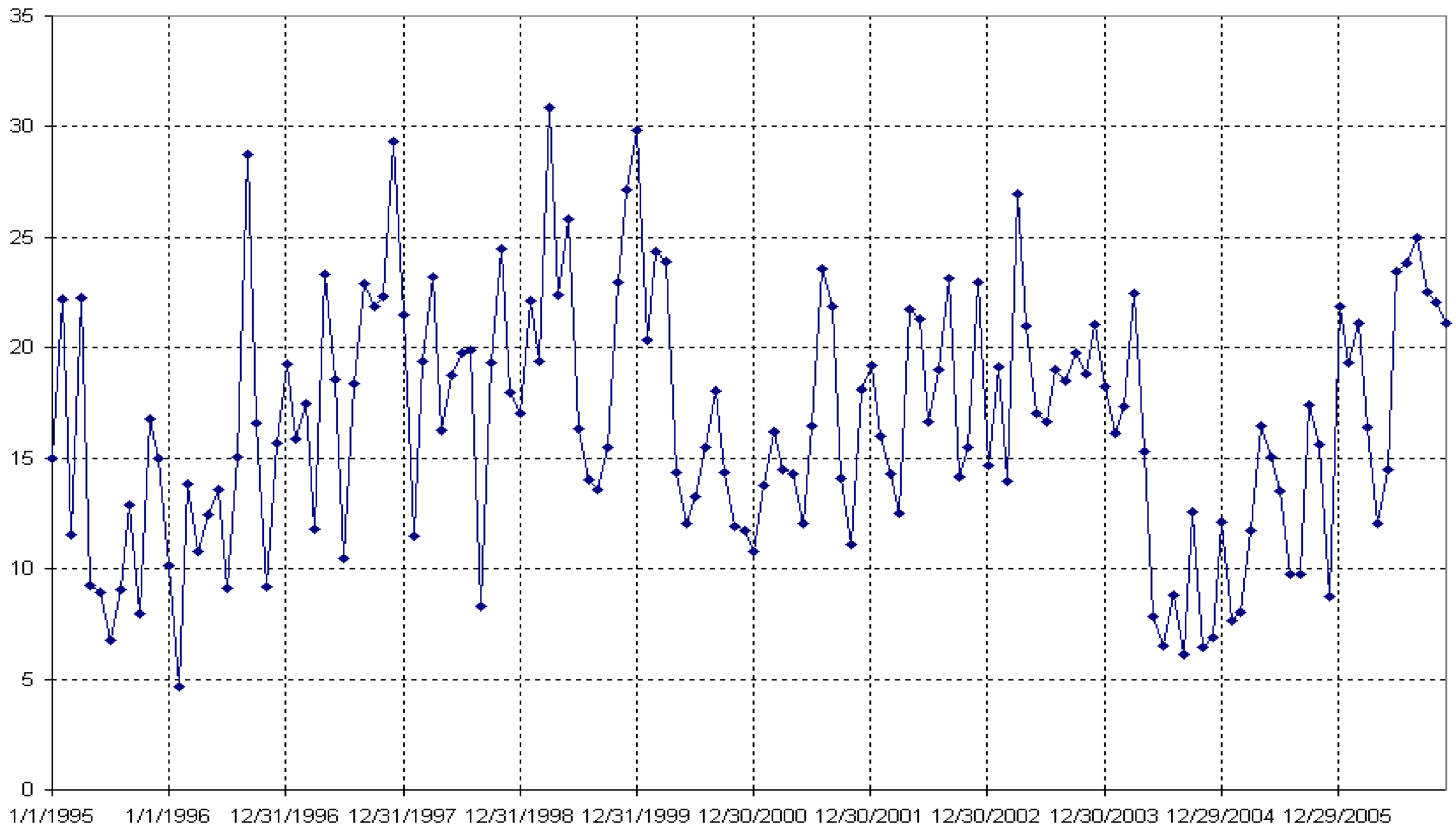
$$autocovariance(k) = \frac{1}{n-1}\sum_{t=k}^{n}(x_t - \bar{x})(x_{t-k} - \bar{x})$$

$$autocorrelation(k) = \frac{autocovariance(k)}{\sqrt{variance(x)}\sqrt{variance(x)}} = \frac{autocovariance(k)}{variance(x)}$$

PhredTek, Inc.
logistics consulting services

Autocorrelation of lag 1

Example 1 - 144 Monthly Sales Observations



Phredlek, Inc.
logistics consulting services

Autocorrelation of lag 1

date	sales	lag 1 sales	sales - mean	lag sales -mean	Covar
1/1/1995	15.00				
2/1/1995	22.21	15.00	5.62	-1.59	-8.95
3/1/1995	11.53	22.21	-5.06	5.62	-28.46
4/1/1995	22.26	11.53	5.67	-5.06	-28.69
5/1/1995	9.27	22.26	-7.33	5.67	-41.50
6/1/1995	8.90	9.27	-7.70	-7.33	56.39
7/1/1995	6.79	8.90	-9.81	-7.70	75.49
8/1/1995	9.06	6.79	-7.53	-9.81	73.84
9/1/1995	12.87	9.06	-3.73	-7.53	28.06
10/1/1995	7.98	12.87	-8.61	-3.73	32.09
11/1/1995	16.76	7.98	0.17	-8.61	-1.45
12/1/1995	14.98	16.76	-1.61	0.17	-0.27
1/1/1996	10.14	14.98	-6.45	-1.61	10.38
2/1/1996	4.66	10.14	-11.93	-6.45	76.94
3/1/1996	13.84	4.66	-2.76	-11.93	32.87
4/1/1996	10.77	13.84	-5.82	-2.76	16.04
5/1/1996	12.43	10.77	-4.17	-5.82	24.24
6/1/1996	13.59	12.43	-3.00	-4.17	12.50
7/1/1996	9.15	13.59	-7.45	-3.00	22.35
8/1/1996	15.03	9.15	-1.57	-7.45	11.66
9/1/1996	28.75	15.03	12.16	-1.57	-19.05
10/1/1996	16.58	28.75	-0.01	12.16	-0.10
11/1/1996	9.21	16.58	-7.38	-0.01	0.06
12/1/1996	15.71	9.21	-0.88	-7.38	6.52
1/1/1997	19.24	15.71	2.65	-0.88	-2.34
2/1/1997	15.88	19.24	-0.71	2.65	-1.88
3/1/1997	17.46	15.88	0.86	-0.71	-0.61
4/1/1997	11.82	17.46	-4.77	0.86	-4.12
5/1/1997	23.31	11.82	6.72	-4.77	-32.09
6/1/1997	18.57	23.31	1.98	6.72	13.29
7/1/1997	10.45	18.57	-6.14	1.98	-12.14
8/1/1997	18.34	10.45	1.75	-6.14	-10.73
9/1/1997	22.89	18.34	6.30	1.75	11.01
10/1/1997	21.88	22.89	5.28	6.30	33.29
11/1/1997	22.29	21.88	5.70	5.28	30.13

Mean = 16.59
Variance = 30.67
Sum of Covariance/(n-1) = 14.75
Autocorrelation lag 1 = 14.75/30.67 = .481

Sample computations for the first 35 observations

Autocorrelation of lag 2

date	sales	lag 1 sales	sales - mean	lag sales -mean	Covar
1/1/1995	15.00				
2/1/1995	22.21				
3/1/1995	11.53	15.00	-5.06	-1.59	8.06
4/1/1995	22.26	22.21	5.67	5.62	31.84
5/1/1995	9.27	11.53	-7.33	-5.06	37.09
6/1/1995	8.90	22.26	-7.70	5.67	-43.61
7/1/1995	6.79	9.27	-9.81	-7.33	71.84
8/1/1995	9.06	8.90	-7.53	-7.70	57.95
9/1/1995	12.87	6.79	-3.73	-9.81	36.55
10/1/1995	7.98	9.06	-8.61	-7.53	64.82
11/1/1995	16.76	12.87	0.17	-3.73	-0.63
12/1/1995	14.98	7.98	-1.61	-8.61	13.85
1/1/1996	10.14	16.76	-6.45	0.17	-1.08
2/1/1996	4.66	14.98	-11.93	-1.61	19.20
3/1/1996	13.84	10.14	-2.76	-6.45	17.77
4/1/1996	10.77	4.66	-5.82	-11.93	69.44
5/1/1996	12.43	13.84	-4.17	-2.76	11.48
6/1/1996	13.59	10.77	-3.00	-5.82	17.47
7/1/1996	9.15	12.43	-7.45	-4.17	31.01
8/1/1996	15.03	13.59	-1.57	-3.00	4.70
9/1/1996	28.75	9.15	12.16	-7.45	-90.53
10/1/1996	16.58	15.03	-0.01	-1.57	0.01
11/1/1996	9.21	28.75	-7.38	12.16	-89.76
12/1/1996	15.71	16.58	-0.88	-0.01	0.01
1/1/1997	19.24	9.21	2.65	-7.38	-19.56
2/1/1997	15.88	15.71	-0.71	-0.88	0.63
3/1/1997	17.46	19.24	0.86	2.65	2.29
4/1/1997	11.82	15.88	-4.77	-0.71	3.40
5/1/1997	23.31	17.46	6.72	0.86	5.81
6/1/1997	18.57	11.82	1.98	-4.77	-9.44
7/1/1997	10.45	23.31	-6.14	6.72	-41.27
8/1/1997	18.34	18.57	1.75	1.98	3.45
9/1/1997	22.89	10.45	6.30	-6.14	-38.68
10/1/1997	21.88	18.34	5.28	1.75	9.23
11/1/1997	22.29	22.89	5.70	6.30	35.92

Mean = 16.59
Variance = 30.67
Sum of Covariance/(n-1) = 8.51
Autocorrelation lag 2 = 8.51/30.67 = .277

Sample computations for the first 35 observations

Partial Autocorrelation

- Autocorrelation shows the relationship between x_t and x_{t-k} without regard to anything that happened in between.
- Partial autocorrelation shows the marginal contribution of the kth lag after factoring out the effects of lags 1 to k-1.
- ACF can be thought of as the regression coefficient of the kth lag when performing a simple regression of the time series on lag k.
- PACF can be thought of as the regression coefficient of the kth lag when performing a multiple regression of the time series on lags 1 to k.

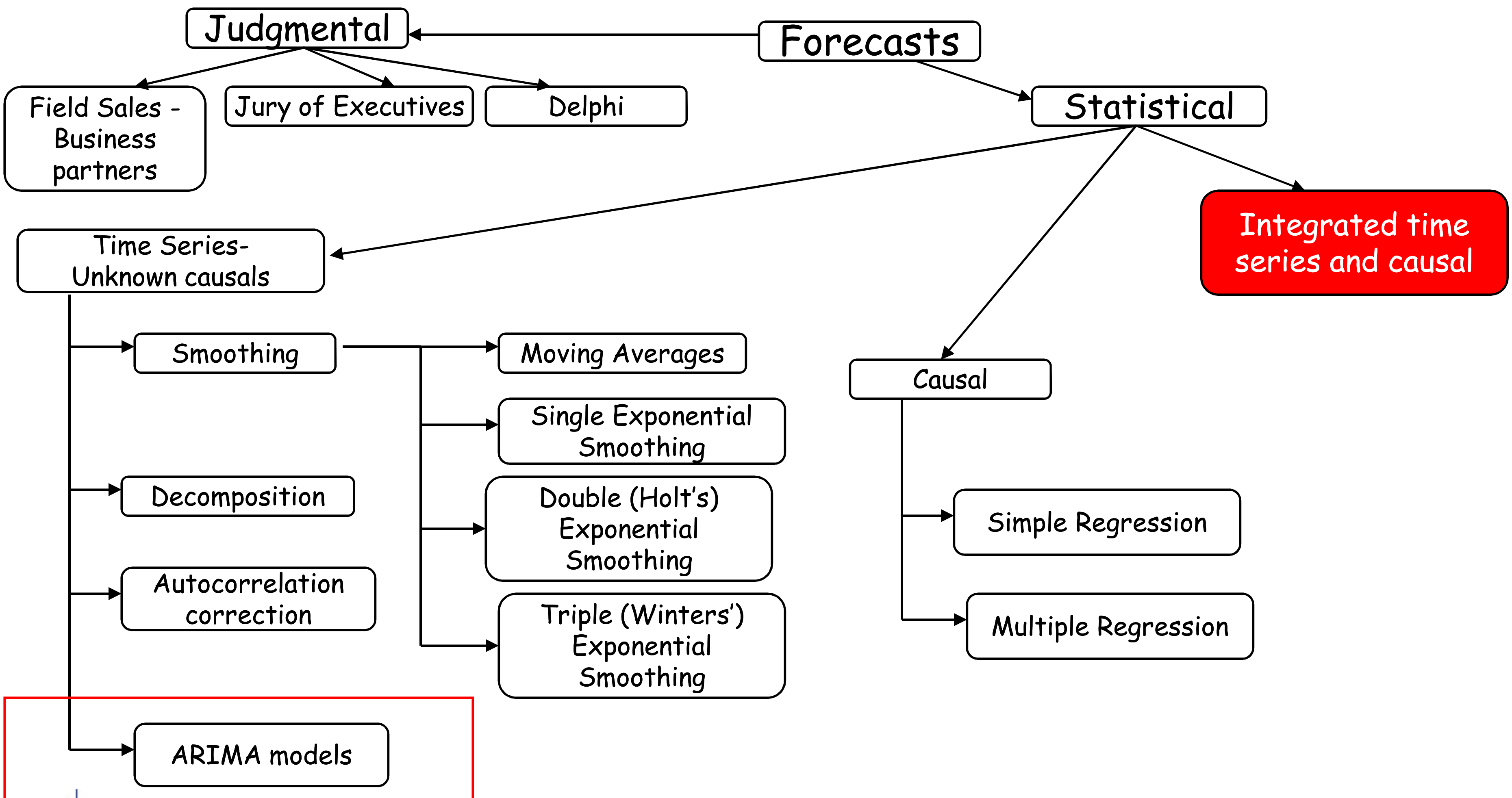
ACF and PACF for Example 1

lag	ACF	autocov	PACF
0	1.000	30.672	
1	0.481	14.754	0.481
2	0.277	8.511	0.060
3	0.282	8.657	0.167
4	0.297	9.113	0.129
5	0.201	6.169	-0.017
6	0.176	5.406	0.047
7	0.192	5.879	0.053
8	0.190	5.843	0.044
9	0.161	4.932	0.022
10	0.099	3.035	-0.047
11	0.069	2.122	-0.029
12	0.034	1.049	-0.052

The Family of ARIMA Models

PhredTek, Inc.
logistics consulting services

Overview of Forecasting Methods



PhredTek, Inc.
logistics consulting services

The Lag Operator B

The lag operator lags a variable by one time period
Examples for the variable Y

$$BY_t = Y_{t-1}$$

$$B^2 Y_t = Y_{t-2}$$

$$B^{-1} Y_t = Y_{t+1}$$

$$(1 - a_0 B) Y_t = Y_t - a_0 Y_{t-1}$$

$$(1 - a_0 B - a_1 B^2) Y_t = Y_t - a_0 Y_{t-1} - a_1 Y_{t-2}$$

The AR1 Model

Autoregressive model with one lag ARIMA(1,0,0)

$$Y_t = c + \Phi_1 Y_{t-1} + a_t$$

$$Y_t = \text{sales at time t}$$

$$a_t = \text{error at time t}$$

$$\Phi_1 = \text{autoregressive coefficient}$$

$$c = \text{constant}$$

PhredTek, Inc.
logistics consulting services

The AR1 Model

Introduce the lag operator

$$Y_t = c + \Phi_1 B Y_t + a_t$$

Combine the Y_t terms

$$(1 - \Phi_1 B) Y_t = c + a_t$$

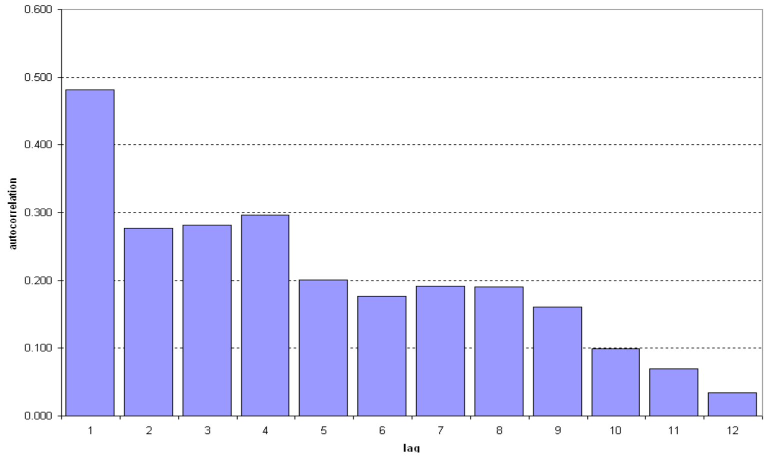
Divide by lag polynomial to get standard form

$$Y_t = \frac{c}{(1 - \Phi_1)} + \frac{1}{(1 - \Phi_1 B)} a_t = c' + \frac{1}{(1 - \Phi_1 B)} a_t$$

$$c' = \text{adjusted constant}$$

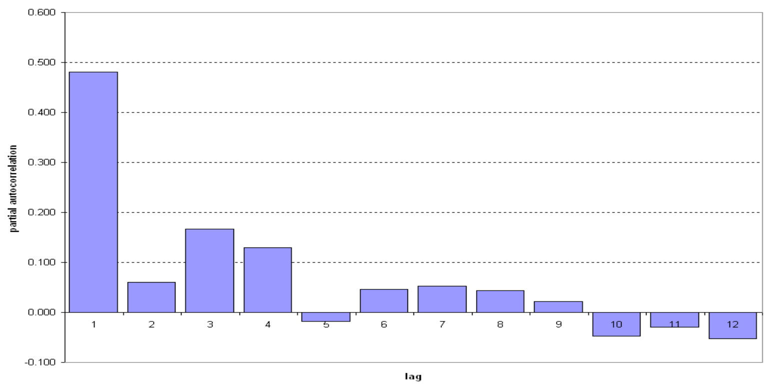
PhredTek, Inc.
logistics consulting services

ACF and PACF for Example 1



Autocorrelation

lag	ACF	autocov	PACF
0	1.000	30.672	
1	0.481	14.754	0.481
2	0.277	8.511	0.060
3	0.282	8.657	0.167
4	0.297	9.113	0.129
5	0.201	6.169	-0.017
6	0.176	5.406	0.047
7	0.192	5.879	0.053
8	0.190	5.843	0.044
9	0.161	4.932	0.022
10	0.099	3.035	-0.047
11	0.069	2.122	-0.029
12	0.034	1.049	-0.052



Partial Autocorrelation

This is the classic ACF and PACF for an AR1 process

PhredTek, Inc.
logistics consulting services

The MA1 Model

Moving average model with one lag ARIMA(0,0,1)

$$Y_t = c + a_t - \Theta_1 a_{t-1}$$

$$Y_t = \text{sales at time t}$$

$$a_t = \text{error at time t}$$

$$\Theta_1 = \text{moving average coefficient}$$

$$c = \text{constant}$$

PhredTek, Inc.
logistics consulting services

The MA1 Model

Introduce the lag operator

$$Y_t = c + a_t - \Theta_1 B a_t$$

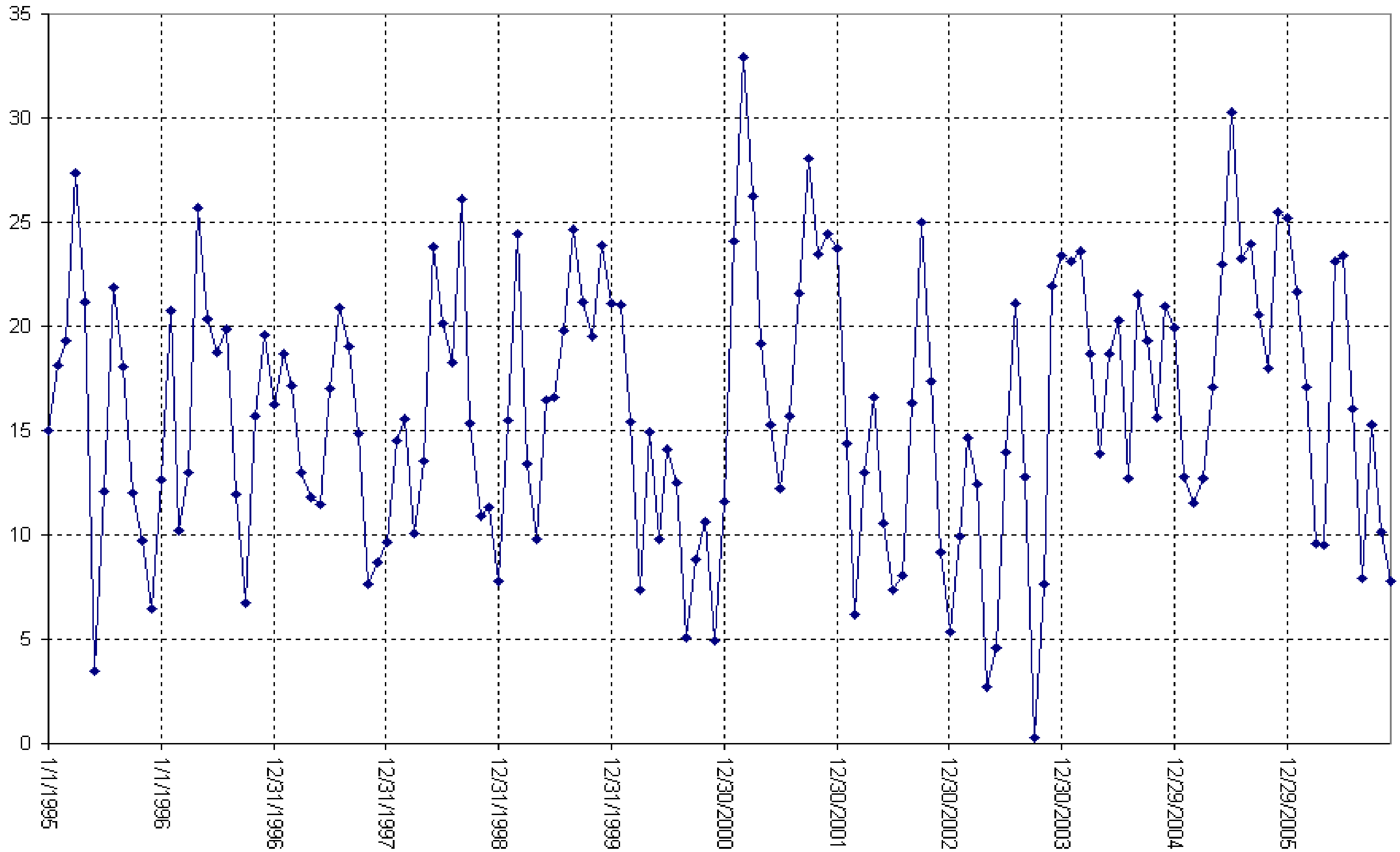
Combine the a_t terms

$$Y_t = c + (1 - \Theta_1 B) a_t$$

PhredTek, Inc.
logistics consulting services

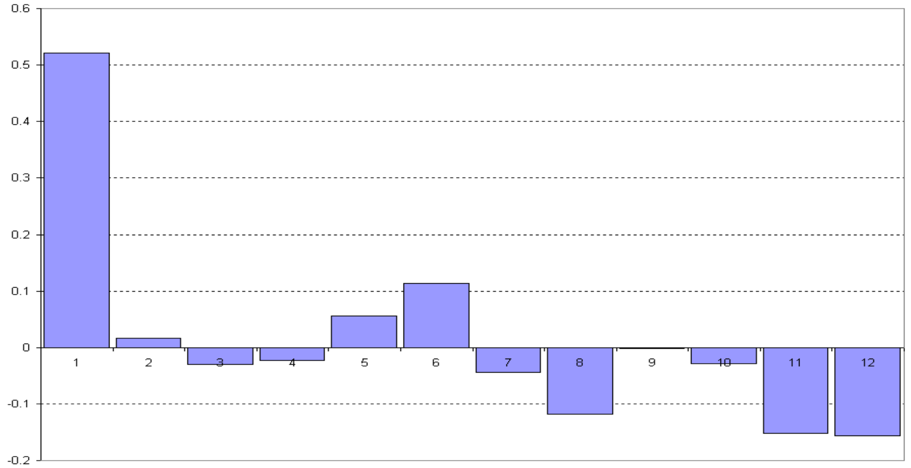
Example 2

144 monthly observations of another sales series



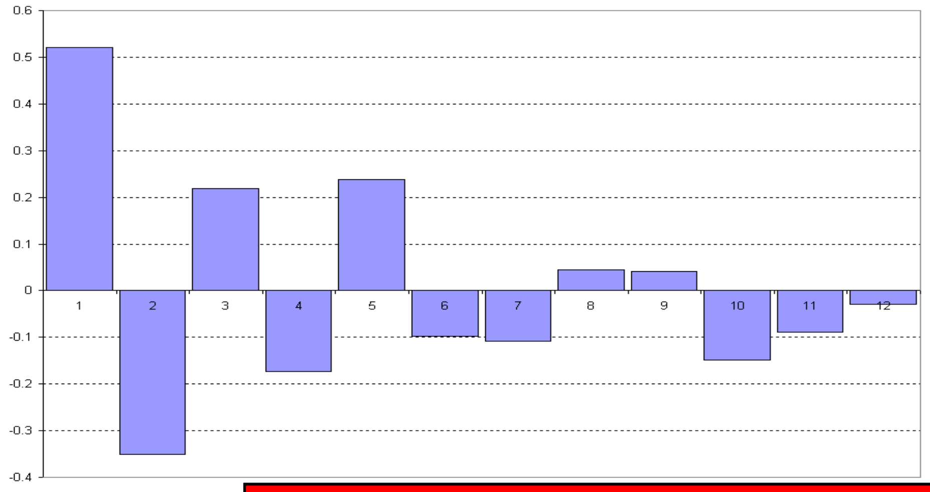
PhredTek, Inc.
logistics consulting services

ACF and PACF for Example 2



Autocorrelation

lag	acf	autocov	pacf
0	1	39.87136	
1	0.521368	20.78766	0.521368
2	0.016968	0.676541	-0.34999
3	-0.03031	-1.20856	0.219446
4	-0.02277	-0.90795	-0.17258
5	0.056143	2.238493	0.238639
6	0.11287	4.500291	-0.09695
7	-0.0443	-1.76624	-0.10824
8	-0.11836	-4.71911	0.044348
9	-0.00219	-0.08724	0.040939
10	-0.02815	-1.12222	-0.14916
11	-0.15265	-6.08634	-0.08952
12	-0.15601	-6.2205	-0.02996



Partial Autocorrelation

This is the classic ACF and PACF for an MA1 process

The First Differencing Model ARIMA(0,1,0)

$$Y_t = Y_{t-1} + c + a_t$$

$$Y_t = \text{sales at time t}$$

$$a_t = \text{error at time t}$$

$$c = \text{constant}$$

Introduce the lag operator

$$Y_t = BY_t + c + a_t$$

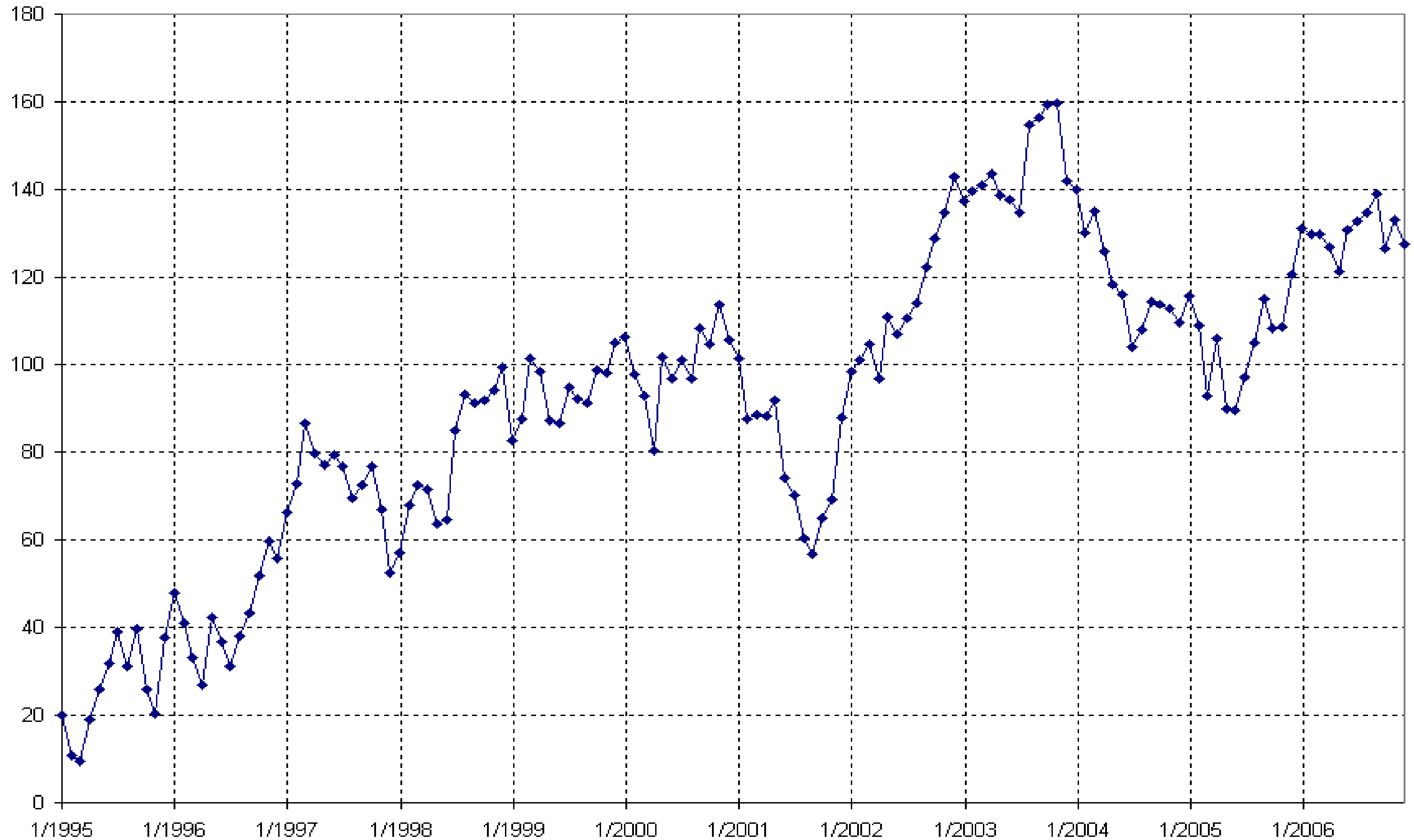
Combine Y_t terms

$$(1 - B)Y_t = c + a_t$$

PhredTek, Inc.
logistics consulting services

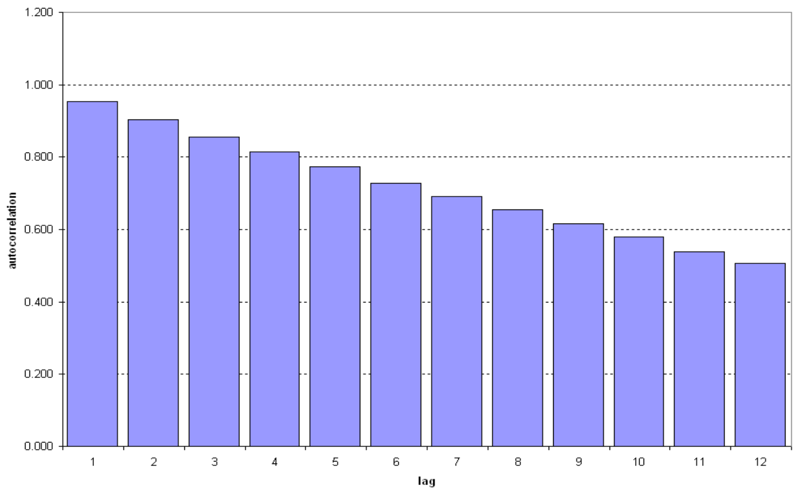
Example 3

144 monthly observations of a third sales series



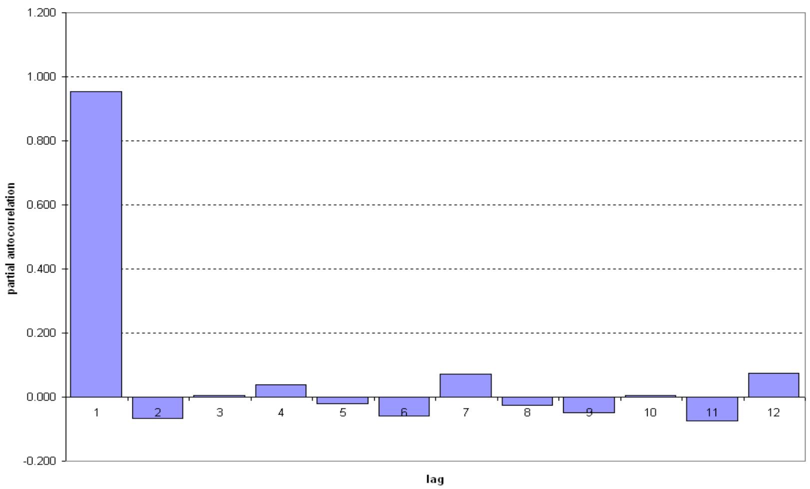
Phredlek, Inc.
logistics consulting services

ACF and PACF for Example 3



Autocorrelation

lag	ACF	autocov	PACF
0	1.000	1224.469	
1	0.953	1167.340	0.953
2	0.903	1105.561	-0.066
3	0.855	1047.143	0.005
4	0.814	996.179	0.039
5	0.772	945.901	-0.021
6	0.728	891.209	-0.060
7	0.691	846.435	0.071
8	0.655	802.325	-0.025
9	0.617	755.500	-0.047
10	0.580	710.328	0.005
11	0.539	659.485	-0.076
12	0.507	620.295	0.075



Partial Autocorrelation

This is the classic ACF and PACF for a first difference process

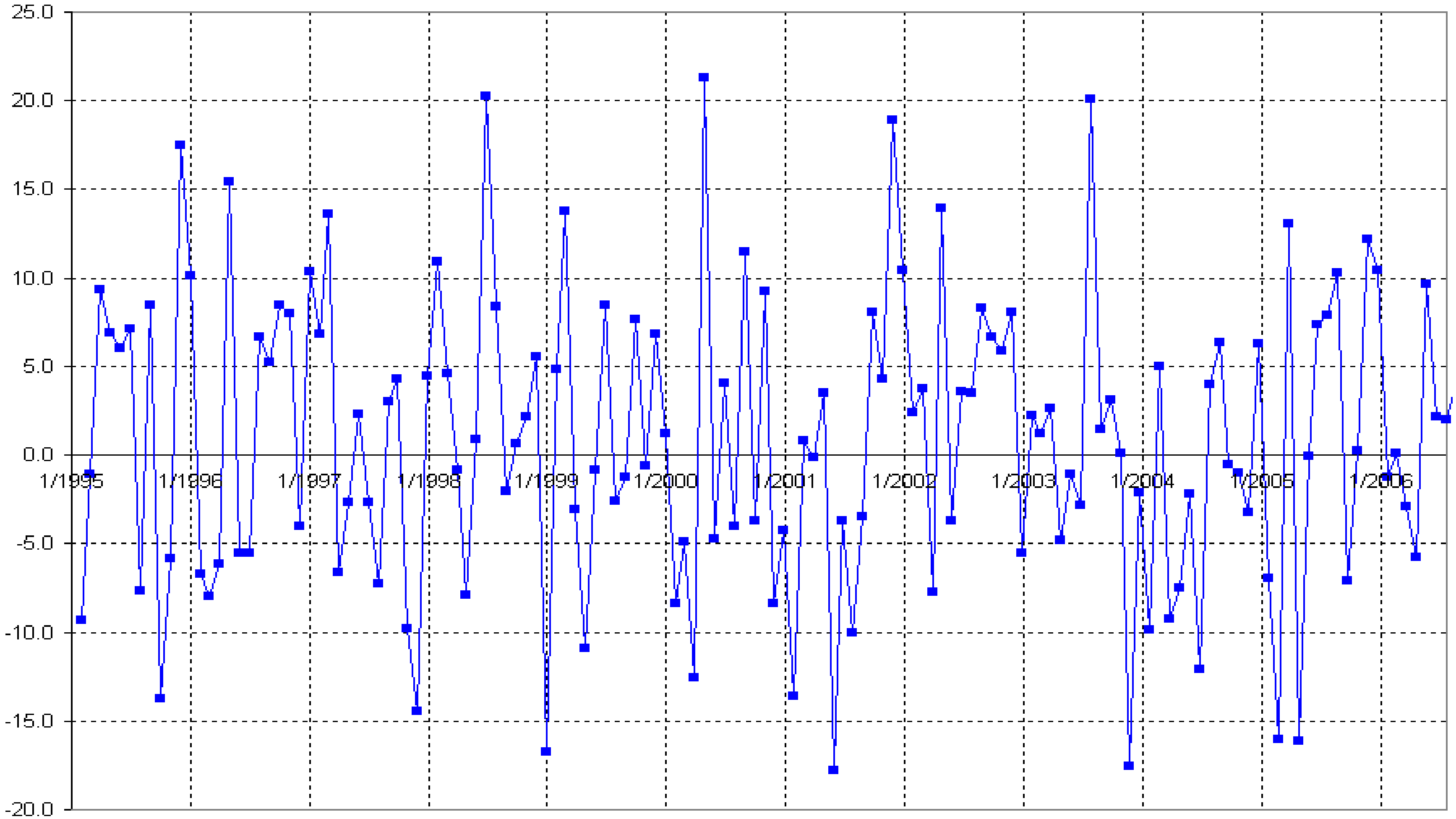
The First Differencing Model

If differencing is needed, we difference the entire series before building models. Here's how we difference example 3

date	sales	first differnce
1/1/1995	20.0	
2/1/1995	10.7	-9.3
3/1/1995	9.6	-1.1
4/1/1995	18.9	9.3
5/1/1995	25.8	6.9
6/1/1995	31.8	6.1
7/1/1995	39.0	7.2
8/1/1995	31.3	-7.7
9/1/1995	39.8	8.5
10/1/1995	26.0	-13.8
11/1/1995	20.2	-5.8
12/1/1995	37.7	17.5
1/1/1996	47.8	10.1
2/1/1996	41.1	-6.7
3/1/1996	33.1	-8.0
4/1/1996	27.0	-6.2
5/1/1996	42.4	15.4
6/1/1996	36.9	-5.5
7/1/1996	31.3	-5.6
8/1/1996	38.0	6.7
9/1/1996	43.2	5.3

Example 3 Differenced

143 Observations of the series differenced



Mixed Models

The AR1, MA1 model – ARIMA(1,0,1)

$$Y_t = c + \Phi_1 Y_{t-1} + a_t - \Theta_1 a_{t-1}$$

$$Y_t = \text{sales at time t}$$

$$a_t = \text{error at time t}$$

$$\Theta_1 = \text{moving average coefficient}$$

$$c = \text{constant}$$

$$\Phi_1 = \text{autoregressive coefficient}$$

PhredTek, Inc.
logistics consulting services

Mixed Models

The AR1, MA1 model – ARIMA(1,0,1)

Introduce the lag operator

$$Y_t = c + \Phi_1 B Y_t + a_t - \Theta_1 B a_t$$

Combine the Y_t and a_t terms

$$(1 - \Phi_1 B) Y_t = c + (1 - \Theta_1 B) a_t$$

Divide by lag polynomial to get standard form

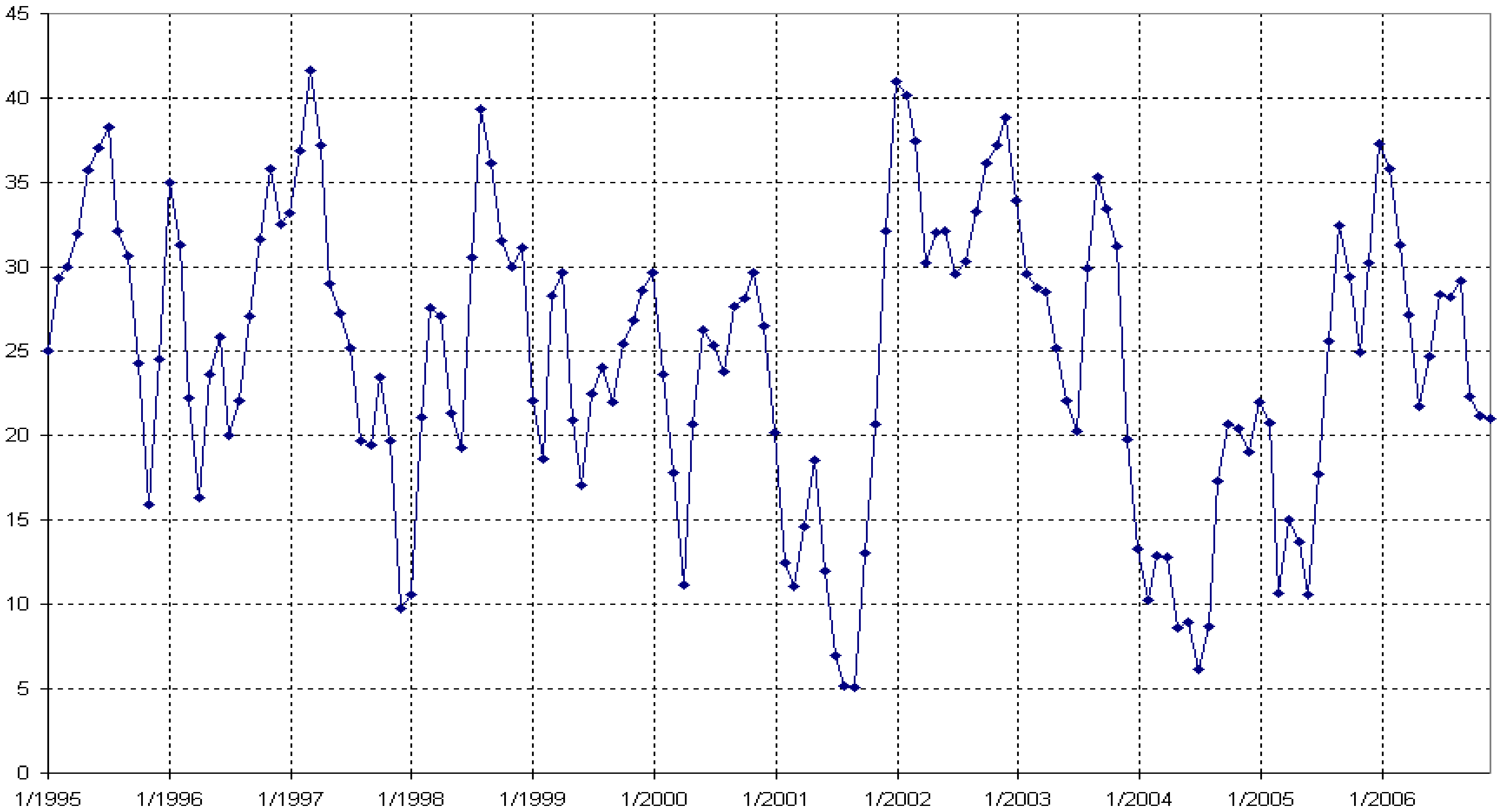
$$Y_t = \frac{c}{(1 - \Phi_1)} + \frac{(1 - \Theta_1 B)}{(1 - \Phi_1 B)} a_t = c' + \frac{(1 - \Theta_1 B)}{(1 - \Phi_1 B)} a_t$$

$$c' = \text{adjusted constant}$$

PhredTek, Inc.
logistics consulting services

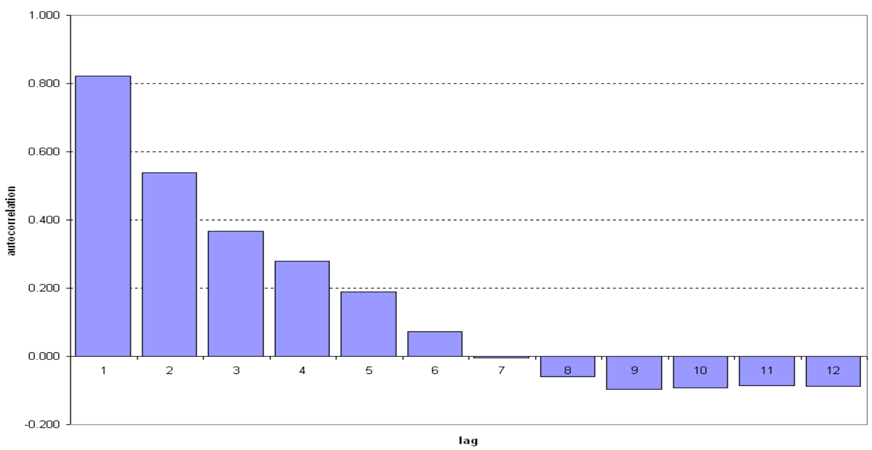
Example 4

144 monthly observations of a fourth sales series



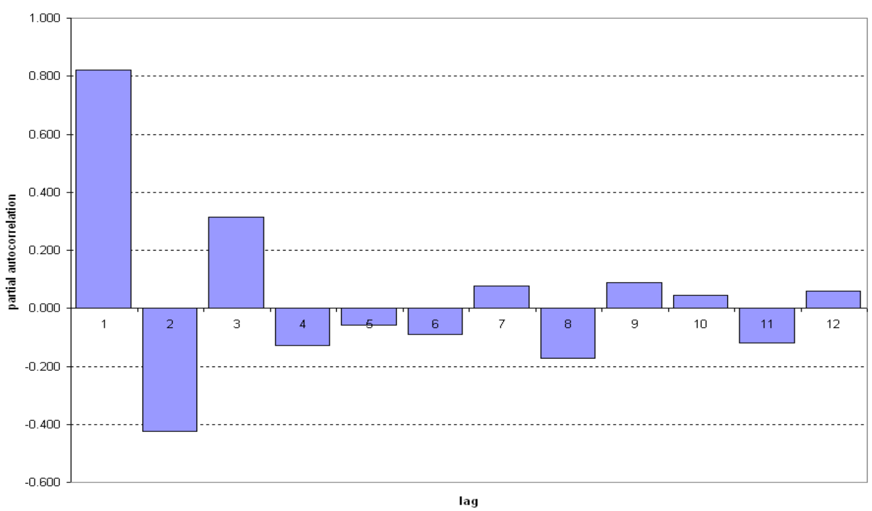
Phredlek, Inc.
logistics consulting services

ACF and PACF for Example 3



Autocorrelation

lag	ACF	autocov	PACF
0	1.000	72.613	
1	0.822	59.688	0.822
2	0.539	39.104	-0.423
3	0.366	26.591	0.315
4	0.279	20.256	-0.129
5	0.188	13.679	-0.056
6	0.073	5.274	-0.089
7	-0.005	-0.368	0.077
8	-0.058	-4.242	-0.172
9	-0.097	-7.027	0.088
10	-0.093	-6.738	0.045
11	-0.086	-6.254	-0.118
12	-0.089	-6.427	0.060



Partial Autocorrelation

This is the classic ACF and PACF for an AR1, MA1 process

The Seasonal AR Model

Autoregressive model with seasonal lag

ARIMA(1,0,0)(0,0,0)

$$Y_t = c + \Phi_1 Y_{t-12} + a_t$$

$$Y_t = \text{sales at time t}$$

$$a_t = \text{error at time t}$$

$$\Phi_1 = \text{autoregressive coefficient}$$

$$c = \text{constant}$$

PhredTek, Inc.
logistics consulting services

The Seasonal AR Model

Introduce the lag operator

$$Y_t = c + \Phi_1 B^{12} Y_t + a_t$$

Combine the Y_t terms

$$(1 - \Phi_1 B^{12}) Y_t = c + a_t$$

Divide by lag polynomial to get standard form

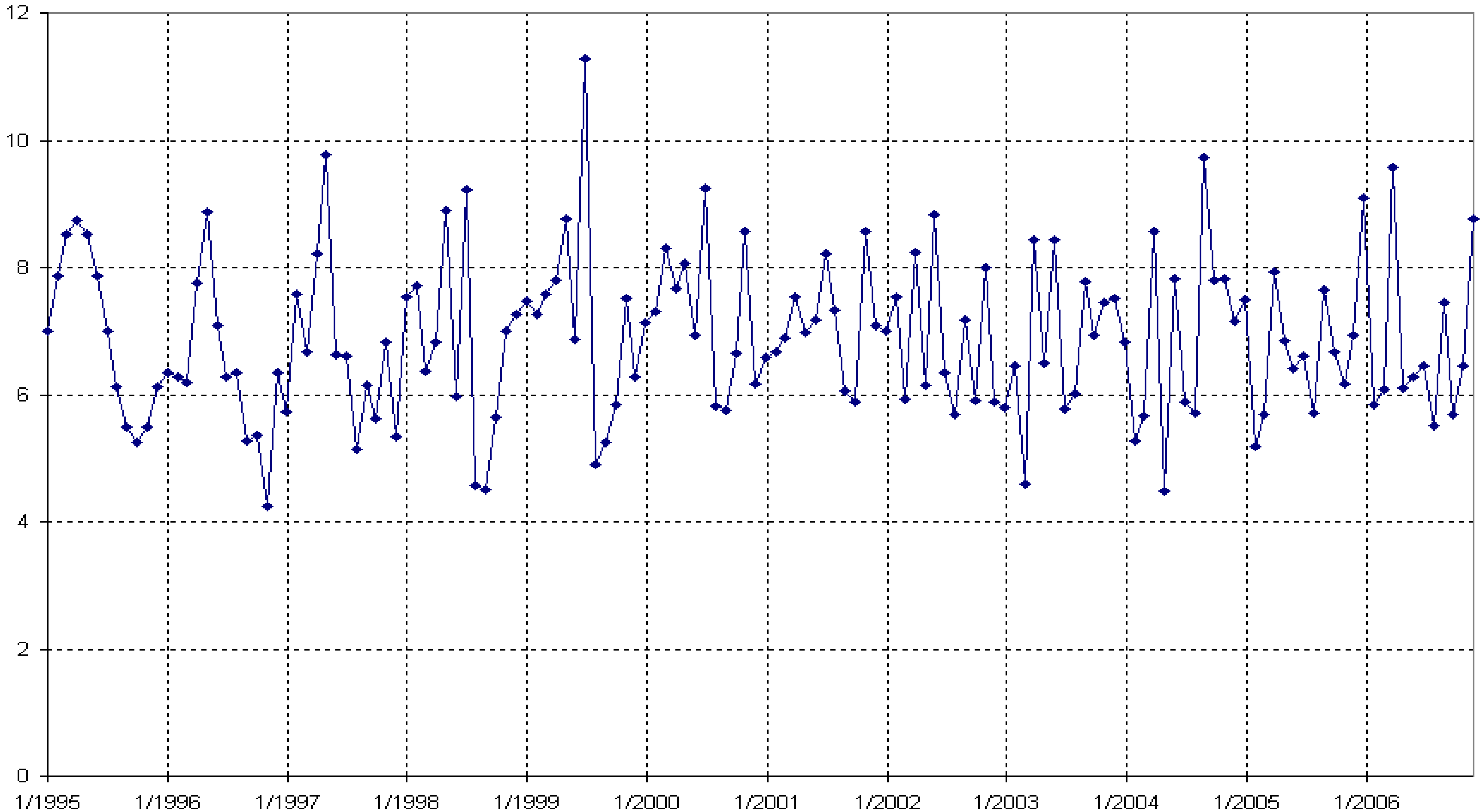
$$Y_t = \frac{c}{(1 - \Phi_1)} + \frac{1}{(1 - \Phi_1 B^{12})} a_t = c' + \frac{1}{(1 - \Phi_1 B^{12})} a_t$$

$$c' = \text{adjusted constant}$$

PhredTek, Inc.
logistics consulting services

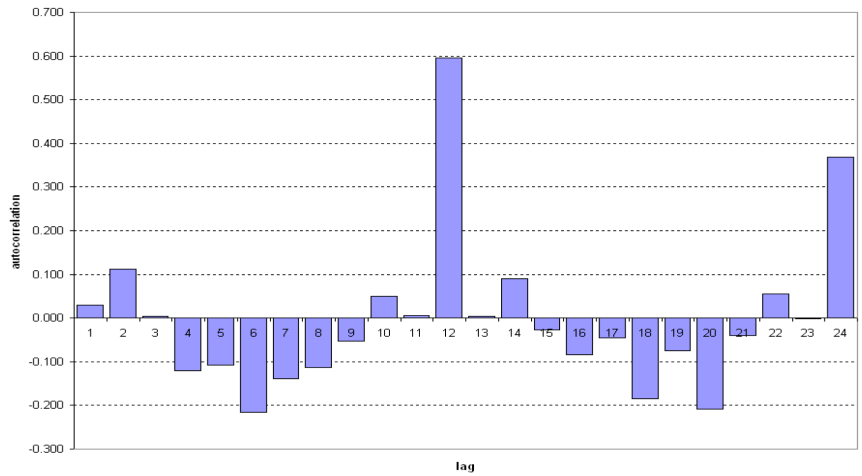
Example 5

144 monthly observations of a fifth sales series

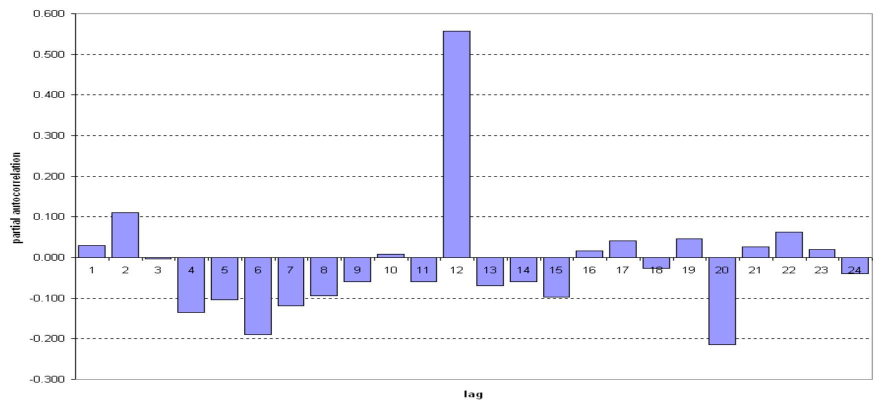


Phredlek, Inc.
logistics consulting services

ACF and PACF for Example 5



Autocorrelation



Partial Autocorrelation

lag	ACF	autocov	PACF
0	1.000	1.556	
1	0.030	0.047	0.030
2	0.112	0.174	0.111
3	0.003	0.005	-0.003
4	-0.120	-0.187	-0.134
5	-0.107	-0.166	-0.103
6	-0.216	-0.335	-0.189
7	-0.138	-0.215	-0.119
8	-0.113	-0.176	-0.093
9	-0.052	-0.081	-0.060
10	0.050	0.078	0.008
11	0.006	0.009	-0.060
12	0.596	0.928	0.557
13	0.005	0.008	-0.070
14	0.090	0.140	-0.059
15	-0.027	-0.042	-0.097
16	-0.083	-0.129	0.017
17	-0.046	-0.071	0.042
18	-0.185	-0.289	-0.027
19	-0.075	-0.116	0.046
20	-0.208	-0.323	-0.215
21	-0.040	-0.062	0.027
22	0.056	0.087	0.063
23	-0.002	-0.003	0.020
24	0.368	0.573	-0.040

This is the classic ACF and PACF for a seasonal AR process

The Seasonal MA Model

Moving average model with seasonal lag
ARIMA(0,0,1)(0,0,0)

$$Y_t = c + a_t - \Theta_1 a_{t-12}$$

$$Y_t = \text{sales at time t}$$

$$a_t = \text{error at time t}$$

$$\Theta_1 = \text{moving average coefficient}$$

$$c = \text{constant}$$

PhredTek, Inc.
logistics consulting services

The Seasonal MA Model

Introduce the lag operator

$$Y_t = c + a_t - \Theta_1 B^{12} a_t$$

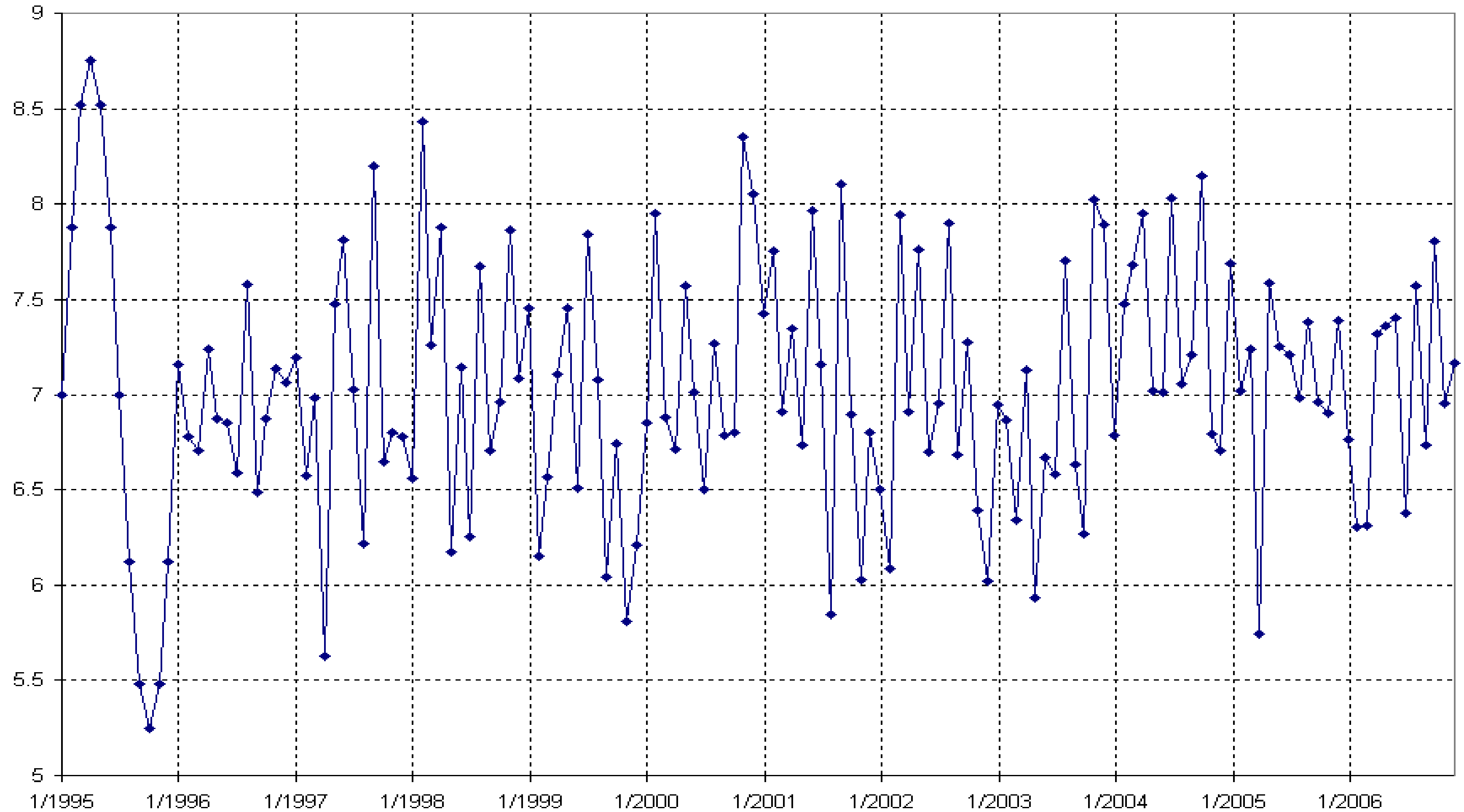
Combine the a_t terms

$$Y_t = c + (1 - \Theta_1 B^{12}) a_t$$

PhredTek, Inc.
logistics consulting services

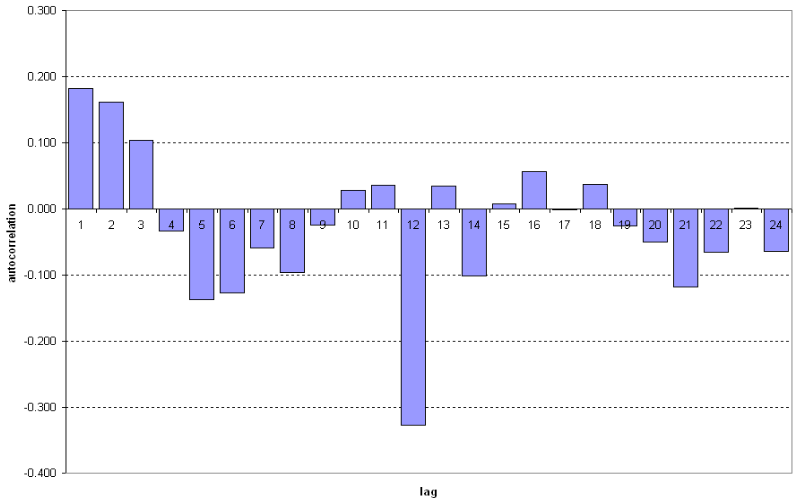
Example 6

144 monthly observations of a sixth sales series

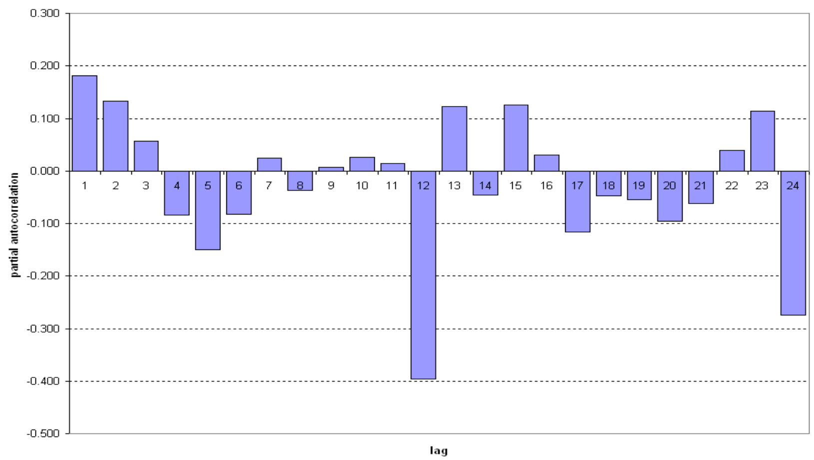


PhredTek, Inc.
logistics consulting services

ACF and PACF for Example 6



Autocorrelation



Partial Autocorrelation

lag	ACF	autocov	PACF
0	1.000	0.458	
1	0.182	0.083	0.182
2	0.162	0.074	0.133
3	0.104	0.047	0.057
4	-0.034	-0.015	-0.084
5	-0.137	-0.063	-0.150
6	-0.127	-0.058	-0.082
7	-0.059	-0.027	0.025
8	-0.096	-0.044	-0.037
9	-0.025	-0.011	0.007
10	0.028	0.013	0.027
11	0.036	0.016	0.014
12	-0.327	-0.150	-0.396
13	0.035	0.016	0.123
14	-0.102	-0.047	-0.045
15	0.007	0.003	0.125
16	0.056	0.026	0.030
17	-0.001	-0.001	-0.116
18	0.037	0.017	-0.048
19	-0.025	-0.012	-0.055
20	-0.049	-0.023	-0.096
21	-0.118	-0.054	-0.062
22	-0.065	-0.030	0.039
23	0.001	0.000	0.113
24	-0.064	-0.029	-0.275

This is the classic ACF and PACF for an seasonal MA process

The Seasonal Differencing Model

ARIMA(0,1,0)(0,0,0)

$$Y_t = Y_{t-12} + c + a_t$$

$$Y_t = \text{sales at time t}$$

$$a_t = \text{error at time t}$$

$$c = \text{constant}$$

Introduce the lag operator

$$Y_t = B^{12} Y_t + c + a_t$$

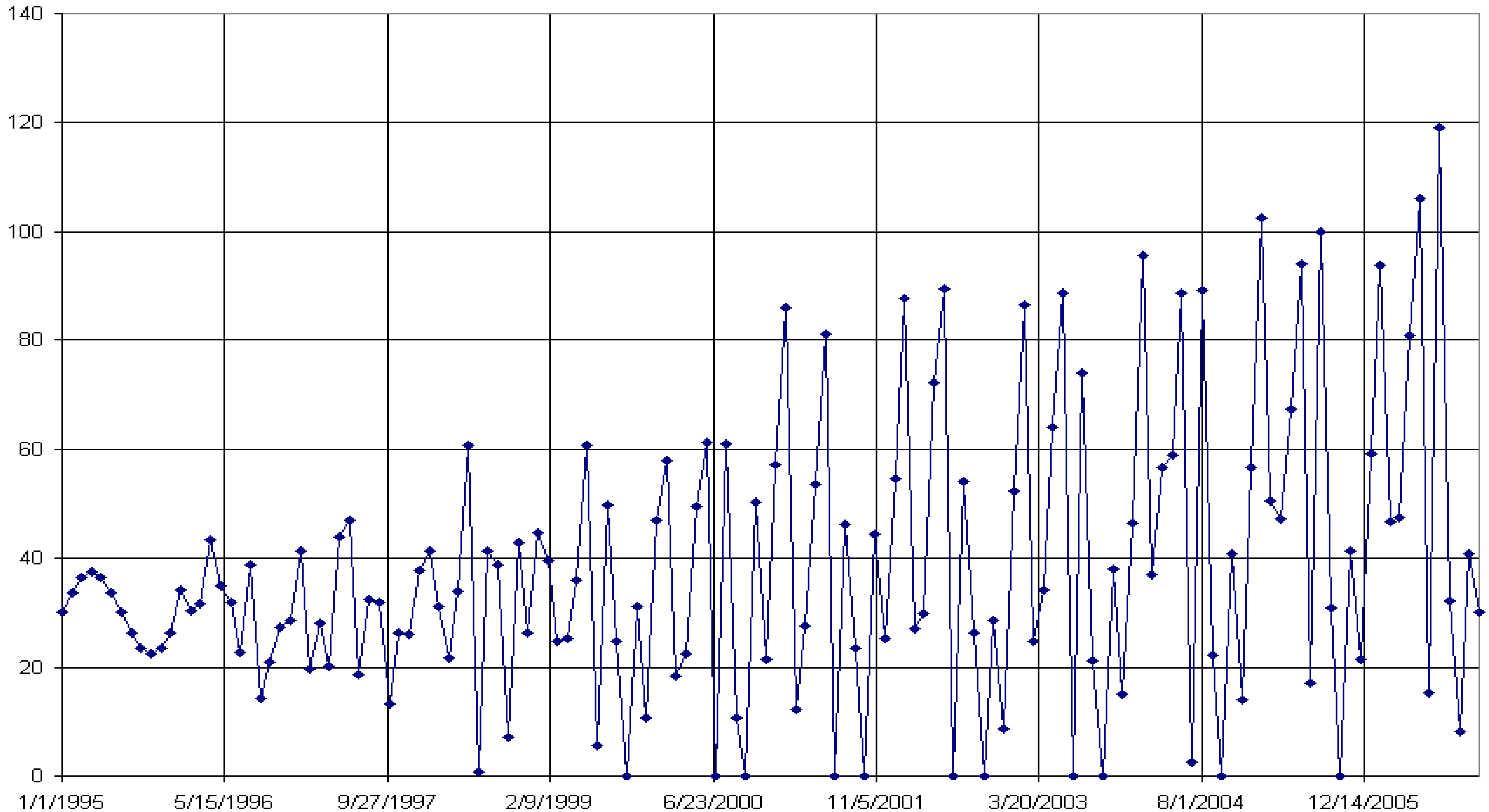
Combine Y_t terms

$$(1 - B^{12}) Y_t = c + a_t$$

PhredTek, Inc.
logistics consulting services

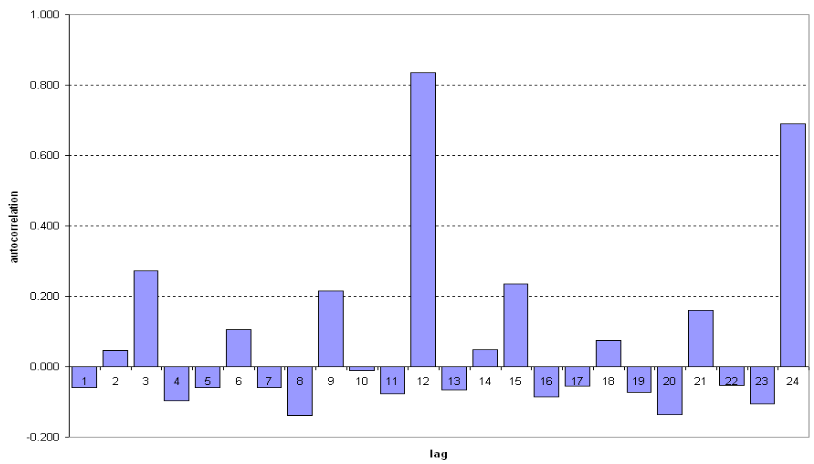
Example 7

144 monthly observations of a seventh sales series

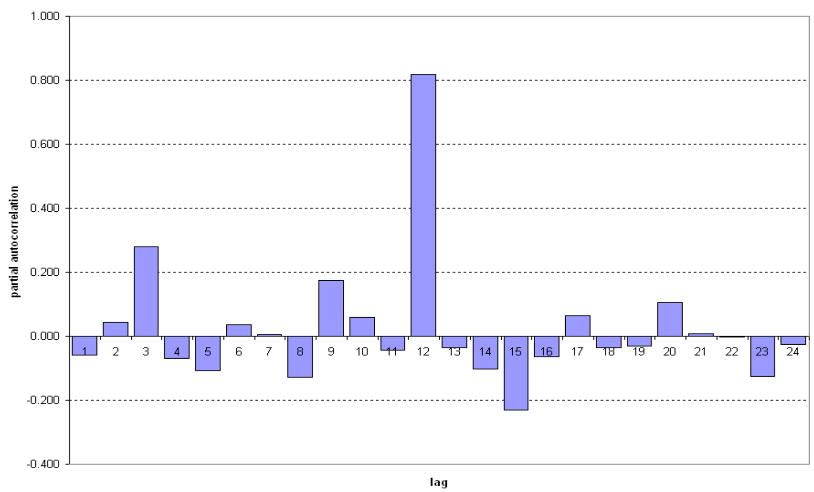


Phredlek, Inc.
logistics consulting services

ACF and PACF for Example 7



Autocorrelation



Partial Autocorrelation

lag	ACF	autocov	PACF
0	1.000	670.879	
1	-0.059	-39.556	-0.059
2	0.046	30.915	0.043
3	0.272	182.407	0.278
4	-0.096	-64.504	-0.070
5	-0.059	-39.547	-0.107
6	0.106	71.103	0.036
7	-0.060	-40.256	0.006
8	-0.139	-93.224	-0.129
9	0.215	144.162	0.173
10	-0.010	-6.693	0.060
11	-0.076	-51.256	-0.043
12	0.835	560.131	0.817
13	-0.065	-43.569	-0.036
14	0.047	31.760	-0.104
15	0.235	157.590	-0.231
16	-0.086	-57.722	-0.064
17	-0.054	-36.538	0.065
18	0.075	50.089	-0.036
19	-0.073	-48.704	-0.030
20	-0.137	-91.741	0.106
21	0.159	106.930	0.007
22	-0.052	-34.726	-0.003
23	-0.105	-70.430	-0.126
24	0.691	463.677	-0.025

This is the classic ACF and PACF for a seasonal differencing process

The General ARMA Model Autoregressive Factors

Autoregressive polynomial

$$\Phi_0(B) = 1 - \Phi_{0,1}B - \Phi_{0,2}B^2 - \Phi_{0,3}B^3 - \ldots$$

Seasonal autoregressive polynomial

$$\Phi_1(B) = 1 - \Phi_{1,1}B^s - \Phi_{1,2}B^{2s} - \Phi_{1,3}B^{3s} - \ldots$$

$$s = \text{seasonality - 12 for months, 4 for quarters, etc.}$$

PhredTek, Inc.
logistics consulting services

The General ARMA Model Moving Average Factors

The moving average polynomial

$$\Theta_0(B) = 1 - \Theta_{0,1}B - \Theta_{0,2}B^2 - \Theta_{0,3}B^3 - \ldots$$

Seasonal moving average polynomial

$$\Theta_1(B) = 1 - \Theta_{1,1}B^s - \Theta_{1,2}B^{2s} - \Theta_{1,3}B^{3s} - \ldots$$

$$s = \text{seasonality - 12 for months, 4 for quarters, etc.}$$

PhredTek, Inc.
logistics consulting services

The General ARMA Model
Putting it together

ARIMA$(m_0,0,m_1)(n_0,0,n_1)$

Express Y_t in terms of the polynomials

$$Y_t = \frac{c}{\Phi_0(B)\Phi_1(B)} + \frac{\Theta_0(B)\Theta_1(B)}{\Phi_0(B)\Phi_1(B)} a_t$$

$$= c' + \frac{\Theta_0(B)\Theta_1(B)}{\Phi_0(B)\Phi_1(B)} a_t$$

Basic MA

Seasonal MA

Basic AR

Seasonal AR

$$c' = \text{adjusted constant}$$

I do not include differencing in the general model because if differencing is needed, we do it up front before modeling

The Restricted Transfer Function Model

$$Y_t = c + \sum_{i=1}^{n} \omega_i X_{i,t} + a_t$$

$$X_i = \text{ith independent (input) variable}$$

It looks like a regression doesn't it - - It is

Introduce a one period lead, one period lag to the independent (input) variables

$$Y_t = c + \sum_{i=1}^{n} (\omega_{i,-1} X_{i,t+1} + \omega_{i,0} X_{i,t} + \omega_{i,1} X_{i,t-1}) + a_t$$

PhredTek, Inc.
logistics consulting services

The RestrictedTransfer Function Model

Introduce the lag operator

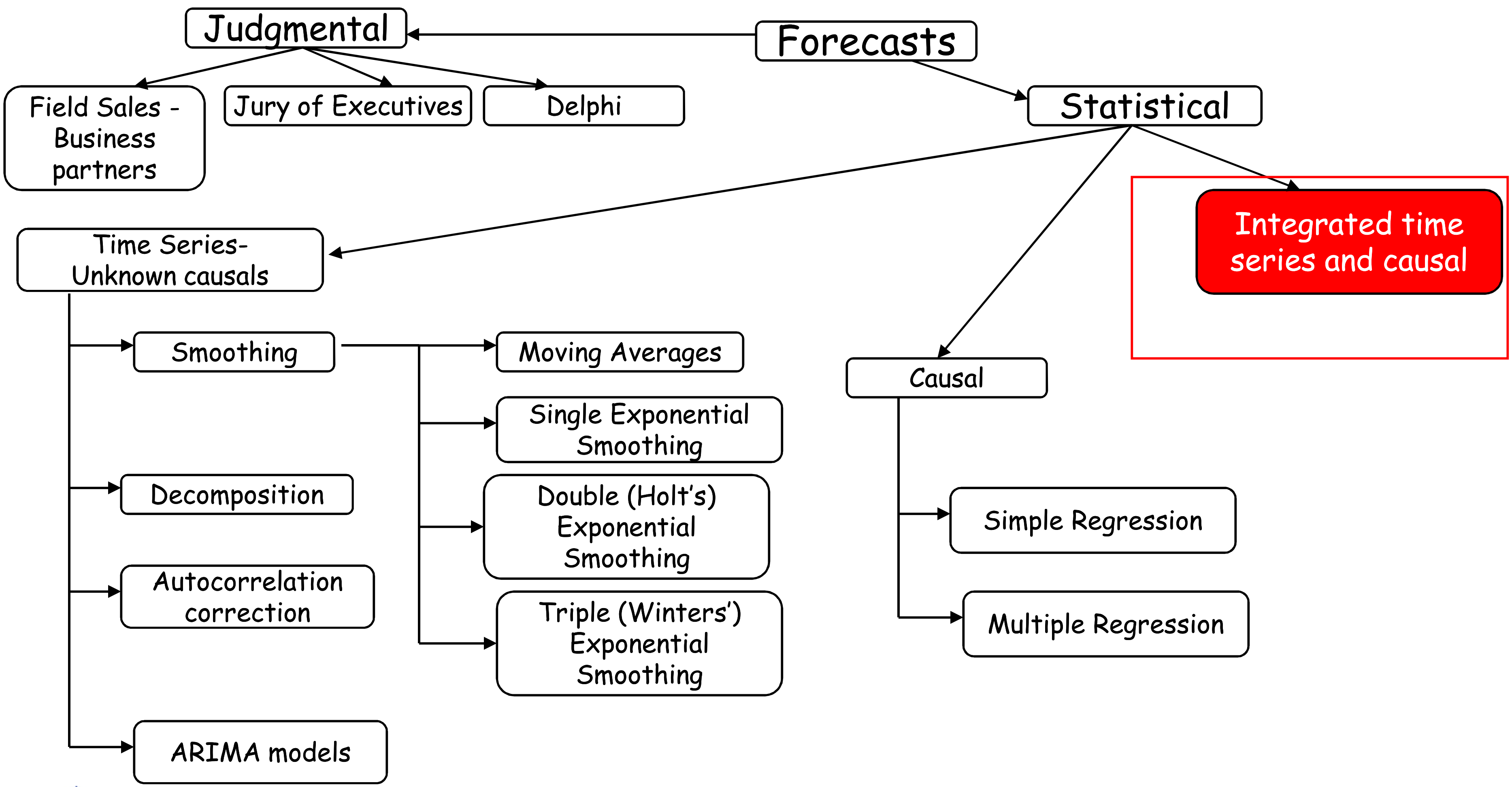
$$Y_t = c + \sum_{i=1}^{n} (\omega_{i,-1} B^{-1} X_{i,t} + \omega_{i,0} X_{i,t} + \omega_{i,1} B X_{i,t}) + a_t$$

Put $X_{i,t}$ terms in a standard summation

$$Y_t = c + \sum_{i=1}^{n} \sum_{l=-1}^{1} \omega_{i,l} B^{l} X_{i,t} + a_t$$

PhredTek, Inc.
logistics consulting services

Overview of Forecasting Methods



PhredTek, Inc.
logistics consulting services

The General Model
The ARIMA Transfer Function Model

General ARMA Model

$$Y_t = c' + \frac{\Theta_0(B)\Theta_1(B)}{\Phi_0(B)\Phi_1(B)} a_t$$

General Transfer Function Model

$$Y_t = c + \sum_{i=1}^{n} \sum_{l=-1}^{1} \omega_{i,l} B^l X_{i,t} + a_t$$

Just stick them together

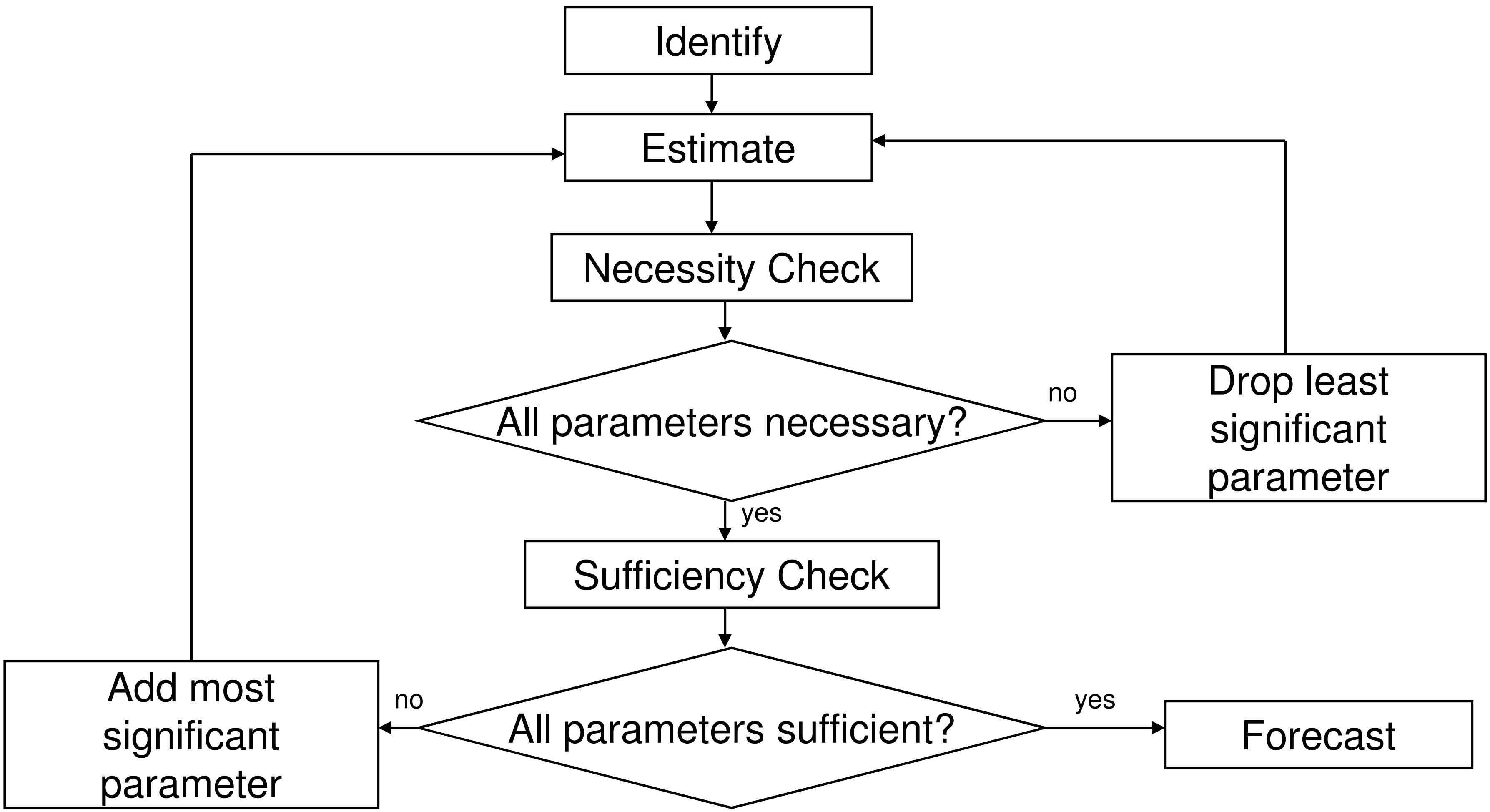
$$Y_t = c' + \sum_{i=1}^{n} \sum_{l=-1}^{1} \omega_{i,l} B^l X_{i,t} + \frac{\Theta_0(B)\Theta_1(B)}{\Phi_0(B)\Phi_1(B)} a_t$$

I do not include differencing in the general model because if differencing is needed, we do it up front before modeling

The Modeling Process

PhredTek, Inc.
logistics consulting services

The Model Building Process

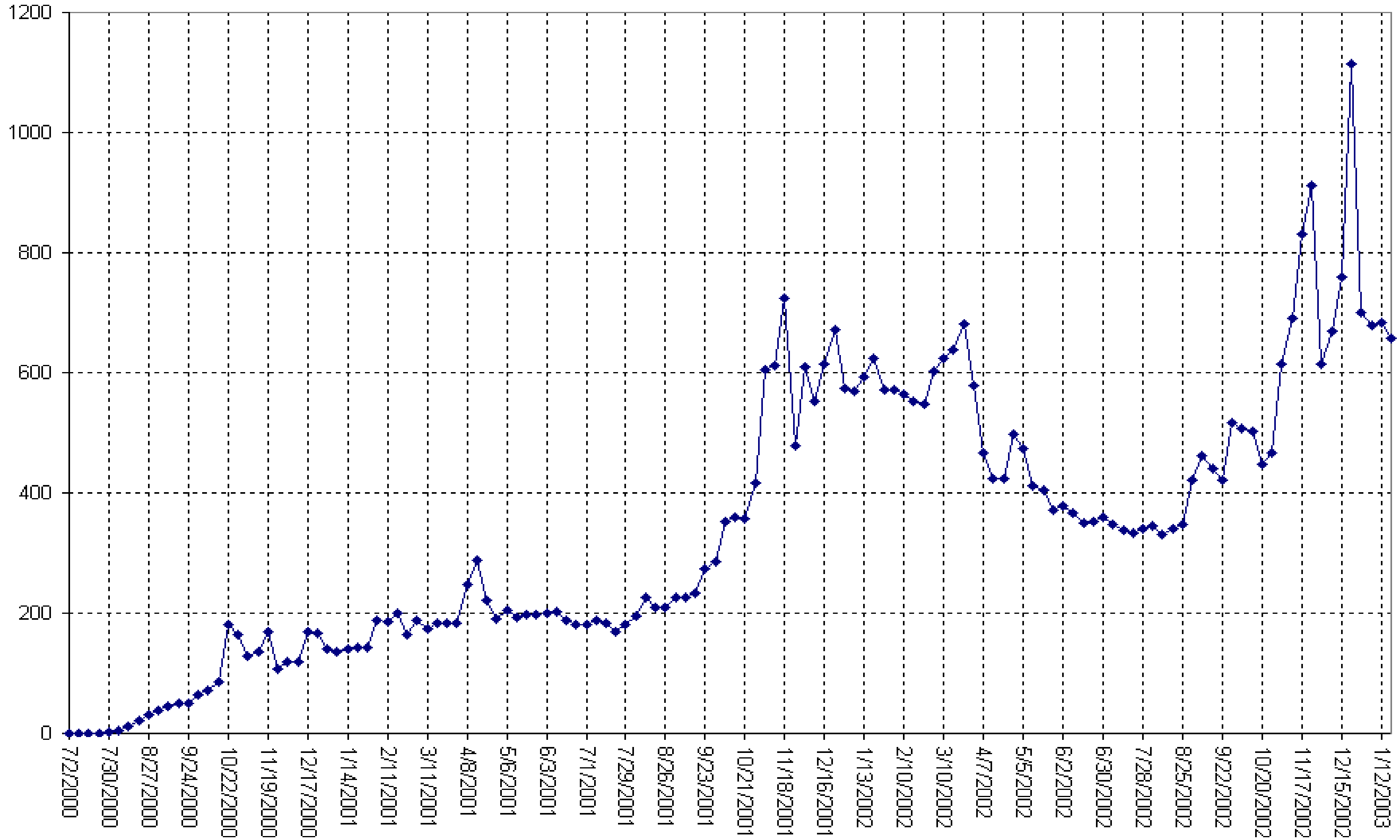


PhredTek, Inc.
logistics consulting services

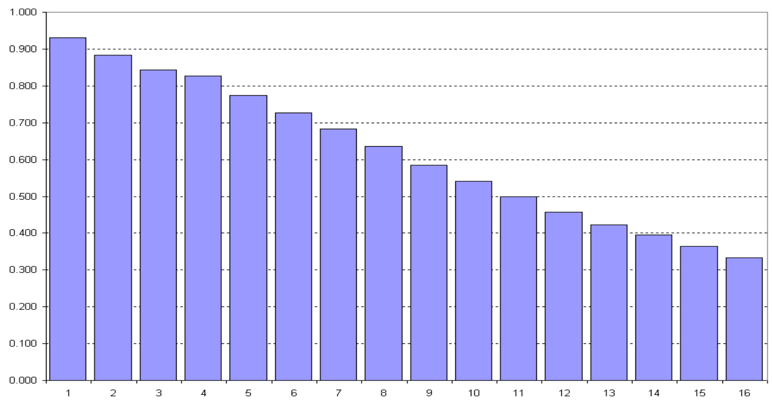
General Mills Biscuits Example

PhredTek, Inc.
logistics consulting services

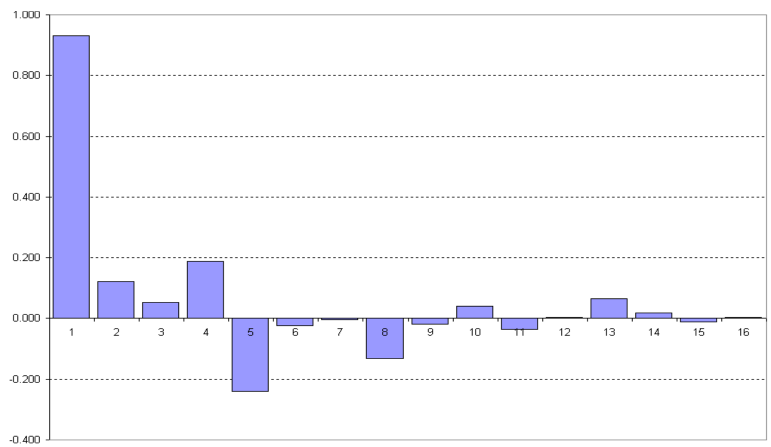
Sales Of General Mills Biscuits Weekly Sales



ACF and PACF for General Mills Biscuits



Autocorrelation



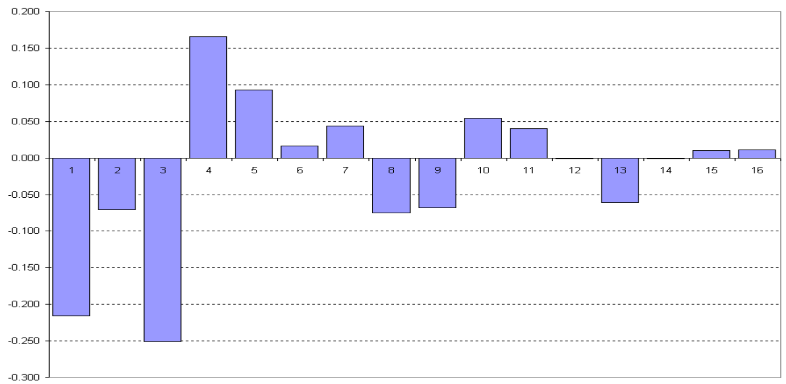
Partial Autocorrelation

lag	acf	pacf
1	0.931	0.931
2	0.884	0.122
3	0.843	0.053
4	0.827	0.188
5	0.774	-0.239
6	0.727	-0.023
7	0.683	-0.003
8	0.636	-0.131
9	0.584	-0.019
10	0.541	0.040
11	0.500	-0.037
12	0.458	0.004
13	0.422	0.066
14	0.395	0.019
15	0.364	-0.011
16	0.333	0.003

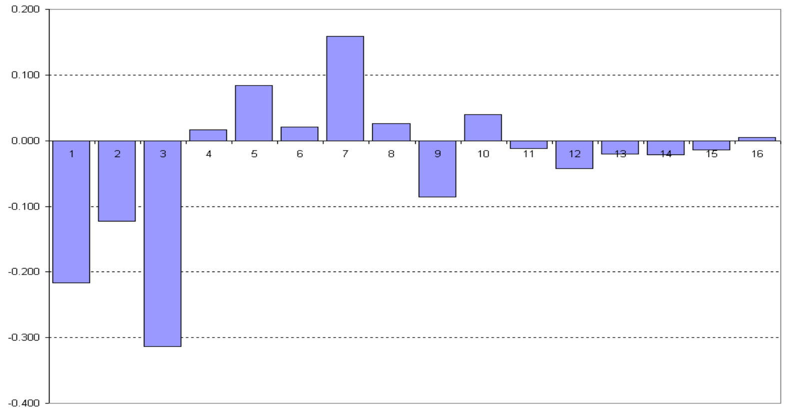
This is the classic ACF and PACF for a differencing process so Autobox differences the series

PhredTek, Inc.
logistics consulting services

ACF and PACF for General Mills Biscuits – Differenced Series



Autocorrelation



Partial Autocorrelation

lag	acf	pacf
1	-0.216	-0.216
2	-0.071	-0.123
3	-0.251	-0.314
4	0.166	0.016
5	0.093	0.084
6	0.016	0.021
7	0.044	0.159
8	-0.075	0.026
9	-0.068	-0.086
10	0.054	0.040
11	0.040	-0.012
12	-0.001	-0.043
13	-0.061	-0.020
14	-0.001	-0.021
15	0.010	-0.014
16	0.011	0.005

Autobox identifies an AR3 as the model that best fits this autocorrelation

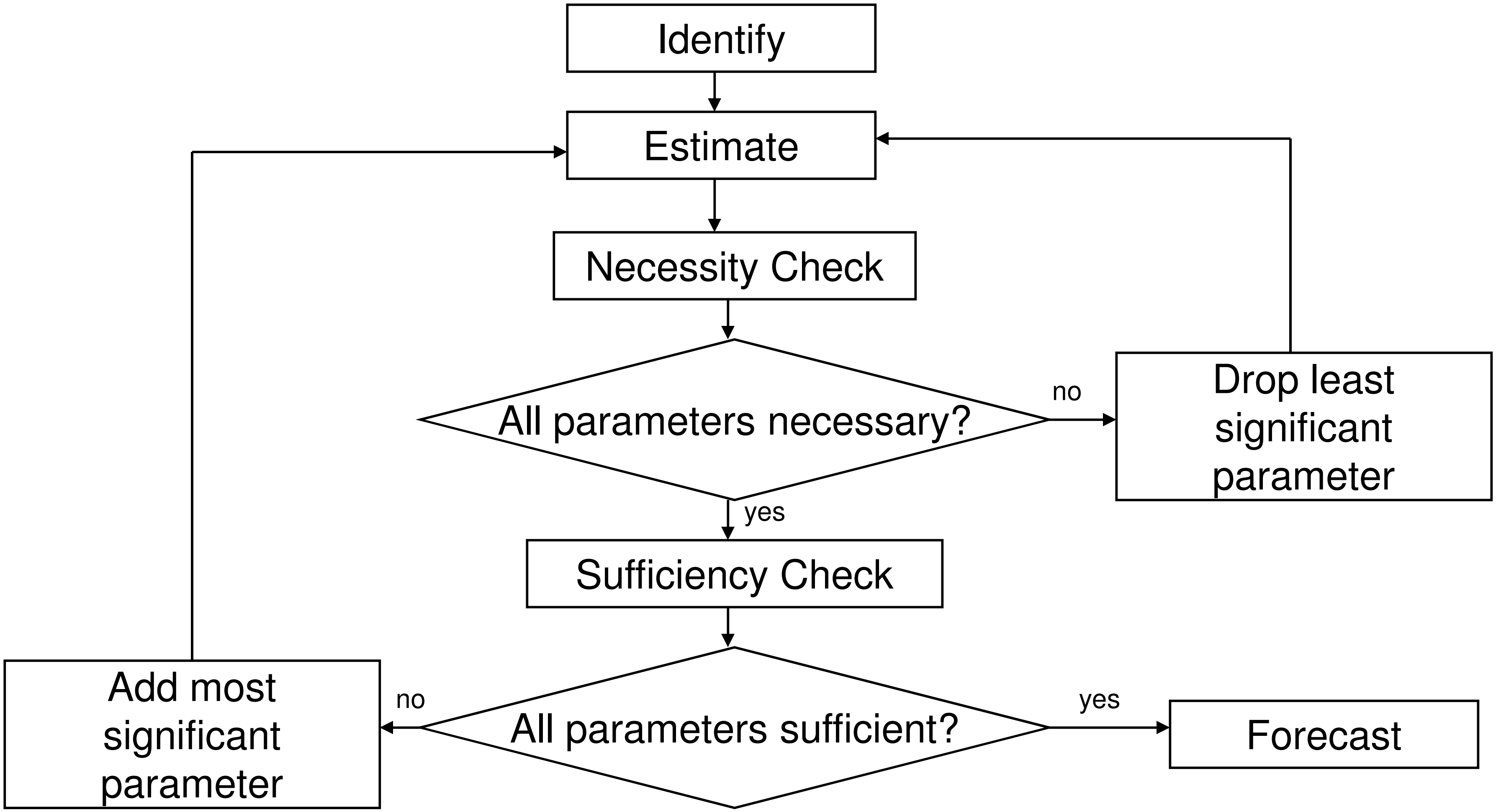
PhredTek, Inc.
logistics consulting services

General Mills Biscuits Identification

Autobox Identified Model

	MODEL COMPONENT	LAG	COEFF
#		(BOP)	
	Differencing	1	
1	CONSTANT		5.0
2	Autoregressive-Factor # 1	1	-0.3
3		2	-0.2
4		3	-0.3

The Model Building Process



PhredTek, Inc.
logistics consulting services

Notes on Estimation of the General Model

- Models other than pure autoregressive are non-linear
- Cannot use standard least squares regression to estimate them
- Still want to find the model that minimizes the sum of squares of the errors
- Marquardt non-linear least squares algorithm is the most commonly used.

General Mills Biscuits Estimation

Model estimated by Autobox

#	MODEL COMPONENT	LAG (BOP)	COEFF	STD ERROR	P VALUE	T VALUE
	Differencing	1				
1	CONSTANT		9.3	6.0	0.126	1.540
2	Autoregressive-Factor # 1	1	-0.3	0.1	0.001	-3.380
3		2	-0.2	0.1	0.020	-2.350
4		3	-0.3	0.1	0.000	-3.780

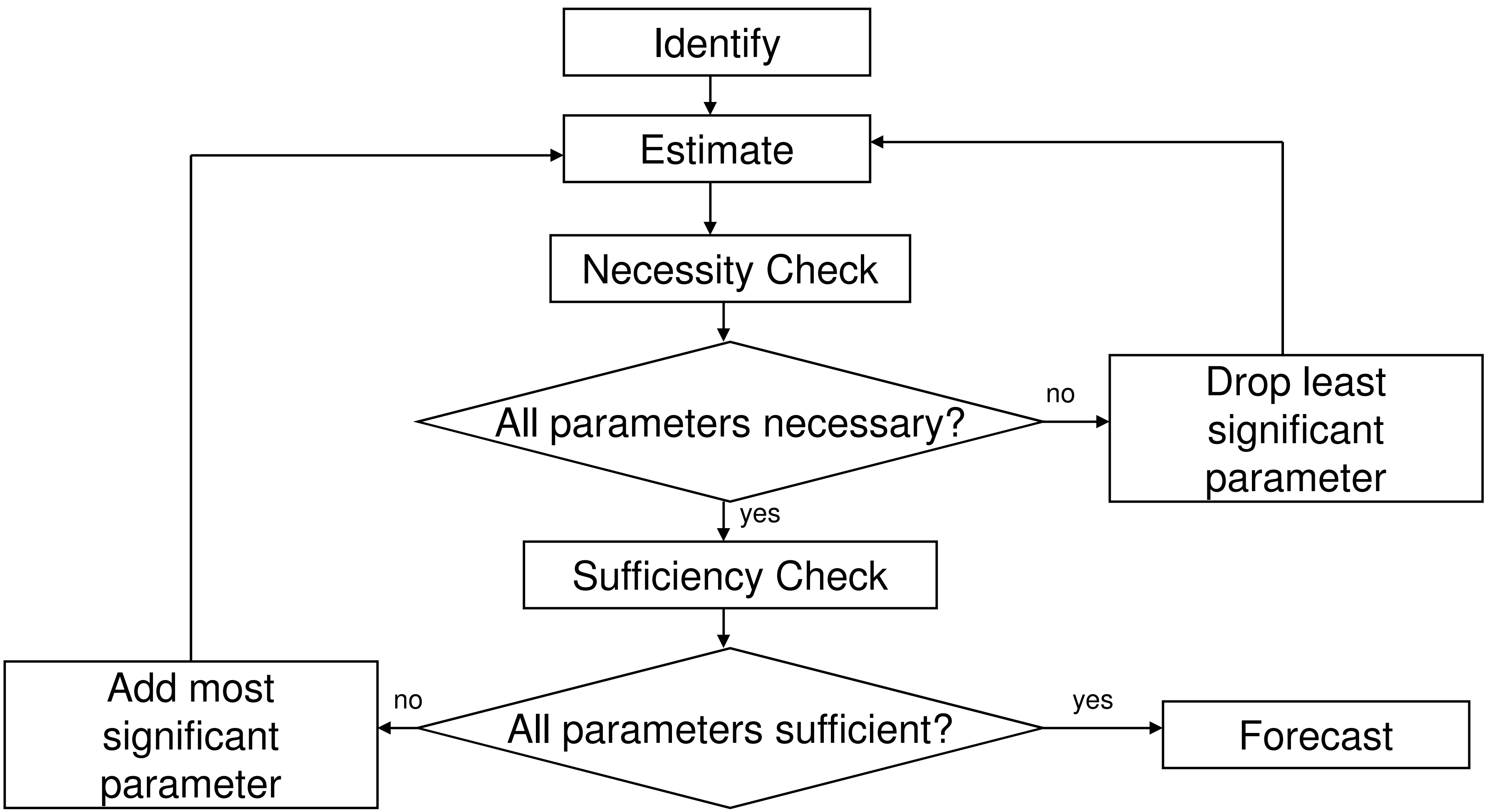
$$Y_t = 5.16 + \frac{1}{(1 + .3B + .2B^2 + .3B^3)} a_t$$

Necessity check

Are the P values of all parameters less than .05? Yes except for the constant. However, it is an ongoing debate in the industry as to whether or not the constant should ever be dropped. Autobox chooses not to drop it. Therefore, the necessity check passes.

PhredTek, Inc.
logistics consulting services

The Model Building Process



PhredTek, Inc.

General Mills Biscuits

Sufficiency Check

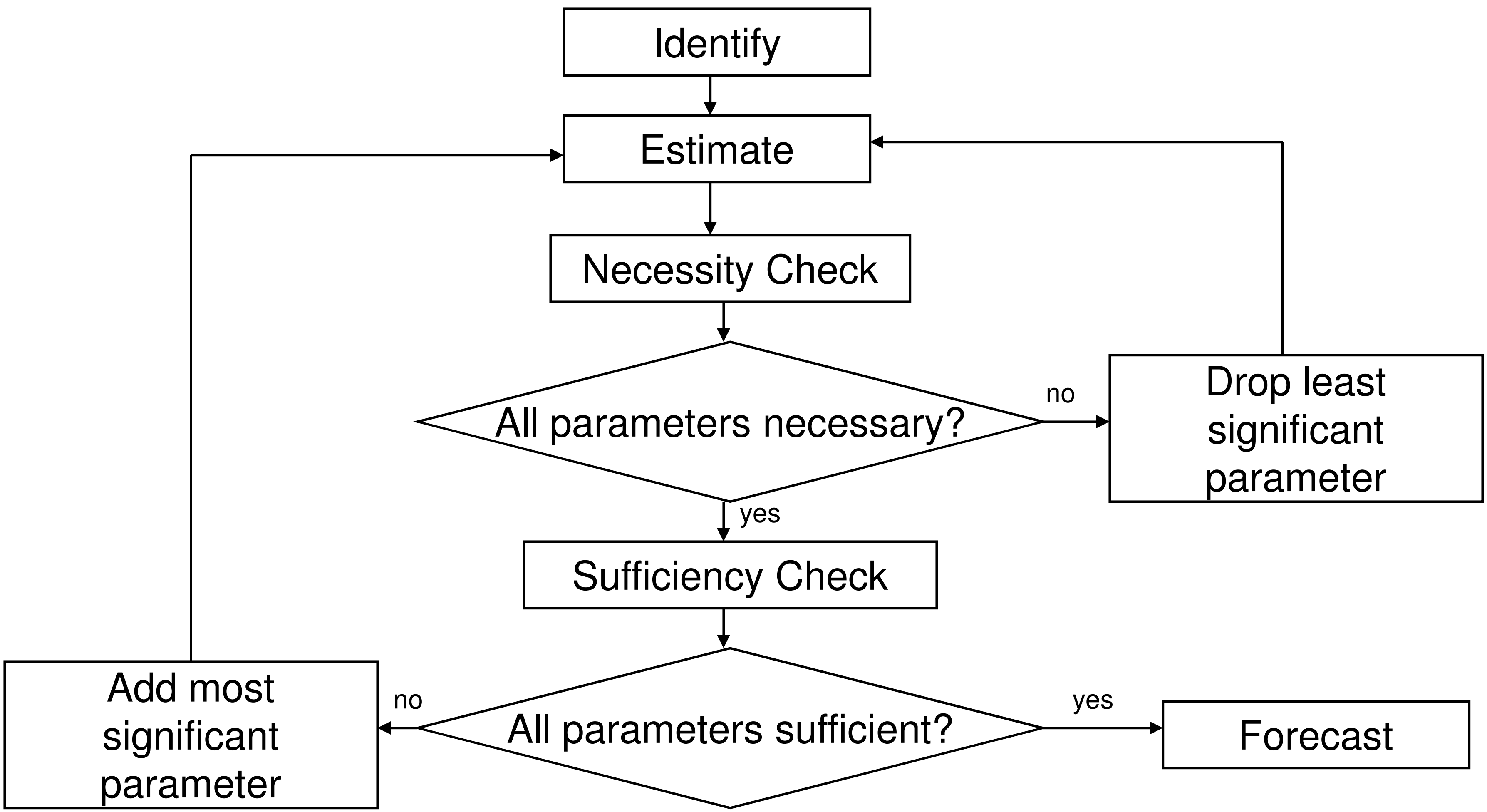
lag	acf	pacf
1	0.00	0.00
2	0.03	0.03
3	0.02	0.02
4	0.07	0.07
5	0.10	0.10
6	0.00	0.00
7	0.13	0.13
8	-0.05	-0.06
9	-0.07	-0.09
10	0.04	0.03
11	0.02	0.00
12	-0.01	-0.03
13	-0.04	-0.02
14	-0.03	-0.04
15	-0.01	-0.01

Autocorrelation and partial autocorrelation of errors of the fit

All the acfs and pacfs of the error are small so it indicates that our model has corrected for all of the autocorrelation in the original series. That means that the model is sufficient.

PhredTek, Inc.
logistics consulting services

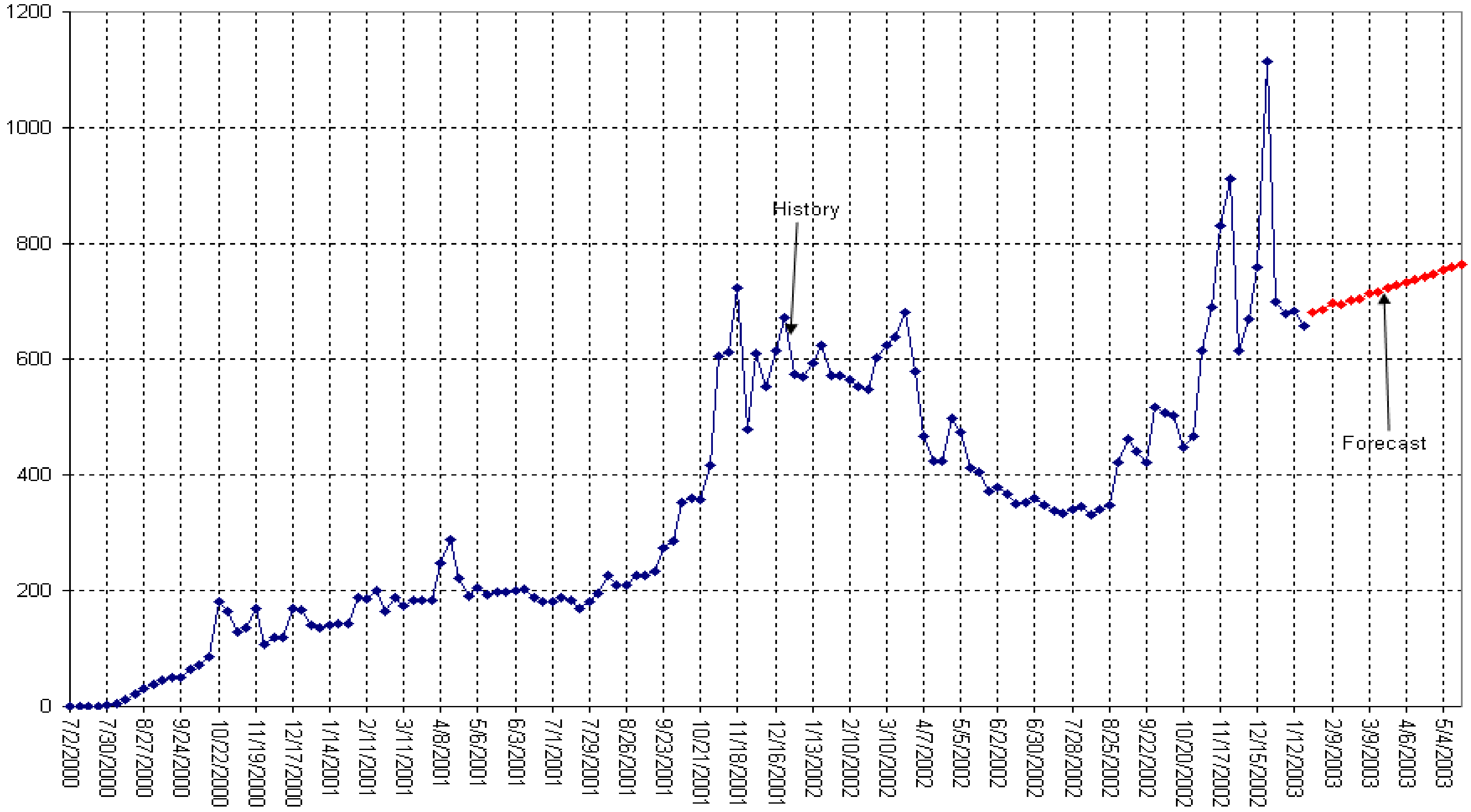
The Model Building Process



PhredTek, Inc.
logistics consulting services

General Mills Biscuits

Forecasting



Phredlek, Inc.
logistics consulting services

General Mills Biscuits Summary Statistics on Final Model

	univariate model
Residual Mean	0.000
Sum of Squares of error	598495
Variance	4500
Standard Deviation	67.1
Mean Absolute Deviation	51.4
R Square	0.904

This model leaves me with an unsettled feeling. There was lots of variation in history that was caused by something. Possible continued variation in those causes is not reflected in the forecast. Can we identify those causes and model them?

PhredTek, Inc.
logistics consulting services

General Mills Biscuits Factors Which Affect Sales

- Price – measured as average unit price
- Holidays – event indicator variables
  - Thanksgiving
  - Easter
  - Christmas
- Quality of merchandising – measured as percent of stores containing merchandising materials
- Temporary price reduction – measured as the percent by which the front line price is reduced.
- TV ads – measured as the number of adds appearing on national television in the week.

PhredTek, Inc.
logistics consulting services

General Mills Biscuits Sample Data

week	sales	avg unit price	Easter	Thanksg	Christm	merch quality	distribu	temp price red	tv adds
11/18/2001	724.11	2.71	0	1	0	29	269	22	6
11/25/2001	479.34	2.78	0	0	0	16	259	24	8
12/2/2001	610.38	2.85	0	0	0	17	272	21	10
12/9/2001	552.64	2.91	0	0	0	12	269	13	12
12/16/2001	613.82	2.92	0	0	0	8	267	15	9
12/23/2001	670.74	2.93	0	0	1	6	264	13	7
12/30/2001	573.76	2.94	0	0	0	6	260	11	6
1/6/2002	568.42	2.94	0	0	0	10	265	12	5
1/13/2002	593.83	2.96	0	0	0	12	264	8	6
1/20/2002	623.46	2.82	0	0	0	7	263	19	8
1/27/2002	571.04	2.84	0	0	0	7	266	22	9
2/3/2002	570.97	2.87	0	0	0	12	265	22	7
2/10/2002	564.44	2.86	0	0	0	20	264	15	6
2/17/2002	552.6	2.88	0	0	0	16	263	15	5
2/24/2002	547.56	2.92	0	0	0	14	264	13	7
3/3/2002	602.94	2.87	0	0	0	17	263	11	9
3/10/2002	623.52	2.89	0	0	0	12	267	11	7
3/17/2002	639.17	2.78	0	0	0	20	268	19	6
3/24/2002	681.5	2.77	0	0	0	18	268	21	5
3/31/2002	578.09	2.8	1	0	0	10	260	16	4

General Mills Biscuits Starting Regression Model

	Coef	*Std err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	728.66	255.25	2.855	0.005	223.50	1233.83
avg unit price	-259.18	89.92	-2.882	0.005	-437.14	-81.22
Easter	79.61	41.89	1.901	0.060	-3.29	162.52
Thanksgiving	113.84	35.00	3.253	0.001	44.57	183.11
Christmas	189.44	33.95	5.580	0.000	122.24	256.63
merch quality	3.01	1.48	2.037	0.044	0.09	5.93
distribution	1.66	0.12	14.301	0.000	1.43	1.89
temp price red	2.17	1.63	1.331	0.186	-1.06	5.41
tv adds	4.48	0.68	6.634	0.000	3.14	5.82

PhredTek, Inc.
logistics consulting services

General Mills Biscuits Final Regression Model

	Coef	Std Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	947.34	227.46	4.165	0.000	497.23	1397.45
avg unit price	-336.54	79.94	-4.210	0.000	-494.73	-178.36
Thanksgiving	113.61	35.22	3.226	0.002	43.91	183.31
Christmas	185.18	34.25	5.407	0.000	117.41	252.95
merch quality	3.82	1.40	2.729	0.007	1.05	6.58
distribution	1.77	0.09	19.026	0.000	1.58	1.95
tv adds	4.56	0.66	6.904	0.000	3.25	5.87

PhredTek, Inc.
logistics consulting services

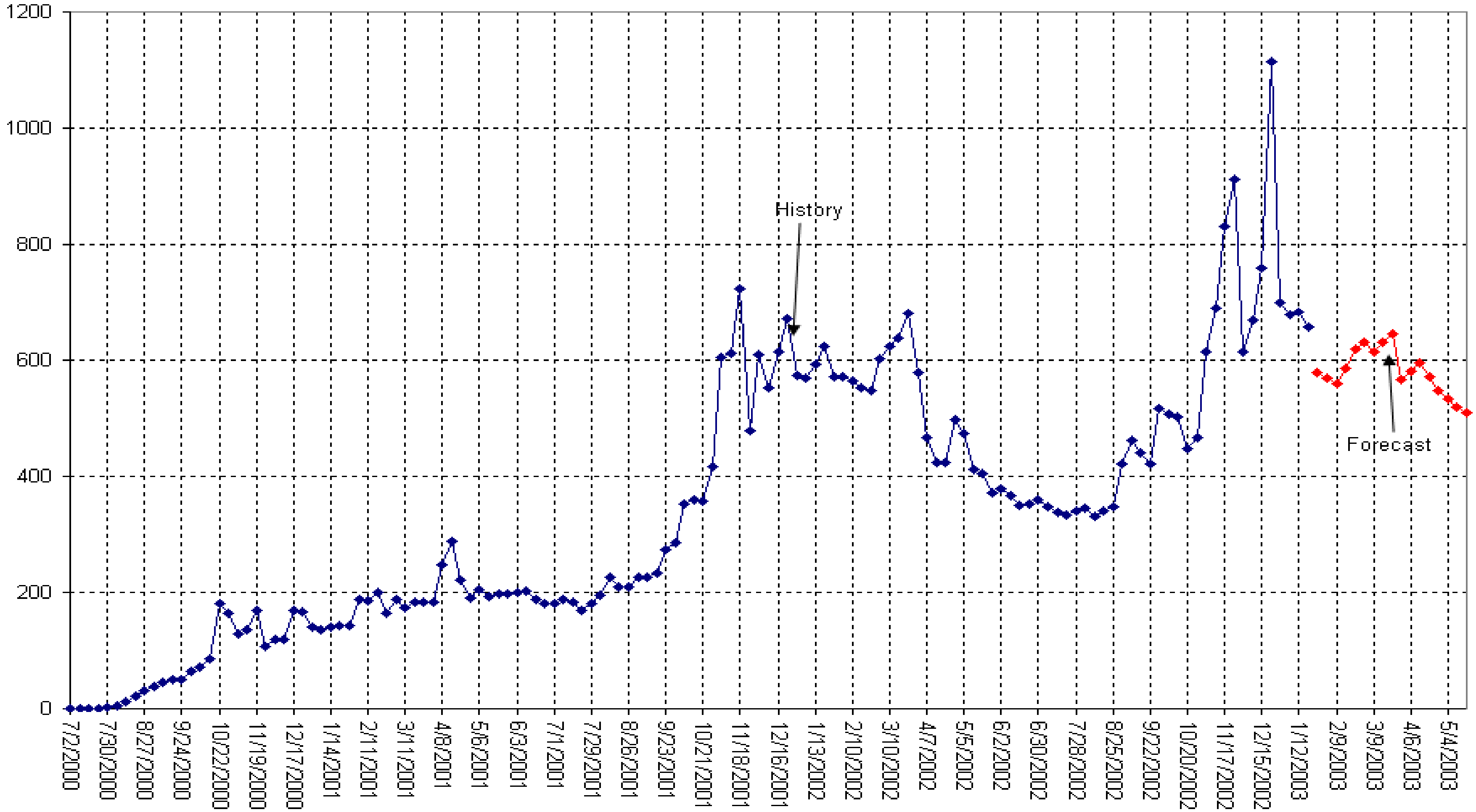
General Mills Biscuits Final Regression Model in Equation Form

$$
\begin{aligned}
Y_t &= 947.34 - 336.54\, price_t \\
&+ 113.61\, Thanksgiving_t \\
&+ 185.18\, Christmas_t \\
&+ 3.82\, merchQuality_t \\
&+ 1.77\, distribution_t \\
&+ 4.56\, TVadds_t + a_t
\end{aligned}
$$

PhredTek, Inc.
logistics consulting services

General Mills Biscuits

Regression Forecast

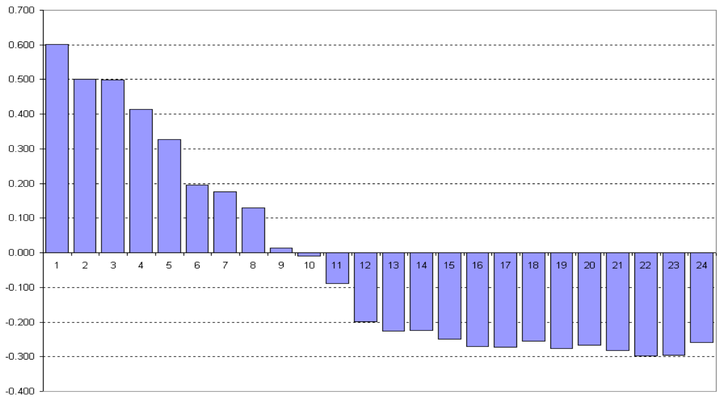


Phredlek, Inc.
logistics consulting services

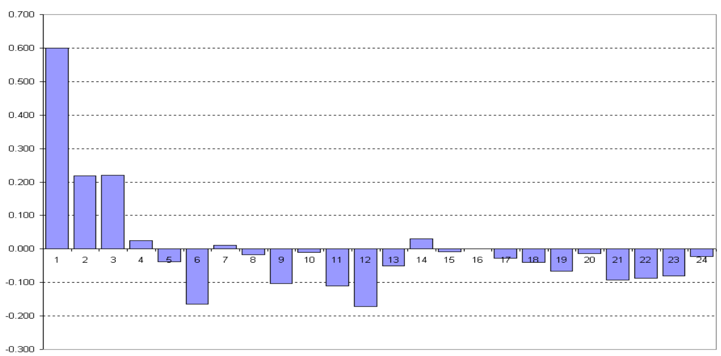
General Mills Biscuits
Summary Model Statistics

	univariate model	pure regression
Residual Mean	0.000	0.000
Sum of Squares of error	598495	424561
Variance	4500	3192
Standard Deviation	67.1	56.5
Mean Absolute Deviation	51.4	43.3
R Square	0.904	0.937

ACF and PACF for General Mills Biscuits – Regression Error



Autocorrelation



Partial Autocorrelation

lag	acf	cov	pacf
0	1	3192	
1	0.601	1917	0.601
2	0.500	1596	0.218
3	0.499	1594	0.220
4	0.414	1323	0.025
5	0.326	1041	-0.038
6	0.195	622	-0.164
7	0.177	565	0.010
8	0.129	413	-0.017
9	0.015	46.4	-0.102
10	-0.010	-31	-0.010
11	-0.089	-283	-0.109
12	-0.199	-635	-0.172
13	-0.226	-721	-0.051
14	-0.224	-716	0.030
15	-0.249	-795	-0.008
16	-0.270	-861	0.000
17	-0.272	-867	-0.028
18	-0.255	-814	-0.040
19	-0.276	-881	-0.066
20	-0.267	-852	-0.013
21	-0.282	-899	-0.093
22	-0.297	-948	-0.087
23	-0.295	-941	-0.080
24	-0.259	-826	-0.023

PhredTek, Inc.
logistics consulting services

General Mills Biscuits
Problems with the regression model

- Autocorrelation is significant
  - Violates the regression assumption that errors are not correlated
  - Places the entire result in question.
- It does not consider potential lead and lag effects on causal variables.

General Mills Biscuits Complete ARIMA Transfer Function Model

Autobox starting model

	MODEL COMPONENT	LAG	COEFF	STD	P	T
#		(BOP)		ERROR	VALUE	VALUE
1	CONSTANT		276.0	205.0	0.180	1.350
	INPUT SERIES X1 avg_unit_price					
2	Omega (input) -Factor # 1	-2	-94.2	71.8	0.192	-1.310
	INPUT SERIES X2 Easter					
3	Omega (input) -Factor # 2	-1	61.5	38.0	0.108	1.620
	INPUT SERIES X3 Thanksgiving					
4	Omega (input) -Factor # 3	0	81.7	33.8	0.017	2.420
5		1	-56.0	35.3	0.115	1.590
	INPUT SERIES X4 Christmas					
6	Omega (input) -Factor # 4	0	202.0	33.0	0.000	6.130
7		1	138.0	32.6	0.000	-4.220
8		2	74.1	31.7	0.021	-2.340

General Mills Biscuits
Complete ARIMA Transfer Function Model

Autobox starting model (continued)

#	MODEL COMPONENT	LAG (BOP)	COEFF	STD ERROR	P VALUE	T VALUE
INPUT SERIES X5 merch_quality						
9	Omega (input) -Factor # 5	-1	-2.8	1.5	0.065	-1.860
		0	-1.2	1.8	0.499	0.680
		1	9.1	1.8	0.000	-5.060
		2	-1.9	1.8	0.276	1.090
INPUT SERIES X6 distribution						
	Omega (input) -Factor # 6	0	0.7	1.3	0.603	0.520
		1	-0.8	1.9	0.658	0.440
		2	-2.8	1.6	0.090	1.710
		3	4.3	1.1	0.000	-3.820
INPUT SERIES X7 temp_price_red						
	Omega (input) -Factor # 7	0	6.7	2.0	0.001	3.290
		1	2.5	2.3	0.278	-1.090
		2	-2.8	2.4	0.241	1.180
		3	3.1	1.9	0.109	-1.620
INPUT SERIES X8 TV_adds						
	Omega (input) -Factor # 8	-2	2.0	0.8	0.012	2.570

PhredTek, In
logistics consulting serv

General Mills Biscuits
Complete ARIMA Transfer Function Model

Autobox model after necessity checks

#	MODEL COMPONENT	LAG (BOP)	COEFF	STD ERROR	P VALUE	T VALUE
1	CONSTANT		5.9	10.4	0.572	0.570
	INPUT SERIES X1 Thanksgiving					
	Differencing	1				
2	Omega (input) -Factor # 1	0	72.1	22.6	0.002	3.190
	INPUT SERIES X2 Christmas					
3	Omega (input) -Factor # 2	0	226.0	31.9	0.000	7.080
4		1	132.0	31.9	0.000	-4.130
5		2	85.4	31.7	0.008	-2.690
	INPUT SERIES X3 merch_quality					
6	Omega (input) -Factor # 3	-1	-2.4	1.2	0.048	-2.000
7		1	8.6	1.5	0.000	-5.740
	INPUT SERIES X4 distribution					
8	Omega (input) -Factor # 4	2	-3.1	1.1	0.005	-2.890
9		3	4.5	1.1	0.000	-4.240
	INPUT SERIES X5 temp_price_red					
	Omega (input) -Factor # 5	0	6.2	1.5	0.000	4.250
	INPUT SERIES X6 TV_adds					
	Omega (input) -Factor # 6	-2	2.7	0.7	0.000	4.080

PhredTek, Inc.
logistics consulting services

General Mills Biscuits Complete ARIMA Transfer Function Model

lag	acf	pacf
1	0.515	0.515
2	0.374	0.147
3	0.239	-0.001
4	0.183	0.032
5	0.141	0.022
6	0.086	-0.024
7	0.146	0.118
8	0.069	-0.066
9	0.001	-0.086
10	0.017	0.048
11	-0.009	-0.024
12	-0.116	-0.164
13	-0.199	-0.118
14	-0.227	-0.080

Sufficiency Check

Autocorrelation drops off gradually and pacf at lag 1 is large then drops off steeply. The model is not sufficient.

Autobox adds an AR1 term

General Mills Biscuits
Complete ARIMA Transfer Function Model

Autobox model with AR1 term added

#	MODEL COMPONENT	LAG (BOP)	COEFF	STD ERROR	P VALUE	T VALUE
1	CONSTANT		6.9	7.2	0.341	0.960
2	Autoregressive-Factor # 1	1	0.4	0.1	0.000	4.810
	INPUT SERIES X1 Thanksgiving					
	Differencing	1				
3	Omega (input) -Factor # 2	0	66.0	12.9	0.000	5.100
	INPUT SERIES X2 Christmas					
4	Omega (input) -Factor # 3	0	131.0	24.2	0.000	5.430
5		1	59.8	25.0	0.018	-2.390
6		2	28.9	22.6	0.202	-1.280
	INPUT SERIES X3 merch_quality					
7	Omega (input) -Factor # 4	-1	-0.8	1.0	0.404	-0.840
8		1	7.4	1.1	0.000	-6.630
	INPUT SERIES X4 distribution					
9	Omega (input) -Factor # 5	2	-0.8	0.8	0.318	-1.000
		3	2.1	0.7	0.006	-2.790
	INPUT SERIES X5 temp_price_red					
	Omega (input) -Factor # 6	0	4.7	1.2	0.000	4.060
	INPUT SERIES X6 TV_adds					
	Omega (input) -Factor # 7	-2	-2.0	0.9	0.032	-2.170
		1	-1.7	1.4	0.237	1.190
		2	8.5	1.5	0.000	-5.690

PhredTek, Inc
logistics consulting services

General Mills Biscuits
Complete ARIMA Transfer Function Model

Autobox model after second necessity checks

#	MODEL COMPONENT	LAG (BOP)	COEFF	STD ERROR	P VALUE	T VALUE
1	CONSTANT		4.6	7.1	0.518	0.650
2	Autoregressive-Factor # 1	1	0.5	0.1	0.000	5.930
	INPUT SERIES X1 Thanksgiving					
	Differencing	1				
3	Omega (input) -Factor # 2	0	68.7	12.4	0.000	5.550
	INPUT SERIES X2 Christmas					
4	Omega (input) -Factor # 3	0	125.0	22.5	0.000	5.570
5		1	46.9	22.2	0.037	-2.110
	INPUT SERIES X3 merch_quality					
6	Omega (input) -Factor # 4	1	7.3	1.1	0.000	6.870
	INPUT SERIES X4 distribution					
7	Omega (input) -Factor # 5	3	1.3	0.1	0.000	14.560
	INPUT SERIES X5 temp_price_red					
8	Omega (input) -Factor # 6	0	4.0	1.2	0.001	3.440
	INPUT SERIES X6 TV_adds					
9	Omega (input) -Factor # 7	-2	-2.2	0.9	0.020	-2.350
		2	7.3	0.9	0.000	-7.920

General Mills Biscuits Autocorrelation of error after second necessity checks

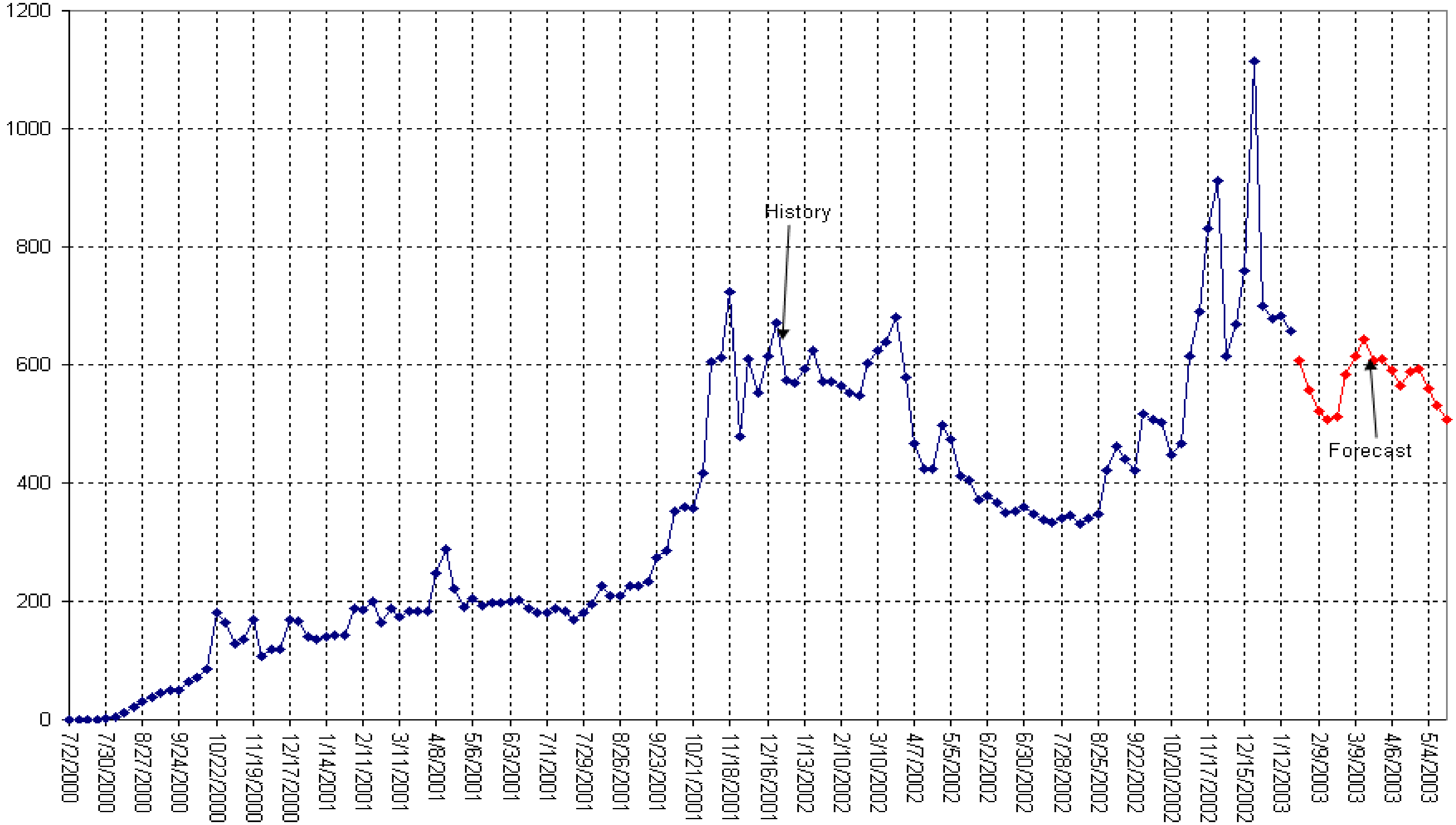
lag	acf	pacf
1	-0.011	-0.011
2	-0.061	-0.061
3	0.138	0.137
4	0.118	0.120
5	-0.043	-0.025
6	-0.091	-0.102
7	0.074	0.036
8	0.124	0.119
9	-0.068	-0.025
10	-0.022	-0.012
11	0.019	-0.041
12	-0.081	-0.105
13	-0.093	-0.067
14	-0.033	-0.021
15	-0.089	-0.097

All acfs and pacfs are small and show no pattern. The model is sufficient

So we're done and now we can forecast

PhredTek, Inc.
logistics consulting services

General Mills Biscuits Sales and Autobox Forecast



Phredlek, Inc.
logistics consulting services

General Mills Biscuits
Complete ARIMA Transfer Function Model

Final Autobox model in equation form

$$
\begin{aligned}
Y_t &= 8.64 \\
&+ (1-B)68.7 Thanksgiving_t \\
&+ (125 + 46.9B) Christmas_t \\
&+ (7.3B) merchQuality_t \\
&+ (1.3B^3) distribution_t \\
&+ (4.0)\, price\mathrm{Re}\, duction_t \\
&+ (-2.2B^{-2} + 7.3B^2) TVadds_t \\
&+ \frac{1}{(1-.471B)} a_t
\end{aligned}
$$

PhredTek, Inc.
logistics consulting services

General Mills Biscuits Summary Model Statistics

	univariate model	pure regression	causal model
Residual Mean	0.000	0.000	0.000
Sum of Squares of error	598495	424561	189651
Variance	4500	3192	1426
Standard Deviation	67.1	56.5	37.8
Mean Absolute Deviation	51.4	43.3	29.0
R Square	0.904	0.937	0.971

Outliers
Types of Outliers

- One time outliers – called pulses
- Level shifts
- Outliers that repeat every season – called seasonal pulses

Outliers are called interventions in the literature

PhredTek, Inc.
logistics consulting services

Outliers
How Outliers Are Modeled

month	pulse	level	seasonal pulse
1/1/2000	0	0	0
2/1/2000	0	0	0
3/1/2000	0	0	0
4/1/2000	0	0	0
5/1/2000	0	0	0
6/1/2000	0	0	0
7/1/2000	0	0	0
8/1/2000	0	0	0
9/1/2000	0	0	0
10/1/2000	0	0	0
11/1/2000	1	0	0
12/1/2000	0	0	0
1/1/2001	0	0	1
2/1/2001	0	1	0
3/1/2001	0	1	0
4/1/2001	0	1	0
5/1/2001	0	1	0
6/1/2001	0	1	0
7/1/2001	0	1	0
8/1/2001	0	1	0
9/1/2001	0	1	0
10/1/2001	0	1	0
11/1/2001	0	1	0
12/1/2001	0	1	0
1/1/2002	0	1	1
2/1/2002	0	1	0
3/1/2002	0	1	0
4/1/2002	0	1	0
5/1/2002	0	1	0
6/1/2002	0	1	0
7/1/2002	0	1	0
8/1/2002	0	1	0
9/1/2002	0	1	0
10/1/2002	0	1	0
11/1/2002	0	1	0

- Dummy variables like events
- Pulse is a one time event
- Level shifts go from 0 to 1 at the onset and then stay 1 forever
- Seasonal pulses are like seasonal dummies but they don't start at the first year of the series. They can be thought of as an adjustment to a seasonal dummy.
- The estimated coefficient will be the size of the outlier

General Mills Biscuits
Autobox Model With Outliers Included

#	MODEL COMPONENT	LAG (BOP)	COEFF	STD ERROR	P VALUE	T VALUE
1	CONSTANT		-4.4	4.6	0.347	-0.940
2	Autoregressive-Factor # 1	1	0.847	0.054	0.000	15.830
INPUT SERIES X1 Thanksgiving						
3	Omega (input) -Factor # 2	0	105.0	14.6	0.000	7.220
4		1	-47.8	13.6	0.001	3.510
INPUT SERIES X2 Christmas						
5	Omega (input) -Factor # 3	0	53.5	15.6	0.001	3.420
6		1	97.8	16.2	0.000	-6.050
7		2	43.3	13.6	0.002	-3.180
INPUT SERIES X3 merch_quality						
8	Omega (input) -Factor # 4	-1	-1.3	0.6	0.039	-2.090
9		1	6.0	0.7	0.000	-8.520
INPUT SERIES X4 distribution						
10	Omega (input) -Factor # 5	0	3.1	0.6	0.000	5.170
11		1	-1.4	0.5	0.014	2.510
12		2	-1.5	0.5	0.007	2.750
13		3	1.3	0.5	0.014	-2.490
INPUT SERIES X5 temp_price_red						
14	Omega (input) -Factor # 6	0	3.4	0.8	0.000	4.270
15		1	2.0	0.9	0.029	-2.220

Base Model

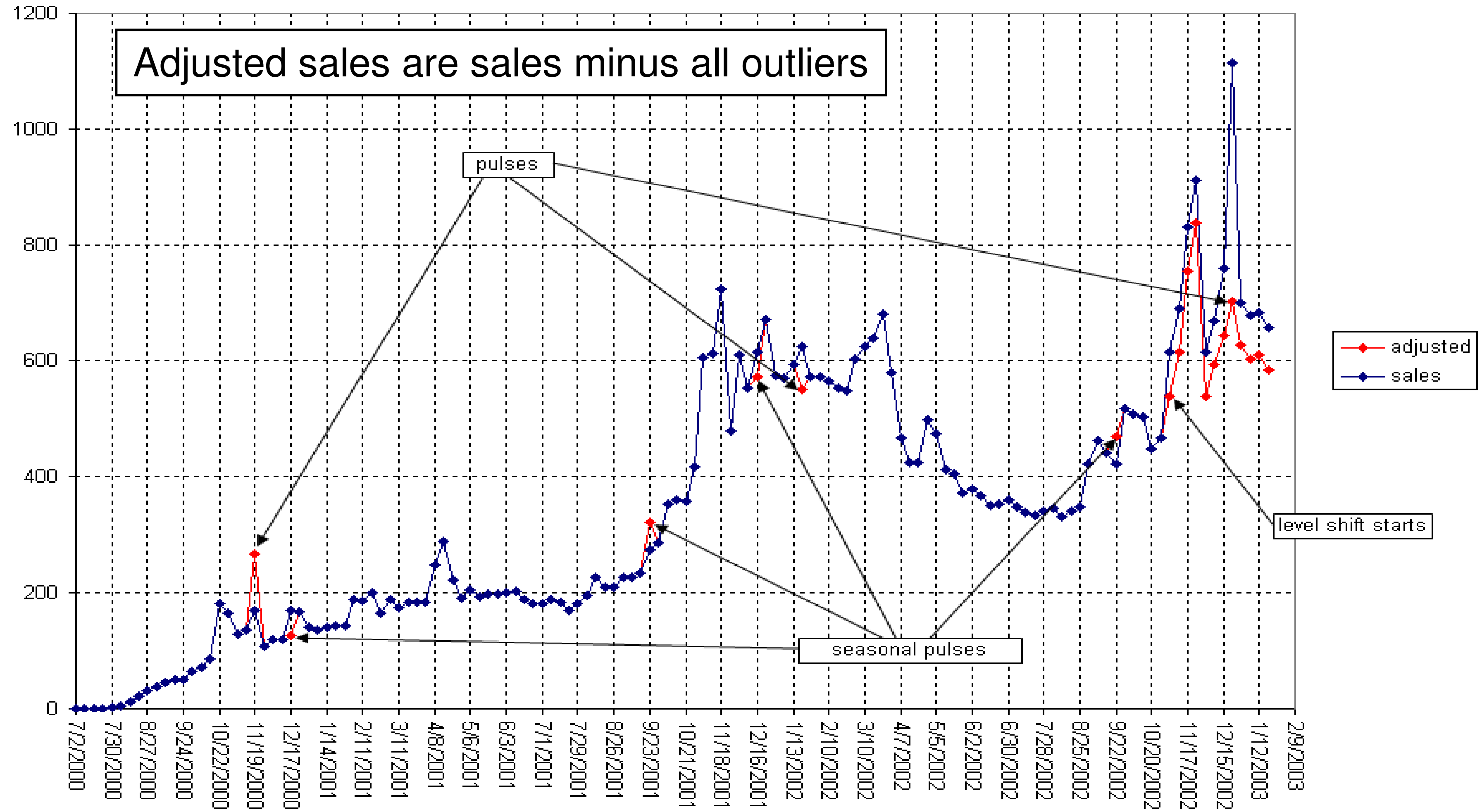
General Mills Biscuits Autobox Model With Outliers Included

#	MODEL COMPONENT	LAG (BOP)	COEFF	STD ERROR	P VALUE	T VALUE
	INPUT SERIES X6 LEVEL Shift beginning 11/3/2002					
16	Omega (input) -Factor # 7	0	75.1	24.4	0.003	3.080
	INPUT SERIES X7 PULSE at 12/22/2002					
17	Omega (input) -Factor # 8	0	335.0	25.2	0.000	13.260
	INPUT SERIES X8 PULSE at 11/19/2000					
18	Omega (input) -Factor # 9	0	-96.9	22.5	0.000	-4.310
	INPUT SERIES X9 Seasonal Pulse beginning 9/23/2001					
19	Omega (input) -Factor # 10	0	-47.7	13.0	0.000	-3.670
	INPUT SERIES X 10 PULSE at 1/20/2002					
20	Omega (input) -Factor # 11	0	72.4	20.1	0.001	3.600
	INPUT SERIES X 11 Seasonal Pulse beginning 12/17/2000					
21	Omega (input) -Factor # 12	0	42.7	12.5	0.001	3.400

Outliers (Interventions)

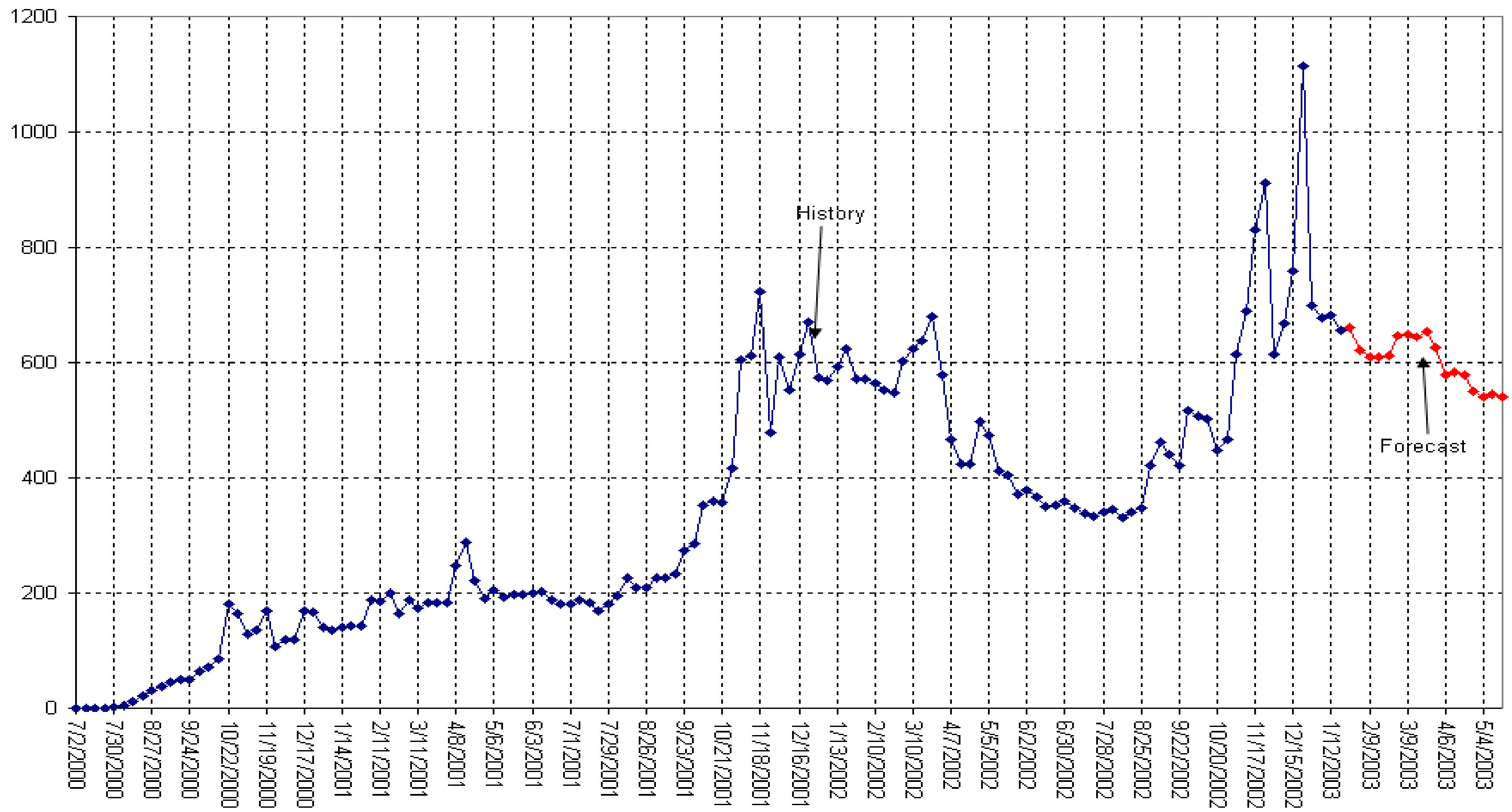
PhredTek, Inc.
logistics consulting services

General Mills Biscuits Outliers Found by Autobox



PhredTek, Inc.
logistics consulting services

General Mills Biscuits
Sales & Autobox Forecast With Outliers Fixed



PhredTek, Inc.
logistics consulting services

General Mills Biscuits Summary Model Statistics

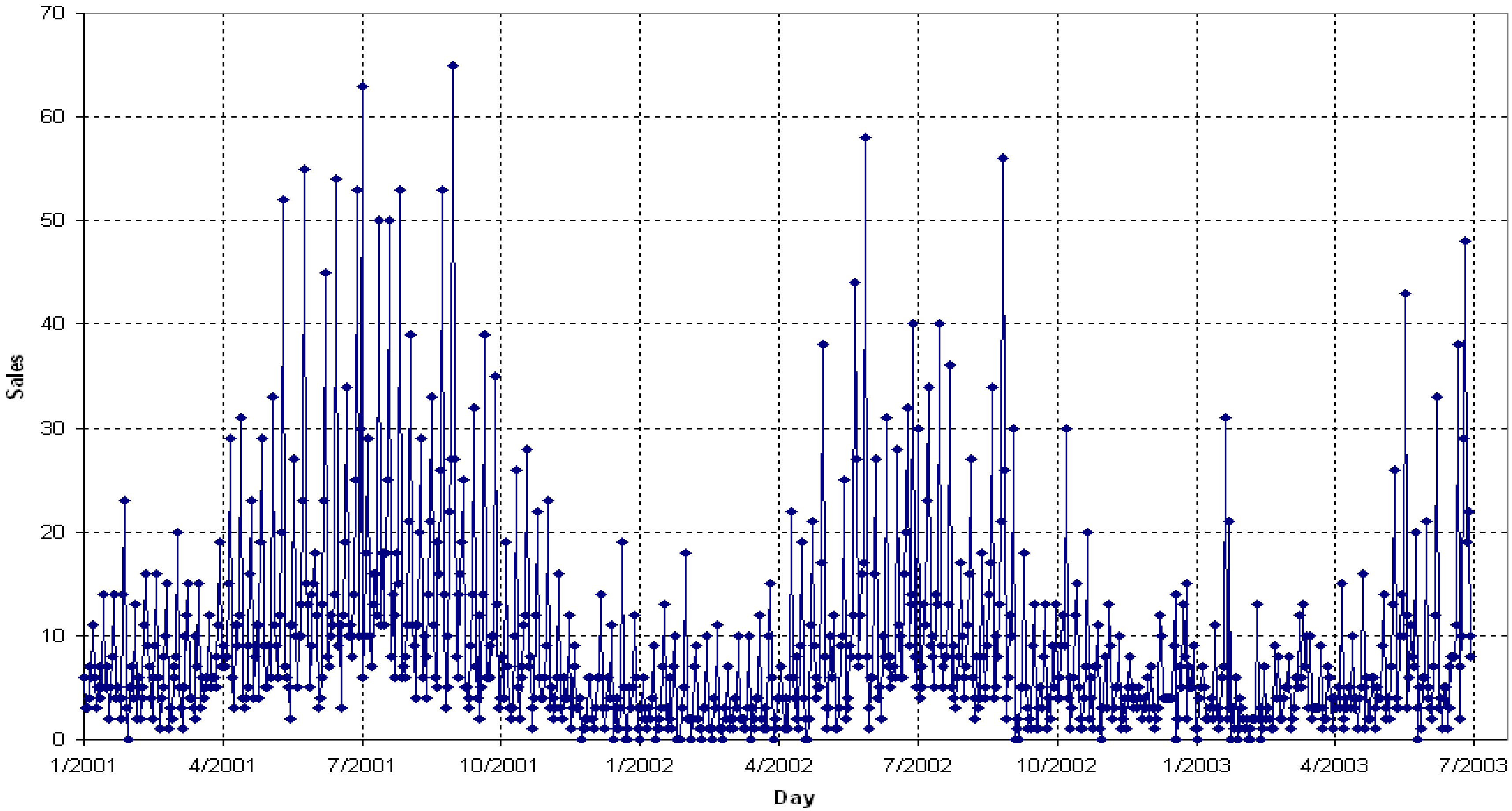
	univariate model	pure regression	causal model	with intervents
Residual Mean	0.000	0.000	0.000	0.000
Sum of Squares of error	598495	424561	189651	87163
Variance	4500	3192	1426	655
Standard Deviation	67.1	56.5	37.8	25.6
Mean Absolute Deviation	51.4	43.3	29.0	19.6
R Square	0.904	0.937	0.971	0.986

PhredTek, Inc.
logistics consulting services

Beer Sales by Brand and Package at One Grocery Store

Example

Daily Sales of 6 Pack Bottles of a Brand of Beer at a Grocery Store Out West



PhredTek, Inc.
logistics consulting services

Important Business Factors

Daily Sales of 6 Pack Bottles of a Brand of Beer at a Grocery Store Out West

- Price
- Holidays
  - New Years
  - Super Bowl
  - Memorial Day
  - July 4^{th}
  - Labor Day
  - Christmas
- High Temperature over 65 degrees F
- Seasonality
  - Day of week
  - Week of year

PhredTek, Inc.
logistics consulting services

Sample Data

Daily Sales of 6 Pack Bottles of a Brand of Beer at a Grocery Store Out West

date	sales	price	New Years	Superbowl	Memorial Day	Labor Day	July 4th	Christmas	High over 65
5/28/2001	15	5.99	0	0	1	0	0	0	15
5/29/2001	13	5.99	0	0	0	0	0	0	9
5/30/2001	5	5.99	0	0	0	0	0	0	0
5/31/2001	9	5.99	0	0	0	0	0	0	7
6/1/2001	14	5.94	0	0	0	0	0	0	14
6/2/2001	15	5.99	0	0	0	0	0	0	24
6/3/2001	18	5.99	0	0	0	0	0	0	4
6/4/2001	12	5.99	0	0	0	0	0	0	1
6/5/2001	3	5.99	0	0	0	0	0	0	8
6/6/2001	4	5.99	0	0	0	0	0	0	15
6/7/2001	6	5.99	0	0	0	0	0	0	12
6/8/2001	13	5.99	0	0	0	0	0	0	21
6/9/2001	23	5.99	0	0	0	0	0	0	23
6/10/2001	45	5.99	0	0	0	0	0	0	28
6/11/2001	8	5.99	0	0	0	0	0	0	28
6/12/2001	7	5.99	0	0	0	0	0	0	25
6/13/2001	12	5.99	0	0	0	0	0	0	0
6/14/2001	10	5.99	0	0	0	0	0	0	0
6/15/2001	11	5.99	0	0	0	0	0	0	17
6/16/2001	14	5.99	0	0	0	0	0	0	23
6/17/2001	54	5.99	0	0	0	0	0	0	27
6/18/2001	9	5.99	0	0	0	0	0	0	26
6/19/2001	11	5.99	0	0	0	0	0	0	7
6/20/2001	3	5.99	0	0	0	0	0	0	14
6/21/2001	11	5.99	0	0	0	0	0	0	13
6/22/2001	12	5.99	0	0	0	0	0	0	24
6/23/2001	19	5.99	0	0	0	0	0	0	29
6/24/2001	34	5.99	0	0	0	0	0	0	27
6/25/2001	10	5.89	0	0	0	0	0	0	28
6/26/2001	11	5.85	0	0	0	0	0	0	27
6/27/2001	8	5.86	0	0	0	0	0	0	23
6/28/2001	10	5.86	0	0	0	0	0	0	27
6/29/2001	14	5.87	0	0	0	0	0	0	30
6/30/2001	25	5.85	0	0	0	0	0	0	33
7/1/2001	53	5.86	0	0	0	0	0	0	36
7/2/2001	10	5.86	0	0	0	0	0	0	32
7/3/2001	30	5.85	0	0	0	0	0	0	29
7/4/2001	63	5.86	0	0	0	1	0	0	31

Phredlek, Inc.
logistics consulting services

Autobox Model

Daily Sales of 6 Pack Bottles of a Brand of Beer at a Grocery Store Out West

#	MODEL COMPONENT		LAG (BOP)	COEFF	STANDARD ERROR	P VALUE	T VALUE
1	CONSTANT			18.800	2.570	0.000	7.310
2	Autoregressive-Factor #	1	7	0.721	0.027	0.000	26.520
3	Autoregressive-Factor #	2	7	-0.361	0.038	0.000	-9.620
INPUT SERIES X1 Price							
4	Omega (input) -Factor #	3	0	-4.860	0.756	0.000	-6.430
INPUT SERIES X2 Super_bowl							
5	Omega (input) -Factor #	4	0	10.700	2.120	0.000	5.030
INPUT SERIES X3 Memorial_Day							
6	Omega (input) -Factor #	5	5	-5.590	2.330	0.016	-2.400
7			6	-17.300	2.840	0.000	6.100
INPUT SERIES X4 July_4th							
8	Omega (input) -Factor #	6	0	43.500	2.250	0.000	19.360
INPUT SERIES X5 Labor_Day							
9	Omega (input) -Factor #	7	0	20.100	2.640	0.000	7.620
INPUT SERIES X6 High_over_65							
10	Omega (input) -Factor #	8	0	0.135	0.021	0.000	6.330

PhredTek, Inc.
logistics consulting services

Autobox Model

Daily Sales of 6 Pack Bottles of a Brand of Beer at a Grocery Store Out West

Day Seasonality

#	MODEL COMPONENT		LAG (BOP)	COEFF	STANDARD ERROR	P VALUE	T VALUE
INPUT SERIES X7 Monday							
11	Omega (input) -Factor #	9	0	-16.500	1.320	0.000	-12.470
INPUT SERIES X8 Tuesday							
12	Omega (input) -Factor #	10	0	-17.000	1.320	0.000	-12.920
INPUT SERIES X9 Wednesday							
13	Omega (input) -Factor #	11	0	-16.900	1.320	0.000	-12.800
INPUT SERIES X 10 Thursday							
14	Omega (input) -Factor #	12	0	-16.500	1.330	0.000	-12.450
INPUT SERIES X 11 Friday							
15	Omega (input) -Factor #	13	0	-14.400	1.320	0.000	-10.860
INPUT SERIES X 12 Saturday							
16	Omega (input) -Factor #	14	0	-11.600	1.320	0.000	-8.770

PhredTek, Inc.
logistics consulting services

Autobox Model

Daily Sales of 6 Pack Bottles of a Brand of Beer at a Grocery Store Out West

	MODEL COMPONENT		LAG	COEFF	STANDARD	P	T
#			(BOP)		ERROR	VALUE	VALUE
INPUT SERIES X 13 Week 21							
17	Omega (input) -Factor #	15	0	4.450	0.893	0.000	4.990
INPUT SERIES X 14 Week 22							
18	Omega (input) -Factor #	16	0	5.840	1.060	0.000	5.500
INPUT SERIES X 15 Week 24							
19	Omega (input) -Factor #	17	0	1.980	0.917	0.032	2.150
INPUT SERIES X 16 Week 25							
20	Omega (input) -Factor #	18	0	3.060	0.908	0.001	3.380
INPUT SERIES X 17 Week 26							
21	Omega (input) -Factor #	19	0	3.480	0.894	0.000	3.890
INPUT SERIES X 18 Week 28							
22	Omega (input) -Factor #	20	0	4.020	1.040	0.000	3.870
INPUT SERIES X 19 Week 29							
23	Omega (input) -Factor #	21	0	5.050	1.100	0.000	4.610
INPUT SERIES X 20 Week 30							
24	Omega (input) -Factor #	22	0	3.350	1.100	0.003	3.030
INPUT SERIES X 21 Week 32							
25	Omega (input) -Factor #	23	0	2.670	1.040	0.011	2.560
INPUT SERIES X 22 Week 34							
26	Omega (input) -Factor #	24	0	4.530	1.060	0.000	4.260
INPUT SERIES X 23 Week 35							
27	Omega (input) -Factor #	25	0	3.370	1.120	0.003	3.010
INPUT SERIES X 24 Week 41							
28	Omega (input) -Factor #	26	0	2.130	0.983	0.031	2.170

Week Seasonality

Autobox Model

Daily Sales of 6 Pack Bottles of a Brand of Beer at a Grocery Store Out West

#	MODEL COMPONENT		LAG (BOP)	COEFF	STANDARD ERROR	P VALUE	T VALUE
INPUT SERIES X 25 One time outlier on 6/22/2003							
29	Omega (input) -Factor #	27	0	-23.400	3.700	0.000	-6.320
INPUT SERIES X 26 One time outlier on 5/27/2001							
30	Omega (input) -Factor #	28	0	19.400	3.900	0.000	4.970
INPUT SERIES X 27 One time outlier on 6/2/2002							
31	Omega (input) -Factor #	29	0	38.400	4.540	0.000	8.450
INPUT SERIES X 28One time outlier on 9/2/2001							
32	Omega (input) -Factor #	30	0	26.600	3.870	0.000	6.870
INPUT SERIES X 29 One time outlier on 7/3/2001							
33	Omega (input) -Factor #	31	0	20.300	3.720	0.000	5.460
INPUT SERIES X 30 One time outlier on 9/1/2002							
34	Omega (input) -Factor #	32	0	28.100	3.840	0.000	7.320
INPUT SERIES X 31 5/31/2001							
35	Omega (input) -Factor #	33	0	21.500	3.780	0.000	5.670
INPUT SERIES X 32 One time outlier on 7/3/2003							
36	Omega (input) -Factor #	34	0	22.500	4.080	0.000	5.520
INPUT SERIES X 33 One time outlier on 1/29/2003							
37	Omega (input) -Factor #	35	0	18.600	3.670	0.000	5.080

Outlier Correction

PhredTek, Inc.
logistics consulting services