

# *Introduction to Forecasting*

*Thomas Whalen & Jeffrey Churchill, March 31 2010*

Forecasting is perhaps the most fundamental model based activity in an organization. At their core, budgets are forecasts. Sales quotas are forecasts, even if they often are intended to be self-fulfilling prophecies. Financial *pro forma's* are forecasts. Market research studies often produce what amount to forecasts.

Our interest here is in forecasting at a more fundamental level. Often, the types of forecasts we consider are the underlying basis of the forecasts mentioned above. Broadly speaking, we consider **Time Series** models, **Leading Indicator models**<sup>1</sup>, and **Qualitative** models. We will put most of our time and effort into time series models. We will take a brief look at leading indicator modeling. Qualitative modeling is a hodgepodge of approaches from a wide variety of sources; the text will take us as far as we need to go in that area. There are also many valuable forecasting methodologies that are simply beyond the level of this course.

## *Delphi and Other Qualitative Methods*

Qualitative modeling is a hodgepodge of approaches from a wide variety of sources. Some, such as opinion polls or "grass roots forecasting," average the opinions of a large number of people in the hopes that their various biases and misinformation will statistically "wash out" into an objective forecast. At the other extreme, you can rely on one expert (maybe yourself!) who you believe to know more about the future prospects of what you're interested in than anybody else.

But generally the best approach is to try to get the consensus of a group of experts. "Panel Consensus" is the simplest method; the experts get together either in person or on line and discuss the future prospects of the variable to be forecasted. However, group dynamics problems can lead to the group becoming dominated by high status or just loudmouthed members so that you end up with just one person's opinion even though you've paid for a whole group of experts. One very successful methods to make sure you get the benefit of all the participant's thoughts is the Delphi Technique.

In the Delphi technique, each expert contributes an estimate anonymously to the group as a whole; then they all look at the whole collection of estimates and each member submits a revised estimate in the light of this new information. The process continues until consensus (or deadlock!) is reached.

## *Leading Indicator ("Causal") Models*

See the discussion about using housing starts data to forecast sales of landscape blocks on the course website for an example of leading indicator forecasting. Note that housing starts are a useful predictor of landscape blocks because housing starts are a leading indicator for landscape blocks; housing starts are useless for forecasting sales of foundation blocks because they are a simultaneous indicator.

---

<sup>1</sup> Sometimes misleadingly called "**causal** models," though usefulness as a leading indicator does not imply any simple cause-and-effect relation.

# Time Series Models

When we build a time series model, we are letting the passage of time act as a surrogate for whatever is really causing the behavior of the data. In effect, if we are forecasting sales, we are saying that the passage of time causes sales to happen. Now, not being fools, we know that isn't so, but it is a convenient fiction. The real cause system involves a very great many causal factors, not all of which are known to us, and many of which we have no way to measure. And even if we had all that information, we don't have a reliable, affordable methodology to use the information. So instead, we rely on finding a pattern in the sales data that we can relate to the passage of time.

Thus time series models are based on what we call the **Stationary Time Series Assumption**. The Stationary Time Series Assumption says that whatever the true underlying cause system behind the data may be, it will continue to operate in the future as it has in the past. Anybody can cite cases where this simply is not true - the cause system changes. We call this *catastrophic error*. By this we do not necessarily mean that something bad happened, but just that there has been a sharp break in the pattern due to some change in the nature of the system. For military industries, the end of the Cold War induced catastrophic error, but it's pretty hard to describe it as a bad thing.

Our conceptual model of a time series says that  $Y_t = T_t \times C_t \times S_t \pm E_t$ . In longer form, it is

$$Y_t = \text{Trend}_t \times \text{Cyclical}_t \times \text{Seasonal}_t \pm \text{Error}_t.$$

where

$Y_t$  is the original data for time  $t$ , for example unit sales of widgets for February 2007.

**Trend** establishes the time pattern of the basic level of the data. We presume this to really stand for the effects of many things, such as population, incomes, the effects of the product life cycle, and so forth.

**Cyclical** represents any long term swings above and below the trend level. This presumably relates to the business cycle. Political cycles and various cycles of nature may enter in here too.

**Seasonal** is any cycle that has a period of exactly one year for a complete cycle. This component can arise from weather, customs, holidays, and any other influences that repeat in approximately the same way year after year.

**Error** Forecasts are always wrong, but often still useful. We have already looked at catastrophic error, which invalidates the model when it happens. Error can also come from selecting a wrong model to apply. This amounts to self induced catastrophic error. The original data usually contains some errors of measurement. If we estimate coefficients, there will necessarily be estimation errors. And finally, there is the underlying degree of randomness of behavior of the phenomenon that the data measures. We use (Actual - Forecast) as the value of Error for any given period, so:

a positive error means the forecast should have been higher (it was too low) and

a negative error means the forecast should have been lower (it was too high).

## *Measuring the Error in Historical Forecasting*

In deciding on a forecast model to use, certain questions always come up. What does it cost? Will forecasts be available in time to use them? How accurate is it? To answer the last question, we will need standard ways of measuring the error term. A few are special to some particular models, but several are quite general. Let's look at the general ones. For a new forecasting model, we construct these measures using the historical data we based our model on, and on a "backcast" using the model. In other words, we use the model to "predict" the history of the system under the assumption that how well or poorly it "predicts" the past is evidence for how well or [poorly it will predict the future.

**Bias** The Bias is simply the arithmetic mean of the errors. We would, of course, like it to be zero. A positive bias means that the model, on the average, forecasts low. Negative means the reverse.

$$\text{Bias} = \frac{\Sigma(\text{Actual-Forecast})}{n}$$

**Mean Absolute Deviation.** MAD penalizes all errors equally, in direct proportion to their magnitude.

You can interpret the MAD as meaning the amount by which the forecast model missed on the average. (The ABS function makes negative numbers positive and leaves positive numbers alone)

$$\text{MAD} = \frac{\Sigma\text{ABS}(\text{Actual-Forecast})}{n}$$

**Mean Absolute Proportional Error (or Mean Absolute Percent Error).** This is like MAD, but penalizes error on the basis of what proportion of the actual value it is rather than its raw numeric amount.

$$\text{MAPE} = \frac{\Sigma \frac{\text{ABS}(\text{Actual-Forecast})}{\text{Actual}}}{n}$$

**Mean Square Error** MSE, as we usually call it, is very much like the simple variance you learned in statistics. It penalizes larger errors much more heavily than smaller ones.

$$\text{MSE} = \frac{\Sigma(\text{Actual-Forecast})^2}{n}$$

## *Rocky Gold's problem*

We will look at all of our *time series* techniques through this one sample problem, mostly using Excel. You should too. When you try to duplicate my results, remember that except where I tell you differently, only the original data is numbers. Everything else (other than labels) is formulas using cell references as needed.

Rocky Gold started a jewelry store some years ago, and it has grown nicely. To support the kind of volume he now sees, Rocky needs to carry inventories of jewelry larger than his personal assets can support. The manufacturers are happy to "floor plan" their products, but their interest rates combined with lost discounts make this very costly. He needs a cheaper source of cash.

The Last National Bank is willing to consider giving Rocky's Jewelers a relatively low cost line of credit that would solve Rocky's problem. Before they can process the loan, they need Rocky to provide them with financial statements and a sales forecast for the next year. Rocky's accountant is working up the financial statements and Rocky has come to us for help with the forecast. He has given us the store's dollar sales (in thousands) by quarter for the last 5 years. That is the data we'll work with.

## *Naive models*

The simplest of all forecast models is called the naive model. The naive model simply says that whatever happened last time will happen again this time. It is understandable, takes no calculations, and obviously is cheap. Sometimes it is a very good choice.

The naive model's other strength is that it gives us a baseline to measure other models against. Every other forecast model is more complex and error-prone than the naive model. You can't justify using a complicated model that won't outperform the naive model.

If we believe that the data may be seasonal, we sometimes use a slight variation on the naive model. Instead of saying "what happened last period will happen again", we say "what happened in this period last year will happen again".

Let's look at Rocky's data and see how these two versions of the naive model perform.

Looking at the results, it is pretty clear that the model labeled Naive1, which says "Next quarter will repeat this quarter" is very unsatisfactory. Its bias of 9.99 (thousand dollars) isn't bad. (The positive bias reflects the fact that there is an overall upward trend that the Naive model can't keep up with.) However, as we shall see, we can do much better than the MAD of 51.36, MAPR of 31%, and MSE of 4,926.67,

Naive2 says "Next quarter will be the same as the same quarter a year ago". It's not surprising that the bias is larger, since it ignores a whole year's growth in sales. The MAD is smaller by nearly \$26,000 and the MPE for Naive2 is less than half of Naive1's MAPE. with a MSE about one-fifth that of Naive1. Rocky could use this forecast. It makes a fine *benchmark* to compare other forecasts with.

**See if you can duplicate these results. Remember that only the column headed "Raw Data" is numbers. Everything else uses formulas.**

<u>Naive Models</u>			Raw	Naive1	Error	Abs	%	Sq. Err.	Naive2	Error	Abs	%	Sq. Err.
from	to	Qtr	Data			Error	Error				Error	Error	
Jan-97	Mar-97	1	107.8										
Apr-97	Jun-97	2	104	107.8	-3.8	3.8	4%	14.44					
Jul-97	Sep-97	3	111.9	104	7.9	7.9	7%	62.41					
Oct-97	Dec-97	4	189.2	111.9	77.3	77.3	41%	5975.29					
Jan-98	Mar-98	5	99.8	189.2	-89.4	89.4	90%	7992.36	107.8	-8	8	8%	64
Apr-98	Jun-98	6	132.7	99.8	32.9	32.9	25%	1082.41	104	28.7	28.7	22%	823.69
Jul-98	Sep-98	7	103.5	132.7	-29.2	29.2	28%	852.64	111.9	-8.4	8.4	8%	70.56
Oct-98	Dec-98	8	175.5	103.5	72	72	41%	5184	189.2	-13.7	13.7	8%	187.69
Jan-99	Mar-99	9	114.5	175.5	-61	61	53%	3721	99.8	14.7	14.7	13%	216.09
Apr-99	Jun-99	10	114.7	114.5	0.2	0.2	0%	0.04	132.7	-18	18	16%	324
Jul-99	Sep-99	11	114.1	114.7	-0.6	0.6	1%	0.36	103.5	10.6	10.6	9%	112.36
Oct-99	Dec-99	12	215.5	114.1	101.4	101.4	47%	10281.96	175.5	40	40	19%	1600
Jan-00	Mar-00	13	134.7	215.5	-80.8	80.8	60%	6528.64	114.5	20.2	20.2	15%	408.04
Apr-00	Jun-00	14	131.4	134.7	-3.3	3.3	3%	10.89	114.7	16.7	16.7	13%	278.89
Jul-00	Sep-00	15	126.8	131.4	-4.6	4.6	4%	21.16	114.1	12.7	12.7	10%	161.29
Oct-00	Dec-00	16	284.5	126.8	157.7	157.7	55%	24869.29	215.5	69	69	24%	4761
Jan-01	Mar-01	17	164.2	284.5	-120.3	120.3	73%	14472.09	134.7	29.5	29.5	18%	870.25
Apr-01	Jun-01	18	177.2	164.2	13	13	7%	169	131.4	45.8	45.8	26%	2097.64
Jul-01	Sep-01	19	186.8	177.2	9.6	9.6	5%	92.16	126.8	60	60	32%	3600
Oct-01	Dec-01	20	297.6	186.8	110.8	110.8	37%	12276.64	284.5	13.1	13.1	4%	171.61
Jan-02	Mar-02	21	297.6	297.6					164.2				
Apr-02	Jun-02	22											
				Bias	MAD	MAPE		MSE	Bias	MAD	MAPE		MSE
				10	51	31%		4927	20	26	15%		984

## Exponential Smoothing Models

Exponential Smoothing models are slightly more complex than the naive model, but not a lot. They have advantages and disadvantages that you should understand.

Advantages:	Disadvantages:
➤ Requires storing very little data	➤ Simple Exponential Smoothing always lags any trend in the data
➤ Very quick and simple to compute	➤ Double Exponential Smoothing ignores seasonality, is more complex
➤ Emphasizes the most up-to-date information	➤ Winter's Method is complex

## Simple Exponential Smoothing

The **Simple Exponential Smoothing** model is of the form  $F_t = \alpha \times Y_{t-1} + (1-\alpha) \times F_{t-1}$

Where

$F_t$  is the forecast for period t

$F_{t-1}$  is the most recent forecast

$Y_{t-1}$  is the most recent actual data point

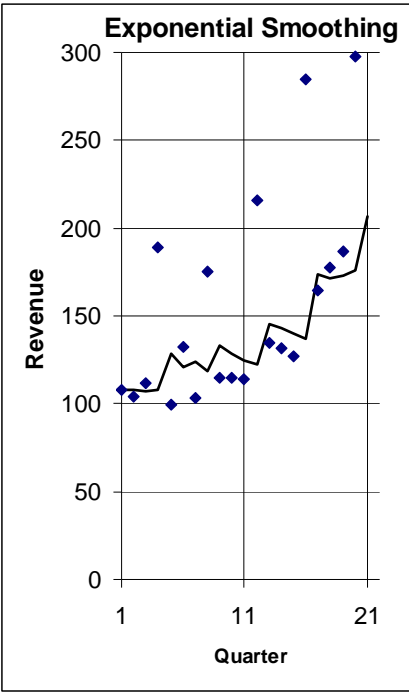
$\alpha$  is the smoothing constant  $0 < \alpha < 1$

Lets see how it performs for Rocky. Look on the next page for my results.

from	to	Qtr	Raw Data	Exponential	Error	Abs. Error	% Error	Squared Error	alpha =
Jan-97	Mar-97	1	107.8	107.8					0.25
Apr-97	Jun-97	2	104	107.8	-3.8	3.8	4%	14.44	
Jul-97	Sep-97	3	111.9	106.850	5.05	5.05	5%	25.50	
Oct-97	Dec-97	4	189.2	108.113	81.09	81.09	43%	6575.18	
Jan-98	Mar-98	5	99.8	128.384	-28.58	28.58	29%	817.07	
Apr-98	Jun-98	6	132.7	121.238	11.46	11.46	9%	131.37	
Jul-98	Sep-98	7	103.5	124.104	-20.60	20.60	20%	424.51	
Oct-98	Dec-98	8	175.5	118.953	56.55	56.55	32%	3197.59	
Jan-99	Mar-99	9	114.5	133.090	-18.59	18.59	16%	345.57	
Apr-99	Jun-99	10	114.7	128.442	-13.74	13.74	12%	188.85	
Jul-99	Sep-99	11	114.1	125.007	-10.91	10.91	10%	118.95	
Oct-99	Dec-99	12	215.5	122.280	93.22	93.22	43%	8689.97	
Jan-00	Mar-00	13	134.7	145.585	-10.88	10.88	8%	118.48	
Apr-00	Jun-00	14	131.4	142.864	-11.46	11.46	9%	131.42	
Jul-00	Sep-00	15	126.8	139.998	-13.20	13.20	10%	174.18	
Oct-00	Dec-00	16	284.5	136.698	147.80	147.80	52%	21845.33	
Jan-01	Mar-01	17	164.2	173.649	-9.45	9.45	6%	89.28	
Apr-01	Jun-01	18	177.2	171.287	5.91	5.91	3%	34.97	
Jul-01	Sep-01	19	186.8	172.765	14.04	14.04	8%	196.98	
Oct-01	Dec-01	20	297.6	176.274	121.33	121.33	41%	14720.07	
Jan-02	Mar-02	21		206.605					
Apr-02	Jun-02	22							

Bias	MAD	MAPE	MSE
21	36	19%	3044



You can see that its performance is not impressive in this case. If you try different values of  $\alpha$ , you may be able to get slightly better results, but they won't be wonderful because Rocky's sales are so seasonal and his business is growing. The simple exponential smoothing model isn't well suited to a case like this. If Rocky were a fastener wholesaler with 30,000 different kinds of nuts, bolts, studs, screws, washers, cotter pins and so forth to keep track of, simple exponential smoothing might give him 30,000 cheap forecasts very satisfactorily.

Exponential smoothing models need to be "jump started" with some initial values. In the simple exponential smoothing model, the initial "Forecast" is usually taken as equal to the actual value for the same period. If you look back at the spreadsheet, you'll see that that is exactly what I did. Now you try to first get the same results I got, then use different smoothing constants ( $\alpha$ 's) to see if you can do better.

### Double Exponential Smoothing

The **Double Exponential Smoothing** model is a little more complicated than the simple exponential smoothing model, because it attempts to stay "on top of" any trend component that may be present in the data. It doesn't cope well with strong seasonality, so we won't expect it to work too well for Rocky. See Appendix 1 for a more complete treatment.

## Simple Moving Average

The **Simple Moving Average** model is of the form  $F_t = \frac{\sum_{i=1}^k Y_{t-i}}{k}$ , where

$F_t$  is the forecast for period  $t$

$Y_{t-1}$  is the most recent actual data point

$Y_{t-2}$  is the data point preceding  $Y_{t-1}$

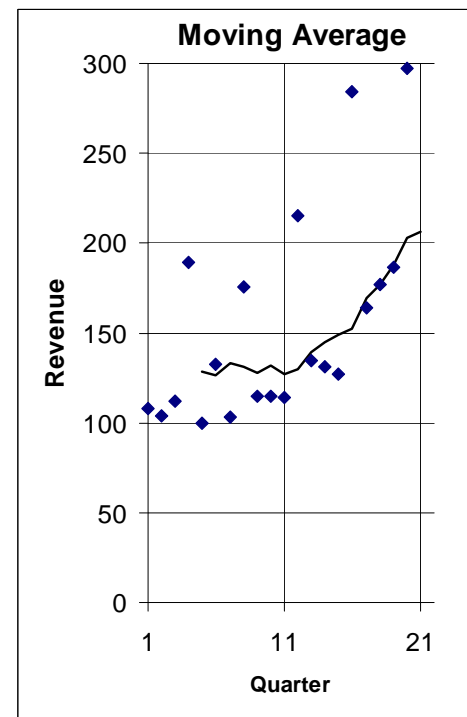
and so on.

In words, we estimate the future value  $Y_t$  using the  $k$  most recent past observations.

In Rocky's case, a four quarter moving average model looks like this:

from	to	Qtr	Raw Data	Moving Average	Abs. Error	% Error	Squared Error	
Jan-02	Mar-02	1	107.8					
Apr-02	Jun-02	2	104					
Jul-02	Sep-02	3	111.9					
Oct-02	Dec-02	4	189.2					
Jan-03	Mar-03	5	99.8	128.225	-28.43	28.43	28%	807.98
Apr-03	Jun-03	6	132.7	126.225	6.47	6.47	5%	41.93
Jul-03	Sep-03	7	103.5	133.400	-29.90	29.90	29%	894.01
Oct-03	Dec-03	8	175.5	131.300	44.20	44.20	25%	1953.64
Jan-04	Mar-04	9	114.5	127.875	-13.38	13.38	12%	178.89
Apr-04	Jun-04	10	114.7	131.550	-16.85	16.85	15%	283.92
Jul-04	Sep-04	11	114.1	127.050	-12.95	12.95	11%	167.70
Oct-04	Dec-04	12	215.5	129.700	85.80	85.80	40%	7361.64
Jan-05	Mar-05	13	134.7	139.700	-5.00	5.00	4%	25.00
Apr-05	Jun-05	14	131.4	144.750	-13.35	13.35	10%	178.22
Jul-05	Sep-05	15	126.8	148.925	-22.13	22.13	17%	489.52
Oct-05	Dec-05	16	284.5	152.100	132.40	132.40	47%	17529.76
Jan-06	Mar-06	17	164.2	169.350	-5.15	5.15	3%	26.52
Apr-06	Jun-06	18	177.2	176.725	0.47	0.47	0%	0.23
Jul-06	Sep-06	19	186.8	188.175	-1.38	1.38	1%	1.89
Oct-06	Dec-06	20	297.6	203.175	94.43	94.43	32%	8916.08
Jan-07	Mar-07	21		206.450				
Apr-07	Jun-07	22						
				Bias	MAD	MAP	MSE	
				13.45	32.02	17%	2428.56	

$k = 4$



Similarly to the results of simple exponential smoothing, the forecast line completely misses the difference between the most important quarter of each year and the other three quarters, and it lags behind the upward trend since it's far below the reality in three quarters out of every four. (Note that there is no forecast and no error measure in the first four quarters since the forecast requires four full quarters of data.)

# Trend Models

Another way to approach the problem is to try to fit a line through the data, and to project that line into the future. In some cases, an "eyeball" fit works pretty well, but there are problems with that approach. First off, it is difficult to computerize. Reasonable people can disagree over which line gives the best fit. And the error measures are all but guaranteed to increase steadily over time. We need a better way, and one is available to us. It's called the *method of least squares*.

A linear trend is of the form  $Y = a + bX$ , where  $Y$  represents the trend value (in the present case, the forecast),  $a$  is an intercept (the value of the trend at time 0),  $b$  is the slope (the amount of increase or decrease per period), and  $X$  stands for time. Normally, we will identify our oldest data point as  $X=1$  and increase  $X$  by one for each period in the data. To project into the future, we just continue increasing the value of  $X$  by 1 for each period we extend the trend. The trick lies in identifying, in some sense, the "best"  $a$  and  $b$  to use. We need a criterion.

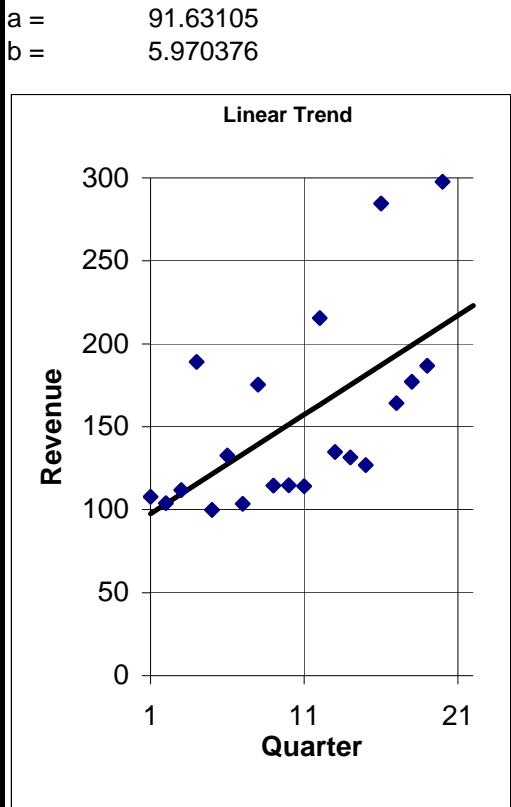
The criterion that is most generally accepted is the least squares criterion. Think of squaring the error terms (Actual - Trend) and adding up all those squared values. For a really bad fit, that sum would be huge. We want to select a value for  $a$  and a value for  $b$  that will produce a sum of squared errors smaller than we would get for any other choice of  $a$  and  $b$ . If we use actual past observed values as our values for  $Y$ , then mathematically, what we want is Minimize  $\Sigma(Y_t - a - bX_t)^2$  with respect to our choice of  $a$  and  $b$ . Excel provides us with two functions,

=INTERCEPT(Y1:Y<sub>N</sub>,X1:X<sub>N</sub>) and SLOPE(Y1:Y<sub>N</sub>,X1:X<sub>N</sub>)  
to find the optimal values of  $a$  and  $b$  respectively.

Let's try it on Rocky's data. It didn't work too badly, but the seasonality is still giving us trouble. As long as we don't somehow capture seasonality, we just aren't going to give Rocky a forecast we can be proud of. Seasonality has even messed up our trend. If you look at the graph, it is plain as the nose on your face that the trend starts too low and rises too fast. Because sales are low at the start of the year and peak during the holidays, we have underestimated the intercept and overestimated the slope. We had better do something about that.



from	to	Qtr	Raw Data	Trend Model	Abs. Error	% Error	Sq. Err.		
Jan-97	Mar-97	1	107.8	97.60	10.20	10.20	9%	104.01	
Apr-97	Jun-97	2	104	103.57	0.43	0.43	0%	0.18	
Jul-97	Sep-97	3	111.9	109.54	2.36	2.36	2%	5.56	
Oct-97	Dec-97	4	189.2	115.51	73.69	73.69	39%	5429.84	
Jan-98	Mar-98	5	99.8	121.48	-21.68	21.68	22%	470.15	
Apr-98	Jun-98	6	132.7	127.45	5.25	5.25	4%	27.53	
Jul-98	Sep-98	7	103.5	133.42	-29.92	29.92	29%	895.43	
Oct-98	Dec-98	8	175.5	139.39	36.11	36.11	21%	1303.64	
Jan-99	Mar-99	9	114.5	145.36	-30.86	30.86	27%	952.61	
Apr-99	Jun-99	10	114.7	151.33	-36.63	36.63	32%	1342.11	
Jul-99	Sep-99	11	114.1	157.31	-43.21	43.21	38%	1866.69	
Oct-99	Dec-99	12	215.5	163.28	52.22	52.22	24%	2727.39	
Jan-00	Mar-00	13	134.7	169.25	-34.55	34.55	26%	1193.42	
Apr-00	Jun-00	14	131.4	175.22	-43.82	43.82	33%	1919.87	
Jul-00	Sep-00	15	126.8	181.19	-54.39	54.39	43%	2957.91	
Oct-00	Dec-00	16	284.5	187.16	97.34	97.34	34%	9475.65	
Jan-01	Mar-01	17	164.2	193.13	-28.93	28.93	18%	836.80	
Apr-01	Jun-01	18	177.2	199.10	-21.90	21.90	12%	479.51	
Jul-01	Sep-01	19	186.8	205.07	-18.27	18.27	10%	333.73	
Oct-01	Dec-01	20	297.6	211.04	86.56	86.56	29%	7492.88	
Jan-02	Mar-02	21		217.01					
Apr-02	Jun-02	22		222.98					
								Bias	0.00
								MAD	36.42
								MAPE	23%
								MSE	1990.75



# Seasonal Decomposition

## Calculating Seasonal Indexes

The conceptual model we started with at the beginning said

$$Y_t = \text{Trend}_t \times \text{Cyclical}_t \times \text{Seasonal}_t \pm \text{Error}_t.$$

Let's return to that idea. We have become pretty knowledgeable about Trend and Error. Now, how about the other 2 components?

First, we'll dispose of the Cyclical component. While cycles are probably real (there is room for debate on the subject), they are hard to measure and harder to predict. Usually there are several cycles operating. Each of these has its own period (length of time for one full cycle) and amplitude (how far up and down it goes). Often the period and amplitude are variable. It takes a lot of data and sophisticated techniques to even measure, much less predict, cycles accurately. And for short term forecasting, even if this is done very well, you have added very little to your forecasting accuracy at great expense. We will assume that any cyclical component that is present will be captured by the trend, and won't deal with it separately. That simplifies our model to

$$Y_t = \text{Trend}_t \times \text{Seasonal}_t \pm \text{Error}_t.$$

While we know that Error is always present, by its definition we aren't trying to forecast it. Let's leave it out for now. That gives us

$$Y_t = \text{Trend}_t \times \text{Seasonal}_t$$

The Seasonal Index  $\text{Seasonal}_t$  for a given quarter such as January-March is the same for January-March of any year in the time period being modeled. It is a quantitative measure of the consistent tendency for Rocky's sales revenue in any January-March quarter to be below the average quarterly sales revenue for the year in which that quarter occurs; specifically, it is the typical ratio of the sales revenue in any January-March quarter divided by the average quarterly sales revenue for the year in which that quarter occurs.

The seasonal index is a very important part of the Time Series Decomposition Model. Not only will we use it as part of our final forecast, but it will help us to measure the Trend component much more accurately. Remember, to get the full benefit of this, you should be trying to match my results.

A very simple approach would be to compare January-March 2007 to the 2007 quarterly average, compare January-March quarter 2008 to the 2008 quarterly average, and so on, then take the average of these ratios. However, notice that this compares January-March to the quarters following it; a similar procedure for October-December quarter would compare the October-December quarter only to the quarters preceding it.

A somewhat better approach, which we will use for conceptual purposes, is to take the ratio of the average of all January-March quarters to the average of all the quarters in the database, and similarly for the other quarters. It's not perfect, but it is close enough, and we can see the big picture without getting bogged down in details.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													Moving Average based Indexes
	from	to	Qtr	Raw Data	Qtr	1997	1998	1999	2000	2001	Average	Seasonal Indexes	
3	Jan-97	Mar-97	1	107.8	1	107.8	99.8	114.5	134.7	164.2	124.20	80.5%	84.5%
4	Apr-97	Jun-97	2	104	2	104.0	132.7	114.7	131.4	177.2	132.00	85.5%	89.0%
5	Jul-97	Sep-97	3	111.9	3	111.9	103.5	114.1	126.8	186.8	128.62	83.3%	80.3%
6	Oct-97	Dec-97	4	189.2	4	189.2	175.5	215.5	284.5	297.6	232.46	150.6%	146.1%
7	Jan-98	Mar-98	5	99.8							Grand Average	154.32	80.5%
8	Apr-98	Jun-98	6	132.7									85.5%
9	Jul-98	Sep-98	7	103.5		<b>Seasonal Indexes by simplified approach</b>							83.3%
10	Oct-98	Dec-98	8	175.5		Upward trend biases first & second quarters down, third & fourth up							150.6%
11	Jan-99	Mar-99	9	114.5									80.5%
12	Apr-99	Jun-99	10	114.7									85.5%
13	Jul-99	Sep-99	11	114.1									83.3%
14	Oct-99	Dec-99	12	215.5									150.6%
15	Jan-00	Mar-00	13	134.7									80.5%
16	Apr-00	Jun-00	14	131.4									85.5%
17	Jul-00	Sep-00	15	126.8									83.3%
18	Oct-00	Dec-00	16	284.5									150.6%
19	Jan-01	Mar-01	17	164.2									80.5%
20	Apr-01	Jun-01	18	177.2									85.5%
21	Jul-01	Sep-01	19	186.8									83.3%
22	Oct-01	Dec-01	20	297.6									150.6%
23	Jan-02	Mar-02	21										80.5%
24	Apr-02	Jun-02	22										85.5%

Columns F through J contain the same raw data as Column D, but arranged into a table by year and season. Cell K3 is the average of all the Quarter 1 (January-March) sales in the historical data. Cell K4 is the average of all the Quarter 2 (April-June) sales in the historical data, and so on for K5 and K6. Cell K7 is the average of all the quarters regardless of what season of the year.

The simplified seasonal index for Quarter 1 (January-March) is found in Cell L3 by dividing K3 by K7, the simplified seasonal index for Quarter 2 (April-June) sales is found in cell L4 by dividing K4 by K7, and similarly for the other two quarters. The rest of column M repeats the Quarter 1 seasonal index for all Quarter 1 rows, the Quarter 2 seasonal index for all Quarter 2 rows, and so on.

See Appendix 2 for a discussion of the more sophisticated method covered in textbooks.

## *Calculating the Deseasonalized Data*

Well, it looks like Christmas accounts for a lot of Rocky's business, with a little help from June brides. Now let's build a better Trend. How can we do that? Remember that we said  $Y_t = \text{Trend}_t \times \text{Seasonal}_t$ . We'll use a little light duty algebra again, and this time we'll get  $\text{Trend}_t = Y_t / \text{Seasonal}_t$

$Y_t / \text{Seasonal}_t$  is just our first estimate of trend; the spreadsheet refers to it as **Deseasonalized Data**. Each number in this column is the quotient of the corresponding number in the column headed "Raw Data" divided by the number in the column headed "Seasonal Indexes."

## *Calculating the Deseasonalized Trend Model*

Deseasonalized data is not a trend that we can extend into the future for forecasting purposes; since we don't have raw data for the future, we can't divide something we don't have by a seasonal index or anything else!. To have a deseasonalized trend useful for forecasting, we need to feed the deseasonalized data into a regression model to get a linear<sup>2</sup> trend model useful for forecasting. Excel's =SLOPE and INTERCEPT functions tell us that the best linear model fit to the Rocky's deseasonalized data is  $\text{Trend}_t = 102.673 + 4.85X_t$ . This is seen in the spreadsheet column headed "Linear Model" based on the coefficients a and b that appear above the graph.

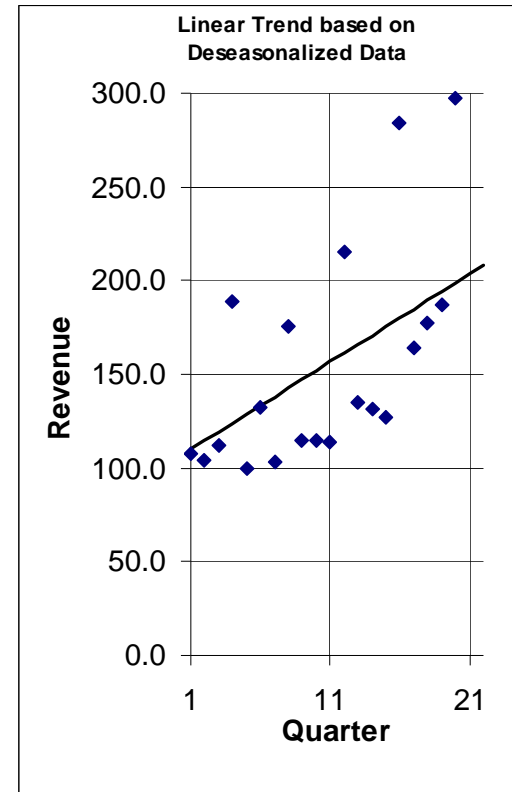
Note that the fit between the raw data and the linear model of the deseasonalized data is not as good as the fit between the raw data and the linear model of the raw data itself. But, when we re-introduce the seasonality in the next step, that will more than make up for it!

---

<sup>2</sup> Actually, we could apply any forecasting technique to the deseasonalized data to get a forecast, from the naive1 model to the fanciest mathematical hocus-pocus. But we'll stick with linear trend modeling for the moment!

from	to	Qtr	Raw Data	Seasonal Indexes	Deseas. Data	Linear Model	Abs. Error	% Error	Sq. Err.
Jan-97	Mar-97	1	107.8	80.48%	133.943	109.913	-2.11	2.11	4.46
Apr-97	Jun-97	2	104.0	85.54%	121.585	114.587	-10.59	10.59	112.09
Jul-97	Sep-97	3	111.9	83.35%	134.259	119.262	-7.36	7.36	54.19
Oct-97	Dec-97	4	189.2	150.64%	125.602	123.936	65.26	65.26	4259.37
Jan-98	Mar-98	5	99.8	80.48%	124.003	128.611	-28.81	28.81	830.05
Apr-98	Jun-98	6	132.7	85.54%	155.138	133.285	-0.59	0.59	0.34
Jul-98	Sep-98	7	103.5	83.35%	124.181	137.959	-34.46	34.46	1187.45
Oct-98	Dec-98	8	175.5	150.64%	116.507	142.634	32.87	32.87	1080.18
Jan-99	Mar-99	9	114.5	80.48%	142.268	147.308	-32.81	32.81	1076.39
Apr-99	Jun-99	10	114.7	85.54%	134.095	151.983	-37.28	37.28	1390.01
Jul-99	Sep-99	11	114.1	83.35%	136.899	156.657	-42.56	42.56	1811.12
Oct-99	Dec-99	12	215.5	150.64%	143.061	161.332	54.17	54.17	2934.21
Jan-00	Mar-00	13	134.7	80.48%	167.366	166.006	-31.31	31.31	980.07
Apr-00	Jun-00	14	131.4	85.54%	153.619	170.681	-39.28	39.28	1542.96
Jul-00	Sep-00	15	126.8	83.35%	152.136	175.355	-48.55	48.55	2357.59
Oct-00	Dec-00	16	284.5	150.64%	188.867	180.029	104.47	104.47	10914.10
Jan-01	Mar-01	17	164.2	80.48%	204.020	184.704	-20.50	20.50	420.41
Apr-01	Jun-01	18	177.2	85.54%	207.163	189.378	-12.18	12.18	148.31
Jul-01	Sep-01	19	186.8	83.35%	224.125	194.053	-7.25	7.25	52.60
Oct-01	Dec-01	20	297.6	150.64%	197.564	199.727	98.87	98.87	9775.83
Jan-02	Mar-02	21		80.48%		203.402			
Apr-02	Jun-02	22		85.54%		208.076			

slope = 4.674  
intercept = 105.238



Bias 0.00  
MAD 36  
MAPE 22%  
MSE 2047

## *Calculating the Reseasonalized Forecast*

Our error measures look a lot better, but remember that I computed them to match the trend and the deseasonalized data. They give us hope, but let's not stop now. We have a Trend function,

$$\text{Deseasonalized Trend} = 105.238 + 4.674X_t$$

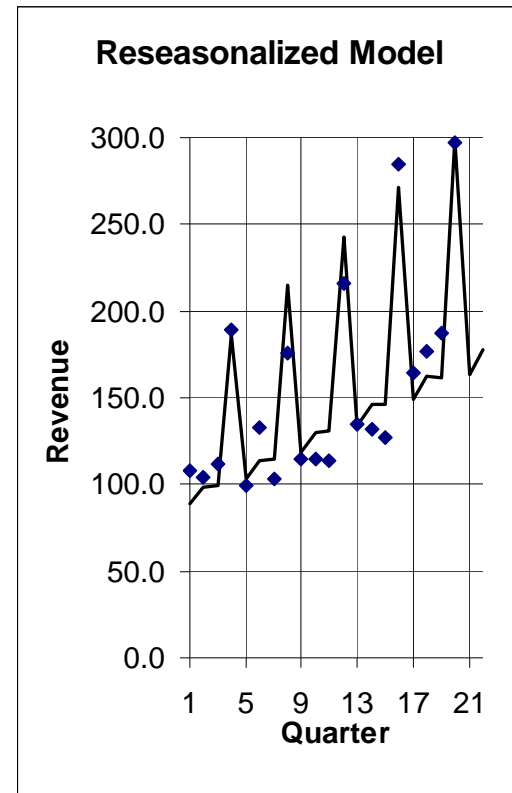
where  $X_t$  stands for the period number.

We also have the 4 Seasonal Indices: we have estimated that  
the sales revenue for the January-March quarter is 80% of the average quarter  
the sales revenue for the April-June quarter is 86% of the average quarter  
the sales revenue for the July-September quarter is 83% of the average quarter  
the sales revenue for the October-December quarter is 155.% of the average quarter  
Now to put together our forecast model all we need to do is compute the Trend value for each quarter and multiply it by that quarter's Seasonal Index. If we want to forecast, that's easy too. Since the past 5 years were periods 1 through 20, next year must be periods 21 through 24. We still, of course, reseasonalize the forecast with the same Seasonal Indices. The spreadsheet shows our result; the model of the past and forecast of the future are in the column headed "Re-Seas. Model." Each number in this column is the product of the corresponding numbers in the columns headed "Linear Model" and "Seasonal Indexes."

Now we have some error measures that really outperform the benchmark Naive2 model:. Rocky can take this to the bank. Alternatively, if he has some good business reason to suspect that this trend might not be linear, he might consider fitting a nonlinear trend to his deseasonalized data. However, this is a very advanced topic, since just "fishing" in your historical data for a good fit without a sound business reason to expect a particular functional form is risking a model that fits the old data so well that that's ALL it fits, destroying the predictive ability that we're really looking for.

from	to	Qtr	Raw Data	Seasonal Indexes	Deseas. Data	Linear Model	Re-Seas Model	Abs. Error	% Error	Sq. Err.	
Jan-97	Mar-97	1	107.8	80.48%	133.943	109.913	88.460	19.34	18%	374.03	
Apr-97	Jun-97	2	104.0	85.54%	121.585	114.587	98.014	5.99	6%	35.83	
Jul-97	Sep-97	3	111.9	83.35%	134.259	119.262	99.400	12.50	11%	156.25	
Oct-97	Dec-97	4	189.2	150.64%	125.602	123.936	186.691	2.51	1%	6.29	
Jan-98	Mar-98	5	99.8	80.48%	124.003	128.611	103.509	-3.71	4%	13.75	
Apr-98	Jun-98	6	132.7	85.54%	155.138	133.285	114.007	18.69	14%	349.41	
Jul-98	Sep-98	7	103.5	83.35%	124.181	137.959	114.984	-11.48	11%	131.88	
Oct-98	Dec-98	8	175.5	150.64%	116.507	142.634	214.857	-39.36	22%	1548.94	
Jan-99	Mar-99	9	114.5	80.48%	142.268	147.308	118.557	-4.06	4%	16.46	
Apr-99	Jun-99	10	114.7	85.54%	134.095	151.983	130.001	-15.30	13%	234.12	
Jul-99	Sep-99	11	114.1	83.35%	136.899	156.657	130.568	-16.47	14%	271.19	
Oct-99	Dec-99	12	215.5	150.64%	143.061	161.332	243.022	-27.52	13%	757.46	
Jan-00	Mar-00	13	134.7	80.48%	167.366	166.006	133.605	1.09	1%	1.20	
Apr-00	Jun-00	14	131.4	85.54%	153.619	170.681	145.994	-14.59	11%	212.99	
Jul-00	Sep-00	15	126.8	83.35%	152.136	175.355	146.152	-19.35	15%	374.50	
Oct-00	Dec-00	16	284.5	150.64%	188.867	180.029	271.187	13.31	5%	177.22	
Jan-01	Mar-01	17	164.2	80.48%	204.020	184.704	148.654	15.55	9%	241.69	
Apr-01	Jun-01	18	177.2	85.54%	207.163	189.378	161.988	15.21	9%	231.41	
Jul-01	Sep-01	19	186.8	83.35%	224.125	194.053	161.736	25.06	13%	628.21	
Oct-01	Dec-01	20	297.6	150.64%	197.564	199.727	299.353	-1.75	1%	3.07	
Jan-02	Mar-02	21		80.48%		203.402	163.702				
Apr-02	Jun-02	22		85.54%		208.076	177.981				
								Bias	MAD	MAPE	MSE
								-1.2	14	10%	288

slope = 4.674  
intercept = 105.238



.Problems

1) If you haven't done it already, replicate my time series work with Excel.

	PER'D	MONTHS	YEAR	SEAS	Million Cases
<p>2) The friendly folks from the Federal ATF Division collected data on beer sales in the US over a number of years. The data is in millions of cases sold, and appears on the next page. They would like to develop a forecast of beer sales for 2002. This would help in working out the Federal budget.</p> <p>Buddy Weiser, Agent In Charge for the Milwaukee Office, has hired you to develop the forecast. Try several methods, and use error measures to decide which one to use for your report to Mr. Weiser. Write him a short letter explaining what the results mean. Include in your letter a printout of your Excel spreadsheet that develops your forecast. Since December and January are lumped together, use half the 93-94 and half the 94-95 forecast for December-January.</p>	1	JUN-JUL	1993	4	176
	2	AUG-SEP	1993	5	172
	3	OCT-NOV	1993	6	141
	4	DEC-JAN		1	144
	5	FEB-MAR	1994	2	136
	6	APR-MAY	1994	3	161
	7	JUN-JUL	1994	4	188
	8	AUG-SEP	1994	5	184
	9	OCT-NOV	1994	6	152
	10	DEC-JAN		1	158
	11	FEB-MAR	1995	2	148
	12	APR-MAY	1995	3	174
	13	JUN-JUL	1995	4	215
	14	AUG-SEP	1995	5	204
	15	OCT-NOV	1995	6	162
	16	DEC-JAN		1	169
	17	FEB-MAR	1996	2	162
	18	APR-MAY	1996	3	183
	19	JUN-JUL	1996	4	222
	20	AUG-SEP	1996	5	214
	21	OCT-NOV	1996	6	179
	22	DEC-JAN		1	187
	23	FEB-MAR	1997	2	182
	24	APR-MAY	1997	3	211
	25	JUN-JUL	1997	4	248
	26	AUG-SEP	1997	5	247
	27	OCT-NOV	1997	6	202
	28	DEC-JAN		1	207
	29	FEB-MAR	1998	2	193
	30	APR-MAY	1998	3	223
	31	JUN-JUL	1998	4	261
	32	AUG-SEP	1998	5	259
	33	OCT-NOV	1998	6	205
	34	DEC-JAN		1	211
	35	FEB-MAR	1999	2	201
	36	APR-MAY	1999	3	235
	37	JUN-JUL	1999	4	271
	38	AUG-SEP	1999	5	275
	39	OCT-NOV	1999	6	224
	40	DEC-JAN		1	234
	41	FEB-MAR	2000	2	227
	42	APR-MAY	2000	3	256
	43	JUN-JUL	2000	4	304
	44	AUG-SEP	2000	5	289
	45	OCT-NOV	2000	6	243
	46	DEC-JAN		1	249
	47	FEB-MAR	2001	2	239
	48	APR-MAY	2001	3	270
	49	JUN-JUL	2001	4	312
	50	AUG-SEP	2001	5	314



3) Calculate Trend and Forecast values for Rocky for each of the first 4 quarters of this year (the data stops at the end of last year). First do it in Excel (because it's easier and you can print your results), and then do it on your calculator. In using your calculator, figure out how to avoid needing to stop and write down intermediate values. And if you find you must write down intermediate values, don't round them. Rounding at intermediate steps can build to huge errors.

4) An old friend of your family, George A. Power, is an electrical supply wholesaler who stocks 20,000 different products. 18,000 of those products trouble him because he continually either runs out for several weeks or wanders into a corner of the warehouse and finds what he suspects is a 10 year supply. If he could forecast demand for these items he could then set up a simple reorder point - reorder quantity inventory system that he believes would let him give his customers better service while reducing his inventory investment.

He maintains his inventory records on a computer system that constantly updates stock on hand whenever an order is received or sold. He has no weekend transactions, so he figures that he could update a forecasting and reordering system over each weekend, probably employing a GSU student who had gotten an A in MgS 3100 to run the system as a part time job. But 18,000 forecasts per weekend sounds like a lot.

He has gotten you weekly sales for a typical one of these products for the last 2 years,

You have agreed to try to identify an effective and economical forecast procedure to use on these 18,000 products. Since George has assured you that #12 bi-polar doppelganger connectors (sales shown above) are absolutely typical, you both agree that anything that works satisfactorily for them will work for the whole line. He doesn't believe that there is anything seasonal about their use, and the data seems to bear that out. You're both grateful for that. You because the methods that measure seasonality are more work, and George because they are a lot more time consuming and expensive to use.

Test the Naive and Simple Exponential models using Excel and make George a recommendation.

Week	Sold	Week	Sold	Week	Sold	Week	Sold
1	52	27	63	53	78	79	81
2	47	28	68	54	68	80	83
3	53	29	67	55	69	81	83
4	55	30	61	56	74	82	77
5	57	31	55	57	65	83	79
6	52	32	63	58	67	84	78
7	49	33	59	59	65	85	84
8	52	34	55	60	75	86	88
9	55	35	59	61	77	87	78
10	60	36	68	62	72	88	84
11	54	37	71	63	66	89	76
12	59	38	62	64	70	90	76
13	56	39	71	65	78	91	83
14	55	40	72	66	75	92	83
15	53	41	63	67	75	93	87
16	54	42	66	68	75	94	80
17	58	43	62	69	68	95	79
18	54	44	73	70	79	96	88
19	59	45	76	71	83	97	84
20	63	46	65	72	85	98	81
21	55	47	65	73	76	99	83
22	53	48	64	74	82	100	93
23	66	49	66	75	79	101	91
24	57	50	64	76	85	102	93
25	61	51	63	77	80	103	92
26	56	52	73	78	81	104	96

## THE SCENARIO

You are now Assistant to the Vice President for Marketing at Significant Concessions, a company whose specialty is operating concession stands at events that involve large public gatherings. If the opportunity looks right, your company will set up and operate concession stands at anything from a football game to a goat roping. Your boss has just learned that the beer concession at the Snellville Dome will soon be coming open. This may or may not be a good opportunity for Significant Concessions.

Your boss has asked you to do some preliminary analysis before she enters serious negotiations with the Snellville Dome management. They have provided her with a copy of a spreadsheet containing the last 10 years' data on total event attendance, by quarter. They have also indicated that they could provide, for the same period, additional data. The additional data they could provide (for each event held in those same ten years) includes *per capita* beer consumption (in 16 ounce cups), the price per cup (in 2004 dollars), the temperature at the event, and the type of event.

She has asked that you submit to her a memorandum dealing with 2 topics:

- First, she wants your forecast of Snellville Dome attendance, by quarter, for the five years following the end of the available data, and she would like some information on how accurate you expect that forecast to be.
- Also, she wants to know what information you think you could squeeze out of the additional data if you had it. She's willing to press them for it, but only if there's a prospect of a worthwhile result. Tell her what you could do, and what statistical approach you would use.

## THE DATA AND THE ANALYSIS

The final step is to implement your forecast *model* and find ways to summarize both your results and the error term. Then you should be ready to write your memo.

Think carefully about what your boss wants to know. You probably could construct some useful summary information about your forecast of the next 5 years attendance at the Snellville Dome while you are there. Oh, and by the way . . . you have remembered to format everything appropriately, haven't you?

## YOUR JOB

Once you have completed the tasks we've just outlined, you are most of the way done. Write your boss the kind of memorandum that might get you a raise. She wants the information mentioned on the first page of this document.

## THE SCENARIO

Congratulations! We are now in the near future, and you have a job. With an office. And a salary! You work for MegaMammoth Properties, a firm which manages commercial real estate for large investors. Your boss, Bill Ding, has been wrestling with a perplexing problem. Mr. Ding knows that you passed this course with flying colors, and figures that if anybody in the office can solve the problem, you can.

The company manages a large office building, Power Place, in a suburb of Buford, Georgia. The building is less than 4 years old, and MegaMammoth has managed it since it opened. Lease provisions for such buildings are often complex and arcane, and Power Place is no exception. The lease for Power Place specifies that if the occupants of the building use more electricity than normal, MegaMammoth can "backcharge" them for the excess. "Normal" is defined in the lease to mean their fair share of the expected annual usage at 95% occupancy. Thus, for example, suppose a tenant occupies 20% of the space and 50% of the building is occupied altogether. Then they are responsible for 20/50 or 40% of any electricity used during the year over and above 50/95 of the annual 95% occupancy usage. Only a lawyer or CPA could dream up such a complicated system, but that's a simplified version of how it works.

Mr. Ding has attempted to apply a "Rule of Thumb" that managers of commercial real estate often apply. He has rejected it on 2 counts. First, he believes that the results it yields are completely unrealistic. Even more important, applying that rule gives a figure that would not permit him to backcharge the tenants. This is probably because of 3 major factors:

1. The building has never even been close to 95% occupied.
2. The "common spaces" in the building seem to use about the same amount of electricity per year at almost any level of occupancy.
3. The usage of electricity appears to be highly seasonal.

The actual Data is on the next page. After examining it, you have hypothesized the influence diagram that appears on this page. The diagram suggests that if you are going to estimate the amount of electricity that would normally be used, you must take into account the month of the year's effect on the weather and the weather's resulting effect on usage. It also suggests, as expected, that you must take into account the proportion of the space occupied.

## THE ANALYSIS

The seasonality in the data will clearly interfere with your ability to measure the effect of occupancy on electricity usage. You must first develop seasonal indices and deseasonalize the data. Then perhaps a simple regression analysis will reveal the effect of occupancy on monthly electricity usage.

If you can estimate a meaningful regression equation of the form  $Y = b_0 + b_1 X$  where  $Y$  is deseasonalized monthly consumption in KWH (kilowatt hours) and  $X$  is proportion of Power Place that is occupied, then you can answer Bill Ding's question. The answer will be

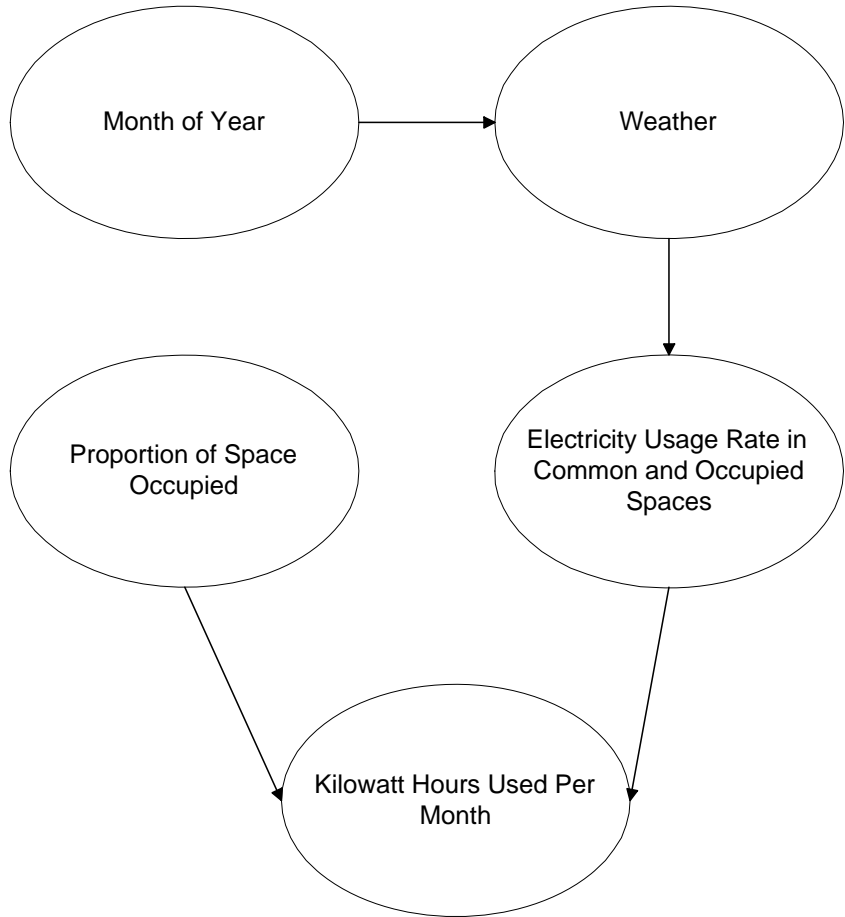
**Annual usage at 95% occupancy =  $12(b_0 + .95 b_1)$  in kilowatt hours.**

## YOUR JOB

Conduct the analysis using any combination of spreadsheet and/or statistical software you prefer. Write Bill Ding a memo presenting and explaining your results. Do not make the memo technical. He will probably need it in convincing tenants that the large bill he will present them is just and proper.

## Power Place

Month Occupancy		KWH
Mar 'X1	15.31%	256,801
Apr 'X1	15.31%	218,144
May 'X1	15.31%	172,389
Jun 'X1	15.31%	204,089
Jul 'X1	15.31%	182,770
Aug 'X1	15.31%	174,790
Sep 'X1	15.31%	191,275
Oct 'X1	15.31%	160,733
Nov 'X1	15.31%	160,393
Dec 'X1	15.31%	250,057
Jan 'X2	15.31%	272,315
Feb 'X2	15.31%	311,146
Mar 'X2	15.31%	229,274
Apr 'X2	15.31%	201,303
May 'X2	15.31%	190,875
Jun 'X2	15.31%	196,769
Jul 'X2	15.31%	233,098
Aug 'X2	15.31%	255,380
Sep 'X2	23.98%	243,214
Oct 'X2	31.98%	255,419
Nov 'X2	32.45%	221,817
Dec 'X2	32.45%	216,251
Jan 'X3	32.45%	399,001
Feb 'X3	32.45%	321,716
Mar 'X3	52.75%	415,620
Apr 'X3	52.75%	333,029
May 'X3	52.75%	335,962
Jun 'X3	52.75%	325,466
Jul 'X3	52.75%	395,868
Aug 'X3	52.75%	348,126
Sep 'X3	56.80%	356,680
Oct 'X3	57.82%	340,037
Nov 'X3	57.82%	287,422
Dec 'X3	57.82%	281,875
Jan 'X4	58.00%	459,952
Feb 'X4	58.00%	429,164
Mar 'X4	60.00%	425,202
Apr 'X4	60.00%	381,084
May 'X4	60.00%	296,641
Jun 'X4	60.00%	337,947



## Appendix 1: Double Exponential Smoothing

The **Double Exponential Smoothing** model is a little more complicated than the simple exponential smoothing model, because it attempts to stay "on top of" any trend component that may be present in the data. It doesn't cope well with strong seasonality, so we won't expect it to work too well for Rocky. There is a version of the model with one smoothing constant, and one with two. We'll look at the 2 constant version. This model says:

$$F_t = C_t + T_t$$

$$C_t = \alpha \times Y_{t-1} + (1-\alpha) \times C_{t-1}$$

$$T_t = \beta \times (C_t - C_{t-1}) + (1-\beta) \times T_{t-1}$$

$F_t$  is the forecast for period  $t$  (and  $t-1$  indicates previous period).

$C_t$  acts like the intercept in a linear equation, but continuously updated.

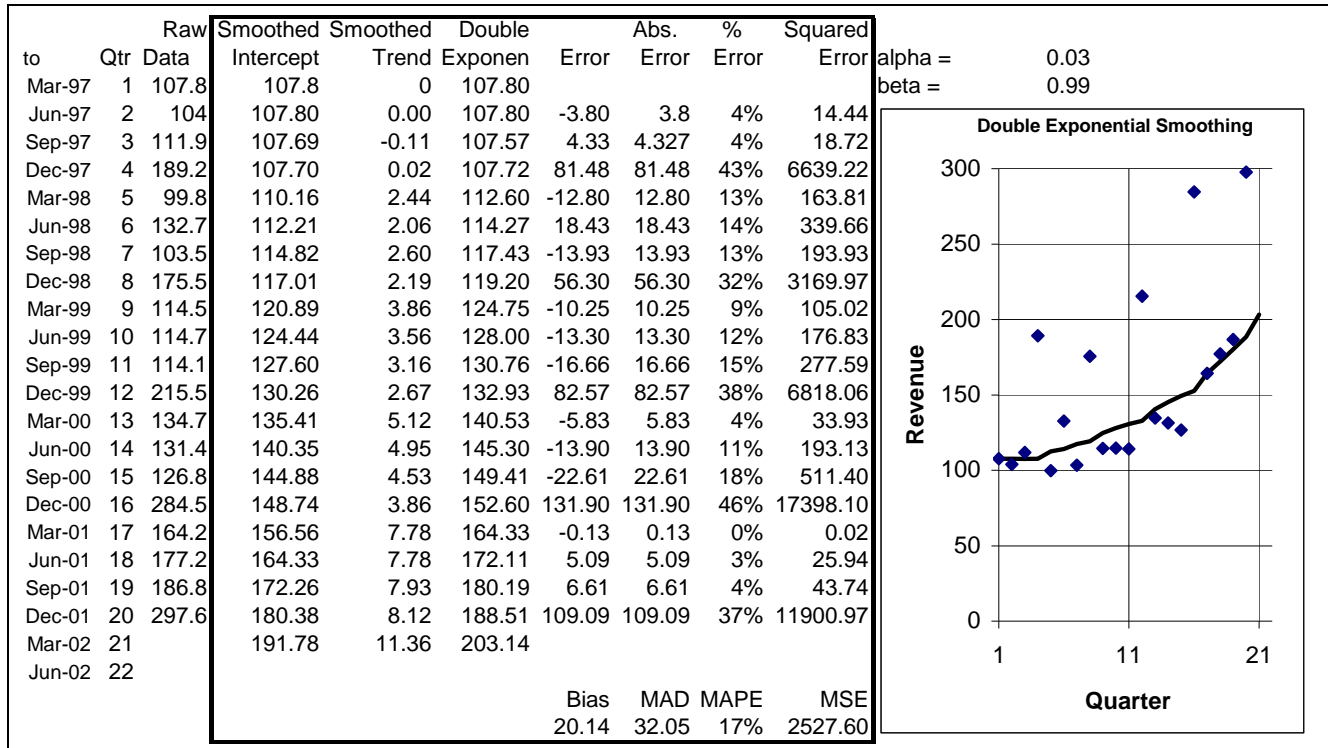
$T_t$  acts like the slope in a linear equation, a smoothed period - to - period growth.

$Y_{t-1}$  is the most recent actual observed value of the time series we are trying to forecast.

$\alpha$  is the smoothing constant for the "intercept"  $C$ .

$\beta$  is the smoothing constant for the "trend"  $T$ . Some versions of this model reuse  $\alpha$  for this purpose.

The following spreadsheet shows how double exponential smoothing (sometimes called trend adjusted exponential smoothing) performed for Rocky. No real improvement is visible, but if you try new values for  $\alpha$  and  $\beta$  you might be able to do a little better. I used the first actual value as the first  $C_t$  (intercept) and 0 as the first  $T_t$  (slope). Those are reasonable values for you to use too. Go ahead and do it.



### Winter's Method

**Winter's Method** adds a third smoothing constant and smoothed Seasonal Indices to the picture. It would work much better for Rocky than what we've seen so far, but it is complex, losing a lot of the advantages of exponential smoothing. We won't try it, but you should be aware of it.

## Appendix 2: Calculation of Seasonal Indexes

Clearly, it would make more sense to define "the year in which that January-March quarter occurs" as the year whose *middle* quarter is the quarter in question, rather than using calendar years.

Unfortunately, due to the fact that there are an even number of quarters to a year, the year whose middle quarter is January-March 2008 is the year starting August 16, 2007, and ending August 15, 2008. This consists of half of the July-September 2007 quarter, the three full quarters October-December 2007, January-March 2008, and April-June 2008. and half of the July-September 2008 quarter. Since we don't have specific data for half-quarters, we have to estimate the two half-quarter sales figures as 50% of the corresponding quarterly totals.

In particular, the total sales revenue of the year whose middle quarter is January-March 2008 is estimated as  $111.9/2 + 189.2 + 99.8 + 132.7 + 103.5/2 = 529.4$

The average quarterly sales revenue for the year whose middle quarter is January-March 2008 is estimated as  $529.4/4 = 132.35$

The ratio of sales revenue for January-March 2008 divided by the average quarterly sales revenue for the year whose middle quarter is January-March 2008 is estimated as  $99.8/132.35 = 75.4\%$

We can't calculate this for January-March 2007 or April-June 2007 because these would require data from 2006, and we can't calculate it for July-September 2001 or October-December 2001 because we don't have the required 2002 data.

In the following spreadsheet, the column headed "Annual Total" shows the total sales revenue of the year whose middle quarter is indicated for each row.

The column headed "Centered Mov. Avg." shows the average quarterly sales revenue for the year whose middle quarter is indicated for each row.

The column headed "Raw Ratio" shows the ratio of sales revenue for the quarter indicated for each row divided by the average quarterly sales revenue for the year for which that quarter is the middle quarter.

The first number in the column headed "Average Ratio" shows the average of the aforementioned ratios for all four January-March quarters for which the ration could be computed (i.e. 2008, 2009, 2000, and 2001).

The second number in the column headed "Average Ratio" shows the average of the aforementioned ratios for all four April-June quarters for which the ration could be computed (i.e. 2008, 2009, 2000, and 2001).

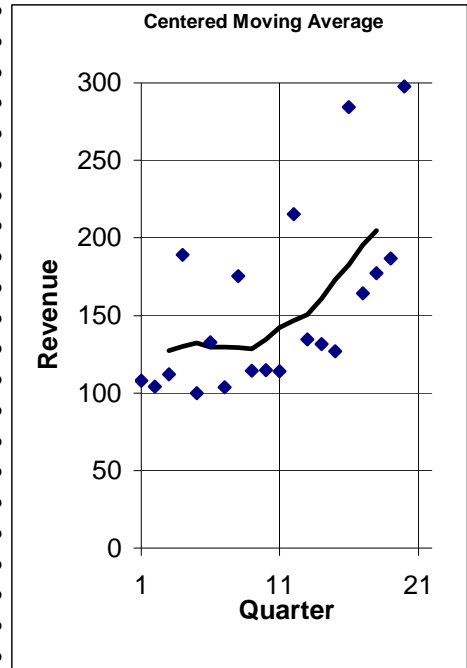
The third number in the column headed "Average Ratio" shows the average of the aforementioned ratios for all four July-September quarters for which the ration could be computed (i.e. 2007, 2008, 2009, and 2000).

The fourth number in the column headed "Average Ratio" shows the average of the aforementioned ratios for all four October-December quarters for which the ration could be computed (i.e. 2007, 2008, 2009, and 2000).

The last number in the column headed "Average Ratio" shows the average of these four averages; this will be used in column I to force the average of the final estimates of the seasonal indexes to be exactly 1.0.

The column headed "Seasonal Indexes" shows the final estimates of the seasonal indexes, found by dividing the average ratios in Column H by their own average to force the average of the final estimates of the seasonal indexes to be exactly 1.0.

from	to	Qtr	Raw Annual Data	Centered Total Mov. Avg.	Raw Average Ratio	Seasonal Indexes
Jan-97	Mar-97	1	107.8		84.50%	84.54%
Apr-97	Jun-97	2	104		88.96%	89.00%
Jul-97	Sep-97	3	111.9	508.9 127.225	88.0%	80.31% 80.35%
Oct-97	Dec-97	4	189.2	519.3 129.813	145.7%	146.04% 146.11%
Jan-98	Mar-98	5	99.8	529.4 132.350	75.4%	Average 84.54%
Apr-98	Jun-98	6	132.7	518.4 129.588	102.4%	of the 89.00%
Jul-98	Sep-98	7	103.5	518.9 129.713	79.8%	Averages 80.35%
Oct-98	Dec-98	8	175.5	517.2 129.300	135.7%	99.95% 146.11%
Jan-99	Mar-99	9	114.5	513.5 128.375	89.2%	84.54%
Apr-99	Jun-99	10	114.7	538.8 134.700	85.2%	89.00%
Jul-99	Sep-99	11	114.1	568.9 142.225	80.2%	80.35%
Oct-99	Dec-99	12	215.5	587.4 146.838	146.8%	146.11%
Jan-00	Mar-00	13	134.7	602.1 150.513	89.5%	84.54%
Apr-00	Jun-00	14	131.4	642.9 160.725	81.8%	89.00%
Jul-00	Sep-00	15	126.8	692.2 173.038	73.3%	80.35%
Oct-00	Dec-00	16	284.5	729.8 182.450	155.9%	146.11%
Jan-01	Mar-01	17	164.2	782.7 195.675	83.9%	84.54%
Apr-01	Jun-01	18	177.2	819.3 204.813	86.5%	89.00%
Jul-01	Sep-01	19	186.8			80.35%
Oct-01	Dec-01	20	297.6			146.11%
Jan-02	Mar-02	21				84.54%
Apr-02	Jun-02	22				89.00%



There is an alternate way to compute the same seasonal indexes that involves less typing and slightly simpler Excel code. Instead of calculating the total sales revenue of the year whose middle quarter is the one specified and dividing by 4, we take a simple 4-quarter moving average, then take a two-item moving average of those moving averages. A little algebra will show that this trick gives us exactly the same seasonal indexes as the method discussed previously. The key calculations are described below:

Period	Quarter	Y	Simple Average	Centered Average	
1	1	$Y_1$			
2	2	$Y_2$			
			$A_{2.5} = \text{AVERAGE}(Y_1, Y_2, Y_3, Y_4)$		
3	3	$Y_3$		$A_3 = \text{AVERAGE}(A_{2.5}, A_{3.5})$	$R_{3.1} = Y_3/A_3$
			$A_{3.5} = \text{AVERAGE}(Y_2, Y_3, Y_4, Y_5)$		
4	4	$Y_4$		$A_4 = \text{AVERAGE}(A_{3.5}, A_{4.5})$	$R_{4.1} = Y_4/A_4$
			$A_{4.5} = \text{AVERAGE}(Y_3, Y_4, Y_5, Y_6)$		
5	1	$Y_5$		$A_5 = \text{AVERAGE}(A_{4.5}, A_{5.5})$	$R_{1.2} = Y_5/A_5$
			$A_{5.5} = \text{AVERAGE}(Y_4, Y_5, Y_6, Y_7)$		
6	2	$Y_6$		$A_6 = \text{AVERAGE}(A_{5.5}, A_{6.5})$	$R_{2.2} = Y_6/A_6$
			$A_{6.5} = \text{AVERAGE}(Y_5, Y_6, Y_7, Y_8)$		
7	3	0		$A_7 = \text{AVERAGE}(A_{6.5}, A_{7.5})$	$R_{3.2} = Y_7/A_7$

Notice that the simple averages really center between quarters. That's why we had to do a moving average of moving averages to get our moving average centered on the data points. While we are noticing things, let's also notice that we couldn't get a centered moving average for the first 2 Quarters. We can't see it here, but we can't get the last 2 Quarters either. The data we'd need to do that doesn't exist. That is the penalty we must pay for using this method. If we have at least 3 years of data it is worth paying the penalty. We can't use the first and last half year in getting our seasonal ratios.

The notation  $R_{ij}$  simply means the raw seasonal ratio for season  $i$  in year  $j$ . To calculate **Seasonal Indexes** from here is the same regardless of which way we got to the centered moving average.

from	to	Qtr	Raw Data	Moving Average	Centered Mov. Avg.	Raw Ratio	Average Ratio	Seasonal Indexes
Jan-97	Mar-97	1	107.8				84.50%	84.54%
Apr-97	Jun-97	2	104.0				88.96%	89.00%
Jul-97	Sep-97	3	111.9	128.225	127.2250	88.0%	80.31%	80.35%
Oct-97	Dec-97	4	189.2	126.225	129.8125	145.7%	146.04%	146.11%
Jan-98	Mar-98	5	99.8	133.400	132.3500	75.4%	Average	84.54%
Apr-98	Jun-98	6	132.7	131.300	129.5875	102.4%	of the	89.00%
Jul-98	Sep-98	7	103.5	127.875	129.7125	79.8%	Averages	80.35%
Oct-98	Dec-98	8	175.5	131.550	129.3000	135.7%	99.95%	146.11%
Jan-99	Mar-99	9	114.5	127.050	128.3750	89.2%		84.54%
Apr-99	Jun-99	10	114.7	129.700	134.7000	85.2%		89.00%
Jul-99	Sep-99	11	114.1	139.700	142.2250	80.2%		80.35%
Oct-99	Dec-99	12	215.5	144.750	146.8375	146.8%		146.11%
Jan-00	Mar-00	13	134.7	148.925	150.5125	89.5%		84.54%
Apr-00	Jun-00	14	131.4	152.100	160.7250	81.8%		89.00%
Jul-00	Sep-00	15	126.8	169.350	173.0375	73.3%		80.35%
Oct-00	Dec-00	16	284.5	176.725	182.4500	155.9%		146.11%
Jan-01	Mar-01	17	164.2	188.175	195.6750	83.9%		84.54%
Apr-01	Jun-01	18	177.2	203.175	204.8125	86.5%		89.00%
Jul-01	Sep-01	19	186.8	206.450				80.35%
Oct-01	Dec-01	20	297.6					146.11%
Jan-02	Mar-02	21						84.54%
Apr-02	Jun-02	22						89.00%