Electric Bill Data
Instructor's Manual

This manual accompanies the article:

McLaren, C. H., and Mclaren, B. J. (2003) "Electric Bill Data," *Journal of Statistics Education* [Online], 11(1). www.amstat.org/publications/jse/v11n1/datasets.mclaren.html

Background

Synopsis
Actual data are provided for a monthly electric bill from January, 1991, through December 2000. As the assignments progress, the concepts needed for analysis of the data become increasingly complex. Six independent statistics assignments, plus a summary assignment, are provided, and instructors could use as many of the parts as appropriate for their classes. Depending on the assignments chosen, this material is suitable for any class from high school through the graduate level. An additional set of assignment questions deals with the use of spreadsheets for break-even analysis and the manipulation of tariff data.

Statistics Assignments

1. The first assignment provides the monthly payment data and asks the student to examine it both statistically and visually. The data is markedly seasonal and moderate trend exists. Students should notice that there is a problem with outliers and missing values, so data cleansing becomes an issue unless the instructor has remedied the problem in advance.

2. The second assignment deals with seasonal decomposition. Students must determine whether an additive or multiplicative model is more appropriate for this time series and how they will judge between the two. A discussion on error measurements should occur.

3. In the third assignment, students concentrate on other univariate forecasting methods. Students are asked to apply simple exponential smoothing and should realize that a seasonal method is more appropriate. If there is access to forecasting software, or if students are adept with spreadsheet operations, other smoothing methods should be applied.

4. ARIMA models are the topic of the fourth assignment. This advanced topic may not be suitable for lower level college classes. The assignment fosters discussion on appropriate model parameters and requires the use of Minitab or another statistical package.

5. The fifth model asks students to use indicator variables for seasonality and a time variable to develop a multiple regression model.

6. In the sixth assignment, the discussion shifts to causal models. By this time, students feel strongly that there are reasons for data fluctuations and are motivated to be able to incorporate explanatory factors. Additional data is provided, and/or students can be asked to

search for information on heating/cooling degree days, electric rates, and other topics that may have a bearing on the size of the bill. Discussion about data sources and data availability for a causal model is appropriate. The instructor should remind students that if the forecasting model is to be used for prediction, the values of the independent variables must be available. The correlation between monthly indicator variables and heating and cooling degree days can be examined for issues of multicollinearity.

7. The concluding statistics assignment, of particular value if the dataset has been used in an ongoing project, asks the student to conduct a tournament among the forecasting techniques and to determine, both quantitatively and qualitatively, which model should be implemented in this situation.

Spreadsheet Assignments

1. In the first assignment, the student is asked to calculate the bill amount based on the amount of energy consumed. A simple cost function is assumed. As a second part of this assignment, students are to consider the question of early replacement of heating/cooling equipment to take advantage of an incentive offered by the power company. This requires students to undertake break-even analysis.

2. The second assignment requires the student to understand the calculation of the bill amount based on the rate information supplied. Students are then asked to produce the bill amount for specific numbers of kilowatt hours used and are finally asked to reverse the process to discover the number of kilowatt hours implied by a specific bill amount. This is an advanced assignment and may incorporate integer programming and the use of Excel's Solver.

Teaching Objectives

Depending on the number of assignments made, students should be able to
- recognize seasonality and trend in a time series
- identify the existence of outliers and apply appropriate techniques for addressing them
- apply a seasonal decomposition model to a time series and use the results to calculate predicted values
- discuss the applicability of smoothing methods, determine appropriate models and their parameters, and use the results to calculate predicted values
- examine a time series to determine the parameters for an ARIMA model, use software to create the model, and use the results to calculate predicted values
- create appropriate indicator variables
- develop a multiple regression model that is statistically sound and appeals to common sense, and use that model to calculate predicted values
- learn to assess forecast errors using various fit measures
- determine and explain which, of many forecasting models, is best suited to a time series and the needs of the decision maker
- build a spreadsheet to determine the break-even point for an equipment replacement problem
- develop an integer programming model and find its solution

Courses and Levels for which the dataset is applicable

These data are suited for quantitative courses at the undergraduate or graduate level, and could be used in a high school statistics or data analysis class. Course titles may include statistics, operations management, forecasting, or operations research/management science. Instructors of lower level courses, or statistics courses that do not incorporate a statistical package, may wish to use only assignments 1, 2, 6, and 7. Instructors of computer tools classes should concentrate on assignment 1 and the spreadsheet assignments.

Note

Variable names are abbreviated in the dataset. Assignment questions are in italics and are followed by their solutions. A separate list of the assignment questions alone appears in the Appendix at the end of this manual.

Discussion Questions and Answers

Statistics and Forecasting Questions
(Reserving year 2000 data as a holdout sample)

Assignment 1 Discussion Questions
Graphing and Data Examination

1.   *Plot the time series. Look at the observations to see if any of them seem to be unusual. How might you arrange them to answer this question? Why is it important to address the issues of missing observations and outliers?*

The time series plot of original values from Minitab is shown below. The missing value in January 1994 leads Minitab to connect the December 1993 (index 36) and February 1994 (index 38) values. The value for August 1999 (index 104) was billed at $0.00.
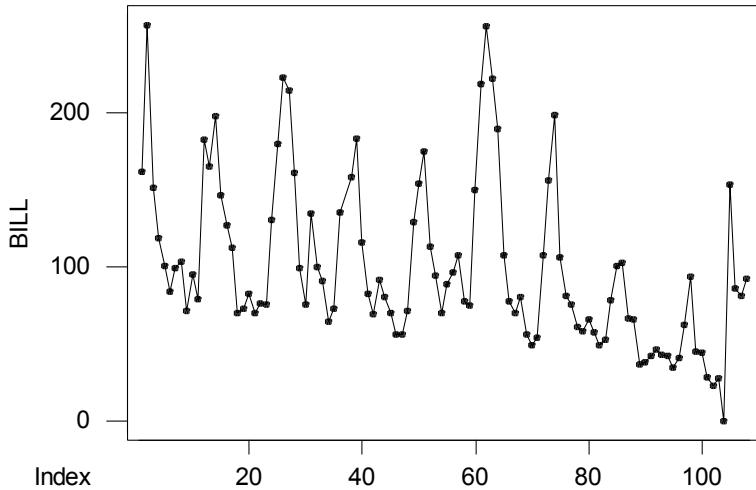
Monthly data from 1991-1999. Year 2000 data is held out.

There is pronounced seasonality in this time series, and there appears to be a noticeable downward trend.

Plot each month's values over time in order to study the consistency from one year to the next for each month. We show three months on each graph to improve legibility. From these graphs, it is easier to determine unusual monthly payments.

January, February, and March
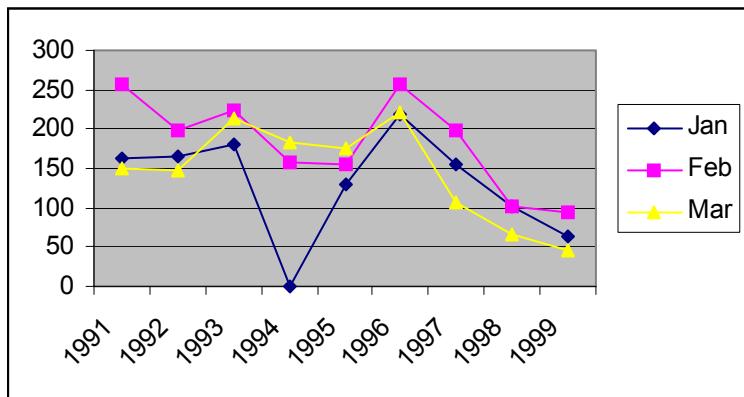
April, May, and June
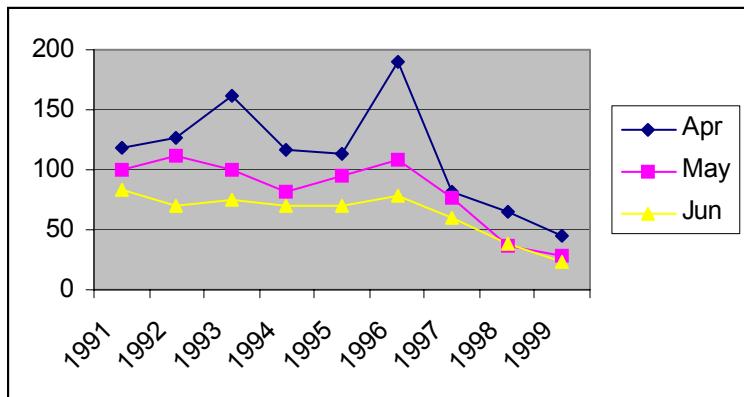
July, August, and September
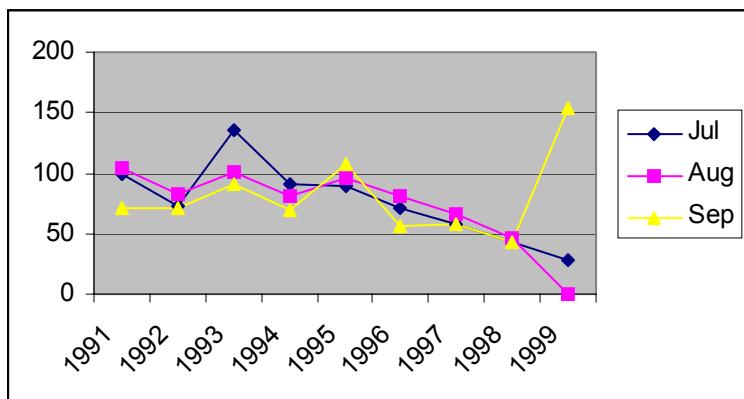
October, November, and December
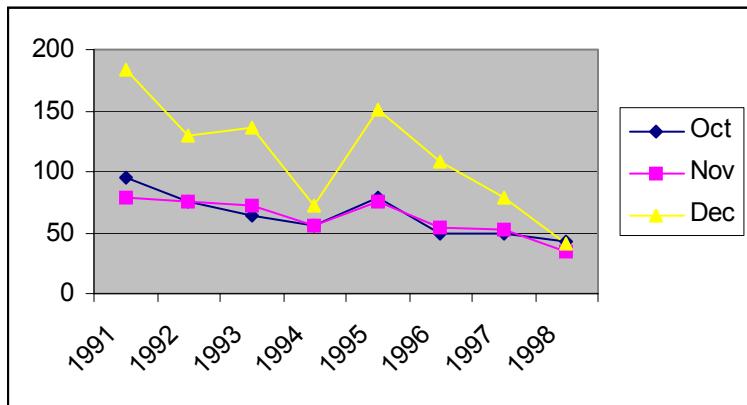
It is important to address the issue of missing values and outliers so that they do not have an undue effect on the model fitted to the time series. The values in 1994 and 1999 seem to be unusual and should be replaced with more typical values.

2.   *There is no record of payment for January 1994, and the bill for August 1999 was $0.00. Do you think the meter reader simply didn't come, and the energy cost was added to the next bill? Do you think the payments were not recorded? What other reasons might there be for the missing values?*

Discussion might provide these ideas for January
- weather could have prevented reading the meter
- the bill was paid but the amount was not recorded
- there was a billing error
- it does not appear that the February payment included the bill from January

The electric meter actually failed and was replaced during August 1999. Students might suggest this, or they might notice the larger bill in September and assume both months' usage was paid at that time.

3.   *Although it is possible to calculate the mean of the billing amount, the result would be of little value. Why is this so?*

The average amount of the bill is $98.89. Instructors should impress upon their students that the mean of a time series with trend and seasonality is a moving target. Knowing the mean would do little to help someone estimate the bill in any particular month. However, the mean of the deseasonalized data can help instructors explain analysis of variance concepts when simple linear trend is calculated from the deseasonalized values.

6

4.   Determine what should be done to handle the values for January 1994 and August 1999. Are there other months whose values seem to be outliers? Whether you decide to ignore these unusual values, or to replace them with other values, justify your choice and make the changes you deem appropriate.

The level of discussion for this question depends on the sophistication of the students. For most classes, a reasonable approach is to replace the missing value in January 1994 with the average of the adjacent January values, or 154.57. The situation in 1999 is more complex. Although the power company maintained that meters fail abruptly, it seems that the meter had been failing for some time. However, this decline in the bill also coincides with the installation of a more efficient heat pump. Without knowing more about the accuracy of the 1999 figures, a reasonable assumption could be to leave the other values as they are and to split the September, 1999, amount between August and September, giving 76.66 in each month. These are the values that will be used in the rest of the analysis.

Assignment 2 Discussion Questions
Seasonal Decomposition

1.   Based on your plot of the time series, do you feel that an additive or multiplicative seasonal decomposition model will be more appropriate?

Usually, if the seasonal swings get wider or narrower over time, a multiplicative model is warranted. There is enough change in the size of the swings that a multiplicative model may work better than an additive model, but both should be examined.

2.   Using Excel, Minitab, or another package, construct both additive and multiplicative seasonal decomposition models for this time series. Use your models to fit values for the nine years of data and to make predictions for the year 2000.

Minitab fits the trend line through the original, rather than the deseasonalized, values. If your students use a different procedure, you will need to calculate the values according to that process. Remind students that although they can make a prediction for any time period using this method, the farther into the future they look, the less confidence they should have in their results. Minitab's results are

Multiplicative Model	Additive Model
Trend Line Equation	Trend Line Equation
$Y_t = 141.630 - 0.757896*t$	$Y_t = 141.630 - 0.757896*t$
Seasonal Indices	Seasonal Indices
Period Index	Period Index
1 1.43736 2 1.73491 3 1.47847 4 1.11709	1 46.5934 2 71.0401 3 55.0598 4 12.8915

5	0.828671		5	-18.7974
6	0.647440		6	-34.6997
7	0.796264		7	-23.3735
8	0.830273		8	-19.4614
9	0.748738		9	-25.5683
10	0.613209		10	-39.6820
11	0.632026		11	-38.6364
12	1.13555		12	14.6340

Accuracy of Model

MAPE: 22.749
MAD: 20.309
MSD: 768.563

Accuracy of Model

MAPE: 24.496
MAD: 21.955
MSD: 870.871

Forecasts

Row Period Forecast

Forecasts

Row Period Forecast

Row	Period	Forecast		Row	Period	Forecast
1	109	84.833		1	109	105.613
2	110	101.079		2	110	129.302
3	111	85.018		3	111	112.564
4	112	63.390		4	112	69.638
5	113	46.396		5	113	37.191
6	114	35.758		6	114	20.531
7	115	43.374		7	115	31.099
8	116	44.598		8	116	34.253
9	117	39.651		9	117	27.388
10	118	32.009		10	118	12.517
11	119	32.512		11	119	12.804
12	120	57.553		12	120	65.317

Performance measures for the multiplicative model are better.

3.   *Calculate and record these error measurements for your models: Average Error, Mean Absolute Error, Average Percentage Error, Mean Absolute Percentage Error, and Mean Squared Error. Discuss the benefits and drawbacks of these error measurements.*

In order to calculate the error measures that are not supplied, students will need to store the residuals and do the arithmetic.

Measurement	Multiplicative Model	Additive Model
Average Error	-.663	0
Mean Absolute Error	20.309	21.955
Average Percentage Error	-7.97	-6.4
Mean Absolute Percentage Error	22.749	24.496
Mean Squared Error	768.563	870.871

The error measurements that are not influenced by the sign of the residuals—mean absolute error, mean absolute percentage error, and mean squared error—are the most useful.

Assignment 3 Discussion Questions
Smoothing Models

1. *Develop a simple exponential smoothing model for this time series, searching to find a good value for the smoothing constant. What does the size of your smoothing constant indicate about the relative emphasis on recent data?*

Minitab's optimal smoothing constant is 1.12168. You should explain to your students that even though most textbooks require the smoothing constant to be between 0 and 1, Minitab's fitting method allows values outside that range. You might point out that values of the smoothing constant near 1 will yield forecasts resembling the latest data values, while values near 0 will produce forecasts resembling the mean. In this case, the value greater than 1 arises because the data are non-stationary, and the fitted smoothing constant wants to increase the forecast *beyond* the latest data values. Also explain that single exponential smoothing looks ahead only one period to construct a forecast.

Single Exponential Smoothing

```
Smoothing Constant
Alpha: 1.12168

Accuracy Measures
MAPE:    24.63
MAD:     24.91
MSD:   1214.15

  Row   Period  Forecast     Lower      Upper

   1      109    93.6049    32.5823    154.627
   2      110    93.6049    32.5823    154.627
   3      111    93.6049    32.5823    154.627
   4      112    93.6049    32.5823    154.627
   5      113    93.6049    32.5823    154.627
   6      114    93.6049    32.5823    154.627
   7      115    93.6049    32.5823    154.627
   8      116    93.6049    32.5823    154.627
   9      117    93.6049    32.5823    154.627
  10      118    93.6049    32.5823    154.627
  11      119    93.6049    32.5823    154.627
  12      120    93.6049    32.5823    154.627
```

The results for other values of alpha indicate that larger smoothing constants are better. This means that the series needs to adjust quickly to changes. Although a large smoothing constant may yield a better fit to the past data, it is suggestive of non-stationarity. Hence, we ought to consider a more appropriate method for trended and/or seasonal data. If your students are doing this without an optimizing routine, suggest that they consider alpha values of .1, .5, and .9 as a first cut for the search. It is also useful to ask the students to consider what happens to the model when the smoothing constant is either 0 or 1. Note: when the smoothing constant is 0, the

forecast for time t is the forecast for time t-1. Therefore, all forecasts will be the same: Minitab's initial forecast value. When the smoothing constant is 1, the model reverts to the naïve model.

Alpha	MAD	MAPE	MSD
0.0	62.11	0.9591	4750.74
0.1	40.97912	0.488382	2271.865
0.5	32.38997	0.331585	1696.145
0.9	26.69394	0.264248	1288.925
1.0	25.18	0.2563	1236.02

2.   *Discuss the limitations of using simple exponential smoothing for this time series.*

Simple exponential smoothing does not work well with seasonality, so we should expect better results with another method.

3.   *If you have access to software, apply other smoothing methods to the time series. Use the resulting models to fit values to the data, and create forecasts for the year 2000. Record the performance statistics for all your models.*

Minitab Method	Parameters	MAPE	MAD	MSD
Double Exponential Smoothing	Alpha (level): 1.23091 Gamma (trend): 0.00942	25.97	26.38	1760.55
Winters' Multiplicative Model	Alpha (level): 0.2 Gamma (trend): 0.2 Delta (seasonal): 0.2	19.629	18.578	604.675
Winters' Multiplicative Model	Alpha (level): 0.5 Gamma (trend): 0.2 Delta (seasonal): 0.2	16.937	16.660	517.679

Students may find better models through their parameter choices.

Assignment 4 Discussion Questions
ARIMA Models

1.   *Develop ARIMA model(s) for this time series. Examine the ACF and PACF plots and perform any operations necessary to obtain a stationary time series. Determine the parameters that you feel are indicated by your results, and use a statistical package such as Minitab to develop your model(s). Record the MAPE, MAD, and MSE measurements. Select the most appropriate ARIMA model for this time series and justify your choice.*

Time Series Plot

Time Series Plot 1991-1999 (outliers adjusted)
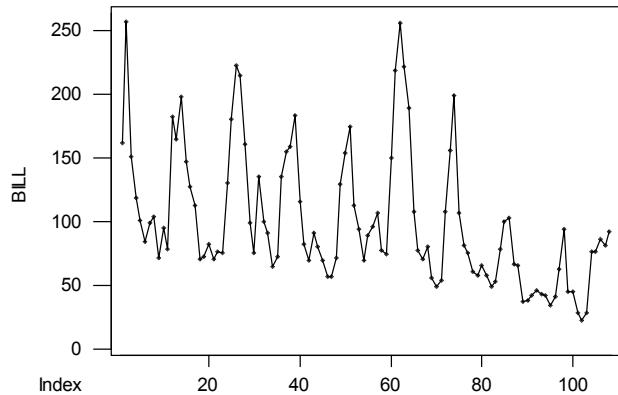
Both seasonality and trend are apparent. This is confirmed by the ACF and PACF.

Autocorrelation Function for BILL
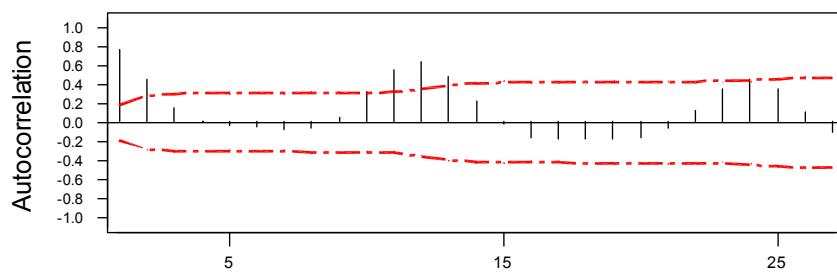
Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.77	7.95	64.98	8	-0.06	-0.40	92.29	15	-0.02	-0.11	230.12	22	0.13	0.58	253.38
2	0.45	3.19	87.89	9	0.05	0.34	92.64	16	-0.17	-0.78	233.68	23	0.35	1.59	270.50
3	0.15	0.99	90.51	10	0.33	2.08	105.59	17	-0.19	-0.87	238.20	24	0.45	2.01	299.59
4	0.01	0.04	90.51	11	0.55	3.36	142.40	18	-0.18	-0.83	242.43	25	0.35	1.49	316.85
5	-0.04	-0.25	90.69	12	0.64	3.59	193.59	19	-0.17	-0.81	246.52	26	0.11	0.47	318.66
6	-0.05	-0.32	90.98	13	0.49	2.44	223.38	20	-0.17	-0.78	250.47	27	-0.11	-0.45	320.37
7	-0.08	-0.54	91.82	14	0.23	1.09	230.05	21	-0.07	-0.33	251.17				

Partial Autocorrelation Function for BILL
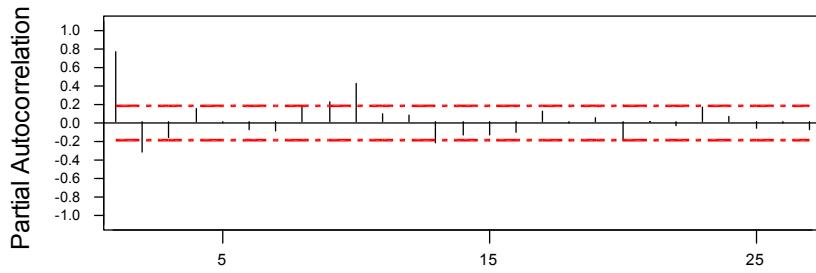
Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.77	7.95	8	0.16	1.69	15	-0.14	-1.42	22	-0.04	-0.42
2	-0.32	-3.34	9	0.23	2.35	16	-0.10	-1.06	23	0.17	1.76
3	-0.16	-1.65	10	0.43	4.45	17	0.13	1.33	24	0.07	0.74
4	0.15	1.56	11	0.09	0.92	18	-0.01	-0.09	25	-0.06	-0.63
5	-0.01	-0.09	12	0.08	0.80	19	0.05	0.55	26	-0.01	-0.06
6	-0.08	-0.84	13	-0.22	-2.32	20	-0.20	-2.04	27	-0.08	-0.79
7	-0.09	-0.93	14	-0.14	-1.42	21	0.01	0.10			

Remind the students of the meaning of stationarity in the mean and in the variation. Explain that although it possible to spot non-stationarity in the original time series plot, examination of the ACF and PACF will provide a more rigorous conclusion by looking for values that are significantly different from 0. When the ACF does not drop to 0 within several lags, the time series is not stationary. Removing non-stationarity is necessary to produce an adequate ARIMA model.

Based on these plots, we will first examine seasonal (12 month lag) differences.

 Plot with seasonal differences

First Seasonal (12) Differences

12
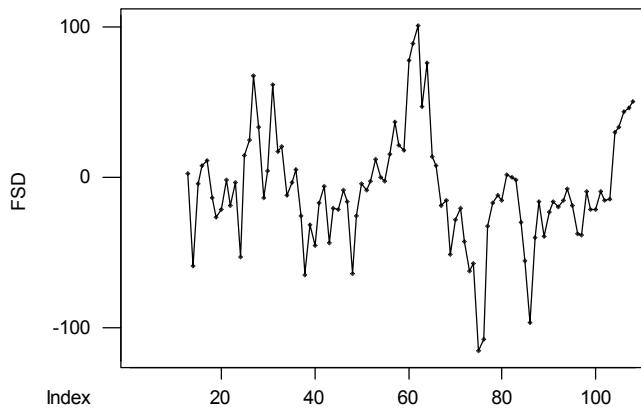
Autocorrelation Function for FSD
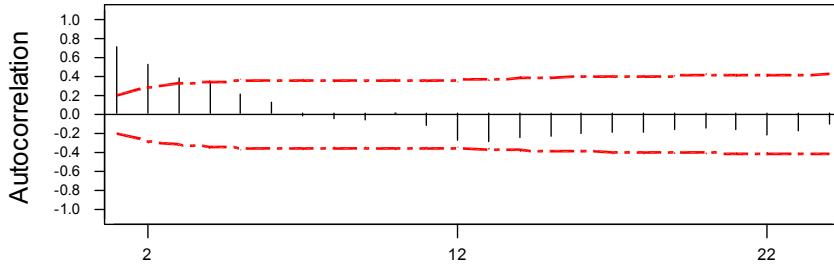
| Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.71	6.91	49.27	8	-0.05	-0.27	108.53	15	-0.24	-1.21	143.39	22	-0.23	-1.09	174.08
2	0.52	3.63	76.63	9	-0.07	-0.39	109.06	16	-0.21	-1.06	148.62	23	-0.18	-0.87	178.38
3	0.39	2.37	91.72	10	0.00	0.01	109.06	17	-0.20	-0.99	153.36	24	-0.11	-0.50	179.85
4	0.33	1.89	102.54	11	-0.12	-0.68	110.76	18	-0.19	-0.95	157.82				
5	0.20	1.14	106.80	12	-0.28	-1.55	119.70	19	-0.16	-0.81	161.15				
6	0.12	0.64	108.20	13	-0.30	-1.60	129.85	20	-0.15	-0.71	163.81				
7	-0.03	-0.16	108.28	14	-0.25	-1.28	136.85	21	-0.17	-0.82	167.47				

Partial Autocorrelation Function for FSD
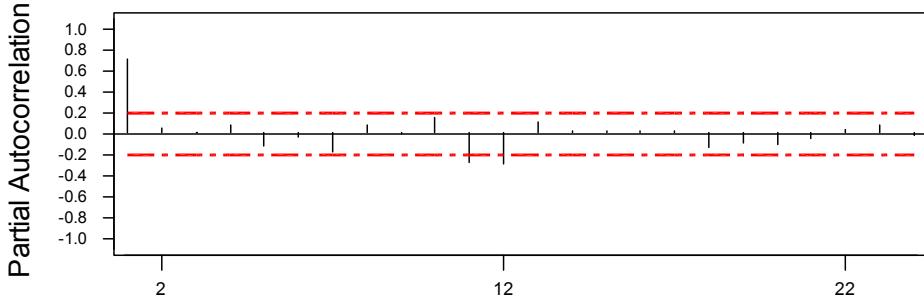
Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.71	6.91	8	0.08	0.77	15	0.02	0.20	22	0.03	0.28
2	0.05	0.49	9	-0.00	-0.04	16	0.02	0.20	23	0.08	0.76
3	0.00	0.01	10	0.15	1.52	17	0.03	0.27	24	-0.02	-0.15
4	0.08	0.78	11	-0.28	-2.78	18	-0.14	-1.35			
5	-0.12	-1.21	12	-0.29	-2.82	19	-0.09	-0.88			
6	-0.03	-0.29	13	0.10	0.99	20	-0.10	-1.00			
7	-0.18	-1.77	14	0.02	0.18	21	-0.05	-0.50			

The ACF has a decline in the first few periods, and shows a decline around period 12, although these values are not significant. The PACF has a significant spike at lags 1 and 12. When the ACF declines and the PACF has a single spike, the indication is for an AR(1) model. Because these patterns repeat at the seasonal lag of 12, a seasonal AR(1) component is appropriate. The result is a seasonal ARIMA (1,0,0) (1,1,0) model with period 12. (The simpler ARIMA (1,0,0)(0,1,0) model converges but its mean squared error is 705, compared to 637.)

13

ARIMA Model

```
ARIMA model for BILL

Estimates at each iteration
Iteration        SSE      Parameters
    0          126324    0.100    0.100    -6.390
    1           98270    0.250    0.028    -5.768
    2           78310    0.400   -0.061    -5.023
    3           65949    0.550   -0.166    -4.088
    4           60744    0.700   -0.291    -2.884
    5           60513    0.726   -0.332    -2.538
    6           60498    0.725   -0.343    -2.533
    7           60496    0.724   -0.348    -2.548
    8           60496    0.724   -0.349    -2.554
    9           60496    0.724   -0.350    -2.556
   10           60496    0.724   -0.350    -2.558
Relative change in each estimate less than  0.0010


Final Estimates of Parameters
Type          Coef        StDev           T          P
AR    1      0.7236      0.0733         9.87      0.000
SAR 12      -0.3501      0.1007        -3.48      0.001
Constant    -2.558       2.582         -0.99      0.324


Differencing: 0 regular, 1 seasonal of order 12
Number of observations:  Original series 108, after differencing 96
Residuals:    SS =  59278.2  (backforecasts excluded)
              MS =    637.4  DF = 93

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                 12        24        36        48
Chi-Square        25.0      35.9      42.9      56.2
DF                   9        21        33        45
P-Value          0.003     0.022     0.115     0.123
```

This model seems to be a good one. It converged, the t statistics are significant, and the Ljung-Box Q statistic indicates that we cannot reject the possibility of white noise. Because the t statistic for the constant term is not strong and differences have been taken, run again without the constant term.

ARIMA Model

```
ARIMA model for BILL

Estimates at each iteration
Iteration        SSE      Parameters
    0          130266    0.100    0.100
    1          101879    0.250    0.038
    2           81197    0.400   -0.041
    3           67873    0.550   -0.139
    4           61625    0.700   -0.259
```

```
    5              61085     0.746    -0.318
    6              61061     0.745    -0.332
    7              61057     0.744    -0.338
    8              61056     0.744    -0.340
    9              61056     0.744    -0.341
   10              61056     0.744    -0.341
Relative change in each estimate less than  0.0010


Final Estimates of Parameters
Type          Coef        StDev          T        P
AR    1      0.7440      0.0703      10.58    0.000
SAR 12      -0.3411      0.1005      -3.40    0.001


Differencing: 0 regular, 1 seasonal of order 12
Number of observations:  Original series 108, after differencing 96
Residuals:    SS =   59966.2  (backforecasts excluded)
              MS =    637.9  DF = 94


Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                 12        24        36        48
Chi-Square        25.8      36.2      43.4      56.9
DF                  10        22        34        46
P-Value          0.004     0.029     0.130     0.130
```

The final model ARIMA(1,0,0)(1,1,0) seems to be a good choice based on its results.

Error Measurements

Mean Absolute Error	17.9859
Mean Squared Error	624.6477
Mean Absolute Percentage Error	19.02%

Forecasted values for the year 2000 are

```
Forecasts from period 108
                              95 Percent Limits
Period      Forecast        Lower        Upper
 109        104.087        54.572      153.602
 110        118.196        56.480      179.913
 111         68.523         0.993      136.052
 112         63.617        -6.925      134.159
 113         40.011       -32.144      112.167
 114         34.503       -38.531      107.536
 115         37.760       -35.755      111.275
 116         69.878        -3.902      143.659
 117         67.824        -6.103      141.750
 118         73.008        -0.999      147.016
 119         66.597        -7.455      140.649
 120         75.799         1.722      149.876
```

Assignment 5 Discussion Questions
Multiple Regression with Indicator Variables

1.   *Build a spreadsheet that incorporates twelve new columns. One of these is a column for time, represented by the observation number from 1 to 108. The other eleven are indicator (dummy) variables for January through November. If the observation is taken from that month, then the value is 1. Otherwise, the value is 0.*

December is the base month. If all eleven dummy variables are 0, then by default the observation is from December. Encourage students to use Minitab's Make Indicator Variables command to establish the dummy variable columns. Note: they will need to delete the automatically produced December column. Be sure they understand why.

2.   *Using time as one independent variable and eleven indicator (dummy) variables for the month, develop a multiple regression model to predict the electric bill. Justify your model based on its statistical results. Use your model to fit values for the nine-year period and calculate performance measurements. Use your model to predict the bills for 2000.*

The regression model shows that all months except April seem to be significant.

Regression Analysis: BILL versus OBSERVATION NUMBER, Jan, ...

```
The regression equation is
BILL = 150 - 0.671 OBSERVATION NUMBER + 30.3 Jan + 65.8 Feb + 29.7 Mar
          - 2.3 Apr - 32.7 May - 50.9 Jun - 37.0 Jul - 31.3 Aug - 40.5 Sept
          - 45.0 Oct - 46.1 Nov

Predictor        Coef      SE Coef           T          P
Constant       150.20        11.76       12.77      0.000
OBSERVAT      -0.67062      0.09565       -7.01      0.000
Jan             30.29        14.56        2.08      0.040
Feb             65.79        14.55        4.52      0.000
Mar             29.68        14.54        2.04      0.044
Apr             -2.29        14.54       -0.16      0.875
May            -32.66        14.53       -2.25      0.027
Jun            -50.87        14.53       -3.50      0.001
Jul            -37.01        14.53       -2.55      0.012
Aug            -31.25        14.52       -2.15      0.034
Sept           -40.49        14.52       -2.79      0.006
Oct            -44.97        14.52       -3.10      0.003
Nov            -46.12        14.52       -3.18      0.002

S = 30.80      R-Sq = 69.1%     R-Sq(adj) = 65.2%

Analysis of Variance

Source           DF          SS          MS          F          P
Regression       12      201608       16801      17.71      0.000
Residual Error   95       90108         949
Total           107      291716
```

Error Measurements

Mean Absolute Error	21.357
Mean Squared Error	834.334
Mean Absolute Percentage Error	23.88%

Forecasts for the year 2000

Year 2000 Forecasts	
Month	Y-hat
Jan	107.39
Feb	142.22
Mar	105.44
Apr	72.80
May	41.76
Jun	22.88
Jul	36.07
Aug	41.16
Sept	31.25
Oct	26.10
Nov	24.28
Dec	69.73

Assignment 6 Discussion Questions:
Incorporating Additional Information with Multiple Regression

1.   *Begin building a spreadsheet that incorporates the time series with the additional information. What effect do you anticipate that each of these potential independent variables will have on the amount of the electric bill?*

Add a variable for the number of people in the household and dummy variables to reflect the installation of the new meter and heat pumps.

2.   *Information is available for the average temperature recorded each month. Would this information provide a useful explanatory variable for the multiple regression model? If so, add it to the model.*

This would not be a good choice because energy usage is not linearly related to temperature. Usage is high for low and high temperatures, and low for medium temperatures.

3.   *Heating Degree Days and Cooling Degree Days are concepts that relate the temperature to energy usage. For USA Today's description of these calculations, visit http://167.8.29.15/weather/askjack/waskdays.htm. Search for monthly values for these measurements in the vicinity of Terre Haute, Indiana, and add them to your spreadsheet, or use the values in the dataset.*

The values provided are from one of the many NOAA web sites that give weather information. The values are long term averages for each month and so are appropriate for use in this sort of model.

4.   *What other explanatory variables do you feel would be useful for this time series? Discuss what you would expect to happen if you had this information, and specify where you would obtain the information.*

Students might consider vacation dates, changes in the utility rates, and some sort of indicator variable for the failing electric meter.

5.   *Use your spreadsheet to develop an effective multiple regression model to use to predict the electric bill. Justify your model based on its statistical results. Use your model to fit values for the nine-year period and calculate performance measurements. Use your model to predict the bills for the year 2000.*

A wide variety of models may be chosen. We will show the results from three cases.

Case I   Including the causal variables and ignoring trend and monthly indicators

Regression Analysis: BILL versus HEATING DEGREE D, COOLING DEGREE D, ...

```
The regression equation is
BILL = - 57.9 + 0.0784 HEATING DEGREE DAYS + 0.0283 COOLING DEGREE DAYS
          + 33.9 NUMBER IN HOUSEHOLD + 49.2 METER? - 24.7 PUMP 1?
          - 26.0 PUMP 2?

Predictor         Coef     SE Coef           T         P
Constant         -57.90       54.75       -1.06     0.293
HEATING         0.07842     0.01211        6.47     0.000
COOLING         0.02830     0.04801        0.59     0.557
NUMBER I          33.92       13.74        2.47     0.015
METER?            49.25       21.26        2.32     0.023
PUMP 1?          -24.66       10.66       -2.31     0.023
PUMP 2?          -26.02       13.47       -1.93     0.056

S = 35.43      R-Sq = 56.5%     R-Sq(adj) = 53.9%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         6      164901       27484     21.89     0.000
Residual Error   101      126815        1256
Total            107      291716
```

Case II   Including the causal variables and eleven monthly indicators

Regression Analysis: BILL versus HEATING DEGREE D, COOLING DEGREE D, ...

```
* Oct is highly correlated with other X variables
* Oct has been removed from the equation
```

18

```
* Nov is highly correlated with other X variables
* Nov has been removed from the equation

BILL = - 97.2 + 0.112 HEATING DEGREE DAYS + 2.08 COOLING DEGREE DAYS
          + 26.6 NUMBER IN HOUSEHOLD + 71.5 METER? - 32.6 PUMP 1?
          - 22.7 PUMP 2? + 19.9 Jan + 80.6 Feb + 73.5 Mar + 81.4 Apr
          - 128 May - 373 Jun - 603 Jul - 458 Aug - 158 Sept

Predictor         Coef       SE Coef          T          P
Constant         -97.22        48.75       -1.99      0.049
HEATING         0.11222      0.03027        3.71      0.000
COOLING           2.084        1.056        1.97      0.051
NUMBER I          26.57        10.54        2.52      0.013
METER?            71.54        15.92        4.50      0.000
PUMP 1?         -32.616        8.000       -4.08      0.000
PUMP 2?         -22.663        9.973       -2.27      0.025
Jan               19.94        15.81        1.26      0.210
Feb               80.58        11.71        6.88      0.000
Mar               73.54        11.29        6.51      0.000
Apr               81.41        17.93        4.54      0.000
May             -127.71        81.22       -1.57      0.119
Jun             -372.6         200.2       -1.86      0.066
Jul             -602.8         323.4       -1.86      0.066
Aug             -458.1         252.8       -1.81      0.073
Sept            -158.31        95.31       -1.66      0.100

S = 26.01       R-Sq = 78.7%     R-Sq(adj) = 75.2%

Analysis of Variance

Source            DF            SS            MS          F          P
Regression        15        229485         15299      22.62      0.000
Residual Error    92         62232           676
Total            107        291716
```

Case III   Including all causal variables, monthly indicators, and the observation number.
The results from Case II indicated the possibility of multicollinearity between the October and
November indicators and other variables. The correlations between October and November and
the other individual values aren't particularly large, but Minitab's warning indicates that these
variables may be linear combinations of the other variables and so should be deleted. The
correlation between Heating Degree Days and Cooling Degree Days is -.758. Although no
warnings were provided, instructors may want to have students run a model that deletes one of
these variables.

Regression Analysis: BILL versus HEATING DEGREE D, COOLING DEGREE D, ...

```
The regression equation is
BILL = - 90.3 + 0.113 HEATING DEGREE DAYS + 2.08 COOLING DEGREE DAYS
          + 26.4 NUMBER IN HOUSEHOLD + 72.2 METER? - 25.1 PUMP 1?
          - 20.1 PUMP 2? + 18.5 Jan + 79.4 Feb + 72.6 Mar + 80.8 Apr
          - 128 May - 372 Jun - 602 Jul - 457 Aug - 158 Sept
          - 0.163 OBSERVATION NUMBER

Predictor         Coef       SE Coef          T          P
Constant         -90.28        49.09       -1.84      0.069
HEATING         0.11262      0.03024        3.72      0.000
```

```
COOLING          2.082      1.054       1.97      0.051
NUMBER I        26.37      10.53        2.50      0.014
METER?          72.22      15.91        4.54      0.000
PUMP 1?        -25.05      10.50       -2.39      0.019
PUMP 2?        -20.12      10.22       -1.97      0.052
Jan             18.46      15.85        1.16      0.247
Feb             79.35      11.75        6.75      0.000
Mar             72.58      11.31        6.42      0.000
Apr             80.76      17.91        4.51      0.000
May           -127.89      81.12       -1.58      0.118
Jun           -372.4      200.0        -1.86      0.066
Jul           -602.1      323.0        -1.86      0.066
Aug           -457.4      252.4        -1.81      0.073
Sept          -157.90      95.19       -1.66      0.101
OBSERVAT        -0.1631     0.1471      -1.11      0.270


S = 25.98      R-Sq = 79.0%     R-Sq(adj) = 75.3%


Analysis of Variance

Source           DF          SS          MS         F        P
Regression       16      230315        14395     21.33    0.000
Residual Error   91       61402          675
Total           107      291716
```

Conclusion: For this analysis, we will use a model that includes all variables from Case III that have a p value less than .10. This maintains the set of causal variables and six of the eleven monthly indicators. It deletes the observation number and so does not consider a trend component. This appears to be a reasonable model.

Regression Analysis: BILL versus HEATING DEGREE D, COOLING DEGREE D, ...

```
The regression equation is
BILL = - 86.3 + 0.104 HEATING DEGREE DAYS + 0.517 COOLING DEGREE DAYS
         + 28.5 NUMBER IN HOUSEHOLD + 68.0 METER? - 31.3 PUMP 1?
         - 22.4 PUMP 2? + 70.1 Feb + 60.9 Mar + 65.8 Apr - 61.2 Jun
         - 108 Jul - 68.3 Aug


Predictor       Coef       SE Coef          T         P
Constant      -86.34        42.22        -2.04     0.044
HEATING         0.10406      0.01458       7.14     0.000
COOLING         0.5173       0.1387        3.73     0.000
NUMBER I       28.50        10.40         2.74     0.007
METER?         68.05        16.17         4.21     0.000
PUMP 1?       -31.264        8.071       -3.87     0.000
PUMP 2?       -22.41        10.16        -2.21     0.030
Feb            70.11        10.25         6.84     0.000
Mar            60.90        10.30         5.91     0.000
Apr            65.82        12.47         5.28     0.000
Jun           -61.18        20.10        -3.04     0.003
Jul          -108.10        34.81        -3.11     0.003
Aug           -68.35        26.15        -2.61     0.010


S = 26.63      R-Sq = 76.9%     R-Sq(adj) = 74.0%


Analysis of Variance

Source           DF          SS          MS         F        P
Regression       12      224345        18695     26.36    0.000
Residual Error   95       67372          709
```

```
Total            107      291716
```

Error Measurements

Mean Absolute Error	19.28
Mean Squared Error	623.81
Mean Absolute Percentage Error	22.63%

Forecasts for the year 2000

Year 2000 Forecasts	
Month	Y-hat
Jan	141.43
Feb	187.60
Mar	150.82
Apr	118.17
May	107.52
Jun	61.92
Jul	75.11
Aug	80.20
Sep	47.35
Oct	28.27
Nov	52.05
Dec	94.19

Assignment 7 Discussion Questions
***Error Analysis, Holdout Sample, and Model Selection***

1.   Combine the statistical results from all of your forecasting models into a table that will let you compare performance measurements. Which model appears to do the best job?

2.   Data from 2000 is shown below. Determine the mean absolute error, the mean absolute percentage error, and the mean squared error for each of your models during the year 2000
.

The chart below shows all summary statistics for the fitted and forecasted values. Based on these results, it appears as though the Winters and ARIMA model performed the best during the within sample period, and the Exponential and Causal models performed better than the other models during the hold-out period.

	Multiplicative Model	Additive Model	Simple Exp. Smoothing	Winters	ARIMA	Regression Models	
Measurement			Alpha: 1.12168	.5, .2, .2		Trend and seasonal	Causal
Mean Absolute Error	20.309	21.955	24.63	16.66	17.986	21.36	19.28

21

Average Percentage Error	-7.97	-6.4	-4.6615	-.726	-5.298	-7.83	-4.43
Mean Absolute Percentage Error	22.749	24.496	24.63	16.937	19.02	23.88	22.63
Mean Squared Error	768.563	870.871	1214.15	517.679	624.65	834.33	623.81

Actual, 2000	Forecasts						
	Multiplicative Model	Additive Model	Simple Exp. Smoothing	Winters	ARIMA	Trend and Seasonal	Causal
126.86	84.83	105.61	93.60	175.62	104.09	107.39	141.43
155.45	101.08	129.30	93.60	228.73	118.20	142.22	187.60
121.19	85.02	112.56	93.60	181.81	68.52	105.44	150.82
190.62	63.39	69.64	93.60	159.41	63.62	72.80	118.17
115.7	46.40	37.19	93.60	119.63	40.01	41.76	107.52
81.35	35.76	20.53	93.60	100.38	34.50	22.88	61.92
91.23	43.37	31.10	93.60	122.22	37.76	36.07	75.11
86.28	44.60	34.25	93.60	145.78	69.88	41.16	80.20
83.97	39.65	27.39	93.60	126.95	67.82	31.25	47.35
77.65	32.01	12.52	93.60	118.96	73.01	26.10	28.27
72.3	32.51	12.80	93.60	119.18	66.60	24.28	52.05
89.49	57.55	65.32	93.60	227.72	75.80	69.73	94.19
MAD	52.16	52.82	26.23	49.73	39.36	47.59	25.80
MAPE	48.64	52.80	20.91	50.45	33.42	47.00	24.23
MSD	3314.83	3633.27	1384.96	3510.99	2696.98	3065.20	1027.14

The chart below compares the Mean Absolute Percentage Errors.
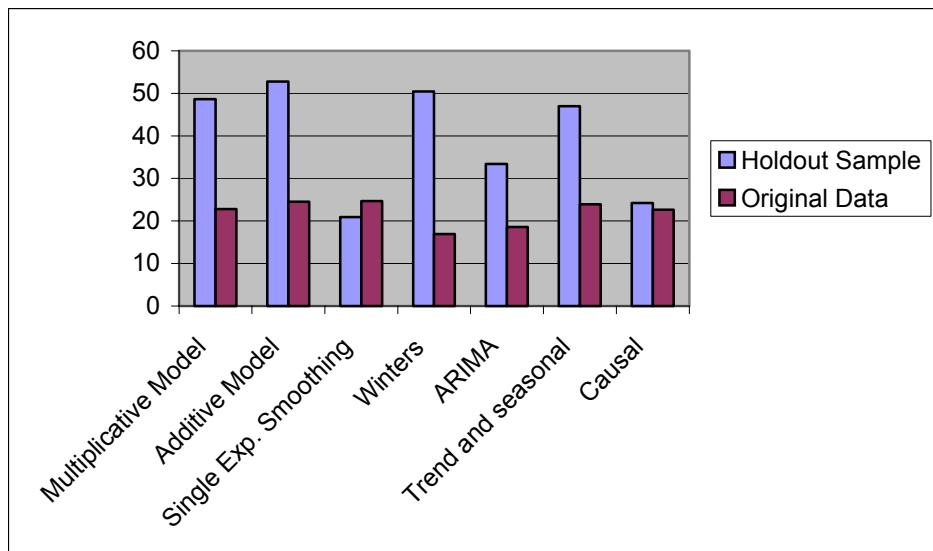
3.   *Based on the results above, and considering the practicality of each of the methods you employed, which forecasting technique would you recommend for this time series? Support your answer.*

The multivariate model has some appeal due to the many apparent causes of energy usage. It also seems reasonable to consider trying different combinations of parameters for the Winters model and comparing the results to ARIMA models. Further analysis might lead to adjustment of the values that seem unusually low in the months prior to the meter replacement.

Spreadsheet Questions

Spreadsheet Assignment 1 (using a subset of the data)

The family has a 13-year-old heat pump (type of energy efficient furnace and air conditioner) that is nearing the end of its useful life, estimated at 8-15 years. Older units require more expensive repairs and are less energy-efficient than newer models. In fact, the new heat pump is estimated to use 30% less energy than the old one.

Assume, on average, that $55 (before tax dollars) of the monthly bill are not *due to heating or cooling requirements (and thus not affected by the heat pump replacement). For this assignment, assume that electricity costs a flat rate of 6 cents per kilowatt hour*

Prepare a table that shows the kilowatt hours used and the expected bill for each month of the year. Add 5% for sales tax, and format all amounts appropriately. Projected kilowatt hours used for the next year are shown below:

Month	KWHs
Jan	3104
Feb	3236
Mar	2934
Apr	2234
May	1987
Jun	1255
Jul	1380
Aug	1655
Sep	1555
Oct	2009
Nov	2256
Dec	2680

Your table should also calculate the average monthly bill for the year and the average monthly bill just for the period October – March.

The cost of a new heat pump is $3150 installed. The local electric utility company offers a $300 rebate in a program that is ending on the last day of this month. This rebate lowers the cost of the unit and encourages customers to be more energy efficient, lowering the capacity that the electric utility must provide at peak times.

Further, the electric utility offers a 20% discount on electric bills from October through March; however, the first 1,000 kilowatt hours are not discounted.

Should the family replace the heat pump now to earn the $300 rebate even though it is still working? How long would it take to pay back the cost of the unit in energy savings and discounted electricity rates? Base your calculation on average monthly cost with and without the replacement heat pump.

As a check on your work, you should find that in a representative year, without replacement the family averaged about $170 per month from October through March with a peak of about $204 in March.

Without Replacement

Parameters

Cost/Kwh	$	0.060
Sales Tax		5%
Non H/AC	$	55.00 (pre tax) = 916.7 kwh

With Replacement

Parameters

Discount	20%
Base Kwhs	1000
Efficiency	30%

Month	KWHs Used	KWH Cost	Monthly Bill	Projected KWHs Used	KWH Cost	Monthly Bill
Jan	3,104.00	$ 186.24	$ 195.55	2,447.80	$ 129.49	$ 135.97
Feb	3,236.00	$ 194.16	$ 203.87	2,540.20	$ 133.93	$ 140.63
Mar	2,934.00	$ 176.04	$ 184.84	2,328.80	$ 123.78	$ 129.97
Apr	2,234.00	$ 134.04	$ 140.74	1,838.80	$ 110.33	$ 115.84
May	1,987.00	$ 119.22	$ 125.18	1,665.90	$ 99.95	$ 104.95
Jun	1,255.00	$ 75.30	$ 79.07	1,153.50	$ 69.21	$ 72.67
Jul	1,380.00	$ 82.80	$ 86.94	1,241.00	$ 74.46	$ 78.18
Aug	1,655.00	$ 99.30	$ 104.27	1,433.50	$ 86.01	$ 90.31
Sep	1,555.00	$ 93.30	$ 97.97	1,363.50	$ 81.81	$ 85.90
Oct	2,009.00	$ 120.54	$ 126.57	1,681.30	$ 92.70	$ 97.34
Nov	2,256.00	$ 135.36	$ 142.13	1,854.20	$ 101.00	$ 106.05
Dec	2,680.00	$ 160.80	$ 168.84	2,151.00	$ 115.25	$ 121.01
Total	26,285.00	$1,577.10	$1,655.96	21,699.50	$1,217.93	$1,278.83
Average	2,190.42	$ 131.43	$ 138.00	1,808.29	$ 101.49	$ 106.57
O-M Avg	2,703.17	$ 162.19	$ 170.30	2,167.22	$ 116.03	$ 121.83

Break-even Analysis:		
HW Cost	$	3,150
Rebate	$	300
Net Cost	$	2,850
Avg Bill Now	$	138.00
New Avg Bill	$	106.57
Savings/mo	$	31.43
Months/BE		90.69
Years/BE		7.56

Spreadsheet Assignment 2
How are the billing amounts determined? The Rider Total column in the dataset and the table below provide rate information obtained from the utility company.

1.   *Use the tiered rate information in the table below and the Rider Total from the dataset to calculate the billing amount for December, 1999, for the following numbers of kilowatt hours used: 100, 500, 750, and 1250. The sales tax rate is 5%.*

Tiered Rate Information

For Jan 1991 to August 2000	BASE	13-100 kwh	101-200 kwh	201-600 kwh	601-1000 kwh	> 1000 kwh
	$7.20	$0.07776	$0.04887	$0.04202	$0.03355	$0.02566
Beginning Sept 2000	BASE	0-300 kwh	300-1000 kwh	> 1000 kwh		
	$8.15	$0.081711	$0.04851	$0.04019		

Legend
BASE is fixed amount of bill regardless of consumption
Other $ amounts are the cost per kilowatt-hour used that each month

Solution

YEAR	MONTH	KWH used	0-12 kwh	13-100 kwh	101-200 kwh	201-600 kwh	601-1000 kwh	> 1000 kwh	Total Riders		
1999	Dec		7.2	0.07776	0.04887	0.04202	0.03355	0.02566	0.000584	**Pre-Tax**	**Bill**
		100	7.2	6.84288	0	0	0	0	0.0584	$ 14.10	$ 14.81
		500	7.2	6.84288	4.887	12.606	0	0	0.292	$ 31.83	$ 33.42
		750	7.2	6.84288	4.887	16.808	5.0325	0	0.438	$ 41.21	$ 43.27
		1250	7.2	6.84288	4.887	16.808	13.42	6.415	0.73	$ 56.30	$ 59.12

2.   *Using the IF function, define a formula that will correctly calculate the billing amount for any number of kilowatt hours used. Calculate the billing amount for 100, 200, 600, and 1000 hours based on the December, 1999, rate information. Demonstrate how you could use these amounts to determine in which category the number of kwh falls. Subtract to find the portion of the bill that applies to the hours in the last category. Determine the total number of hours used.*

Solution

YEAR	MONTH	KWH used	0-12 kwh	13-100 kwh	101-200 kwh	201-600 kwh	601-1000 kwh	> 1000 kwh	Total Riders		
1999	Dec		7.2	0.07776	0.04887	0.04202	0.03355	0.02566	.000584	**Pre-Tax**	**Bill**
		100	7.2	6.84288	0	0	0	0	0.0584	$ 14.10	$ 14.81
		200	7.2	6.84288	4.887	0	0	0	0.1168	$ 19.05	$ 20.00

| | | 600 | | 7.2 | 6.84288 | 4.887 | 16.808 | 0 | | 0 | 0.3504 | $ 36.09 | $ 37.89 |
| | | 1000 | | 7.2 | 6.84288 | 4.887 | 16.808 | 13.42 | | 0 | 0.584 | $ 49.74 | $ 52.23 |

To use the billing amounts from the category breaks, do the following.

a.   Let x = the unknown number of kwh

b.   If Bill > $52.23, then the consumption is over 1000 kwh. The first 1000 hours would be
billed at $49.74, the hours over 1000 at .02566 per kwh, and all hours would be subject to
the rider charge. The rider charge for the first 1000 hours is included in the $49.74.

Bill = 1.05(49.74 + (x-1000)(.02566) + (x-1000)(.000584))

or x = 1000 + [(Bill/1.05) – 49.74] / .026244

For a bill of $65.00, the consumption is 1464 kwh

c.   When $37.89 ≤ Bill ≤ $52.23, the consumption is between 600 and 1000.

Bill = 1.05(36.09 + (x-600)(.03355) + (x-600)(.000584))

For a bill of $45.00, the consumption is

x = [{Bill/1.05)-15.6966] / .034134 = 796.

d.   Continue in a similar fashion for smaller bills.

3.   *Develop an integer programming model that will convert the billing amount to the total
number of kilowatt hours used. Test your model for December 1999 using a bill of
$59.12.*

Solution

Determine the billing categories

Category	If total is between	Hours in category	Charge per hour
0	0-12	Fixed charge	c_0 (base cost)
1	13-100	88	c_1
2	101-200	100	c_2
3	201-600	400	c_3
4	601-1000	400	c_4
5	over 1000	unlimited	c_5

Define the following variables

Let x_i = the number of hours billed at rate category i
        y_i = 1 if rate category i is used, 0 if not

c_i = the cost per kilowatt hour in category i
r_j = the cost per kilowatt hour due to rider j
x = total hours used

Minimize $\Sigma\ y_i$

s.t. $x_1 \leq 88y_1$ $x_1 \geq 88y_2$
         $x_2 \leq 100y_2$ $x_2 \geq 100y_3$
         $x_3 \leq 400y_3$ $x_3 \geq 400y_4$
         $x_4 \leq 400y_4$ $x_4 \geq 400y_5$
         $x_5 \leq 5000y_5$

         $y_1 \geq y_2$
         $y_2 \geq y_3$
         $y_3 \geq y_4$
         $y_4 \geq y_5$

         $x = \Sigma\ x_i + 12$

         Net Bill = (Bill amount/1.05) – c_0 = $\Sigma\ c_i\ x_i + x\ (\Sigma\ r_j)$

         $x_i \geq 0$
         y_i are binary 0/1

Entering the IP problem into Solver results in the solution below, confirming the total consumption at 1250 kwh.

Model (Using December 1999 rate data)

		Rate Bin	For a total	Hours in group		Rider Sum		0.000584
Let xi = number of units billed at rate i		0	0-12	12		Bill	59.12 Net bill	49.10476
Let yi = 1 if rate i is used, 0 if not		1	13-100	88				
Let x = sum of the xi's + 12 = total hours used		2	101-200	100				
		3	201-600	400				
		4	601-1000	400				
		5	over 1000	unlimited				

	x1	x2	x3	x4	x5	x	y1	y2	y3	y4	y5		Objective Function	
Variables														
Values	88	100	400	400	250.0717	1250.072	1	1	1	1	1		Z =	5

Teaching Suggestions

When this set of forecasting assignments was class tested in a college senior level forecasting elective, the original data was posted on the class Web site. The dataset was used throughout the semester, with the successive forecasting assignments (1 through 6) given as we covered each specific technique in the classroom. There is enough ambiguity in the data that students are intrigued to see which model will do the best job with the data. Use in the business statistics

class was limited to assignments 1 and 2. We found that this dataset provided useful alternatives to the textbook exercises.

The spreadsheet break-even assignment is a difficult problem. Determining the bill without replacement (and assuming a fixed charge) is easy, but determining the electric bill for the replacement problem is tricky. First determine the number of kilowatts that are required for non-heating and air conditioning because they will not be reduced by the more efficient heat pump. Reduce the heating and cooling kilowatts by the efficiency factor. Then determine whether this is a month in which the discount (Oct-Mar) is effective. Finally, the first 1000 kilowatts are not discounted. The students who completed this assignment were enrolled in a second-level business computer tools class.

The spreadsheet bill conversion problem has not been class tested but forms an interesting and challenging problem. Students should be in at least a second level business computer tools class, and will need to have studied 0-1 integer programming in a management science class if they are asked to formulate the model in the second part of this question.

We would suggest that you assign only those parts that fit well with your curriculum and the software your students are accustomed to using. Supply time for discussion, both before and after their analysis. These assignments work well for preparation by small teams.

Appendix

Discussion Questions
(Reserving year 2000 data as a holdout sample)

Assignment 1 Discussion Questions
Graphing and Data Examination

1.   Plot the time series. Look at the observations to see if any of them seem to be unusual. How might you arrange them to answer this question? Why is it important to address the issues of missing observations and outliers?

2.   There is no record of payment for January 1994, and the bill for August 1999 was $0.00. Do you think the meter reader simply didn't come, and the energy cost was added to the next bill? Do you think the payments were not recorded? What other reasons might there be for the missing values?

3.   Although it is possible to calculate the mean of the billing amount, the result would be of little value. Why is this so?

4.   Determine what should be done to handle the values for January 1994 and August 1999. Are there other months whose values seem to be outliers? Whether you decide to ignore these unusual values, or to replace them with other values, justify your choice and make the changes you deem appropriate.

Assignment 2 Discussion Questions
Seasonal Decomposition

1.   Based on your plot of the time series, do you feel that an additive or multiplicative seasonal decomposition model will be more appropriate?

2.   Using Excel, Minitab, or another package, construct both additive and multiplicative seasonal decomposition models for this time series. Use your models to fit values for the nine years of data and to make predictions for the year 2000.

3.   Calculate and record these error measurements for your models: Average Error, Mean Absolute Error, Average Percentage Error, Mean Absolute Percentage Error, and Mean Squared Error. Discuss the benefits and drawbacks of these error measurements.

Assignment 3 Discussion Questions
Smoothing Models

1.   Develop a simple exponential smoothing model for this time series, searching to find a good value for the smoothing constant. What does the size of your smoothing constant indicate about the relative emphasis on recent data?

2.   Discuss the limitations of using simple exponential smoothing for this time series.

3.   If you have access to software, apply other smoothing methods to the time series. Use the resulting models to fit values to the data, and create forecasts for the year 2000. Record the performance statistics for all your models.

Assignment 4 Discussion Questions
ARIMA Models

1.   Develop ARIMA model(s) for this time series. Examine the ACF and PACF plots and perform any operations necessary to obtain a stationary time series. Determine the parameters that you feel are indicated by your results, and use a statistical package such as Minitab to develop your model(s). Record the MAPE, MAD, and MSE measurements. Select the most appropriate ARIMA model for this time series and justify your choice.

Assignment 5 Discussion Questions
Multiple Regression with Indicator Variables

1.   Build a spreadsheet that incorporates twelve new columns. One of these is a column for time, represented by the observation number from 1 to 108. The other eleven are indicator (dummy) variables for January through November. If the observation is taken from that month, then the value is 1. Otherwise, the value is 0.

2.   Using time as one independent variable and eleven indicator (dummy) variables for the month, develop a multiple regression model to predict the electric bill. Justify your model based on its statistical results. Use your model to fit values for the nine-year period and calculate performance measurements. Use your model to predict the bills for 2000.

Assignment 6 Discussion Questions:
Incorporating Additional Information with Multiple Regression

1.   Begin building a spreadsheet that incorporates the time series with the additional information. What effect do you anticipate that each of these potential independent variables will have on the amount of the electric bill?

2.   Information is available for the average temperature recorded each month. Would this information provide a useful explanatory variable for the multiple regression model? If so, add it to the model.

3.   Heating Degree Days and Cooling Degree Days are concepts that relate the temperature to energy usage. For USA Today's description of these calculations, visit http://167.8.29.15/weather/askjack/waskdays.htm. Search for monthly values for these

measurements in the vicinity of Terre Haute, Indiana, and add them to your spreadsheet, or use the values in the dataset.

4.   What other explanatory variables do you feel would be useful for this time series? Discuss what you would expect to happen if you had this information, and specify where you would obtain the information.

5.   Use your spreadsheet to develop an effective multiple regression model to use to predict the electric bill. Justify your model based on its statistical results. Use your model to fit values for the nine-year period and calculate performance measurements. Use your model to predict the bills for the year 2000.

Assignment 7 Discussion Questions
Error Analysis, Holdout Sample, and Model Selection

1.   Combine the statistical results from all of your forecasting models into a table that will let you compare performance measurements. Which model appears to do the best job?

2.   Data from 2000 is shown below. Determine the mean absolute error, the mean absolute percentage error, and the mean squared error for each of your models during the year 2000.

3.   Based on the results above, and considering the practicality of each of the methods you employed, which forecasting technique would you recommend for this time series? Support your answer.

Spreadsheet Assignment 1 (using a subset of the data)

The family has a 13-year-old heat pump (type of energy efficient furnace and air conditioner) that is nearing the end of its useful life, estimated at 8-15 years. Older units require more expensive repairs and are less energy-efficient than newer models. In fact, the new heat pump is estimated to use 30% less energy than the old one.

Assume, on average, that $55 (before tax dollars) of the monthly bill are not due to heating or cooling requirements (and thus not affected by the heat pump replacement). For this assignment, assume that electricity costs a flat rate of 6 cents per kilowatt hour

Prepare a table that shows the kilowatt hours used and the expected bill for each month of the year. Add 5% for sales tax, and format all amounts appropriately. Projected kilowatt hours used for the next year are shown below:

Month	KWHs
Jan	3104
Feb	3236
Mar	2934
Apr	2234
May	1987

Jun	1255
Jul	1380
Aug	1655
Sep	1555
Oct	2009
Nov	2256
Dec	2680

Your table should also calculate the average monthly bill for the year and the average monthly bill just for the period October – March.

The cost of a new heat pump is $3150 installed. The local electric utility company offers a $300 rebate in a program that is ending on the last day of this month. This rebate lowers the cost of the unit and encourages customers to be more energy efficient, lowering the capacity that the electric utility must provide at peak times.

Further, the electric utility offers a 20% discount on electric bills from October through March; however, the first 1,000 kilowatt hours are not discounted.

Should the family replace the heat pump now to earn the $300 rebate even though it is still working? How long would it take to pay back the cost of the unit in energy savings and discounted electricity rates? Base your calculation on average monthly cost with and without the replacement heat pump.

As a check on your work, you should find that in a representative year, without replacement the family averaged about $170 per month from October through March with a peak of about $204 in March.

Spreadsheet Assignment 2

How are the billing amounts determined? The Rider Total column in the dataset and the table below provide rate information obtained from the utility company.

1.   Use the tiered rate information in the table below and the Rider Total from the dataset to calculate the billing amount for December, 1999, for the following numbers of kilowatt hours used: 100, 500, 750, and 1250. The sales tax rate is 5%.

Tiered Rate Information

For Jan 1991 to August 2000	BASE	13-100 kwh	101-200 kwh	201-600 kwh	601-1000 kwh	> 1000 kwh
	$7.20	$0.07776	$0.04887	$0.04202	$0.03355	$0.02566
Beginning Sept 2000	BASE	0-300 kwh	300-1000 kwh	> 1000 kwh		
	$8.15	$0.081711	$0.04851	$0.04019		

Legend
BASE is fixed amount of bill regardless of consumption
Other $ amounts are the cost per kilowatt-hour used that month

2. Using the IF function, define a formula that will correctly calculate the billing amount for any number of kilowatt hours used. Calculate the billing amount for 100, 200, 600, and 1000 hours based on the December, 1999, rate information. Demonstrate how you could use these amounts to determine in which category the number of kwh falls. Subtract to find the portion of the bill that applies to the hours in the last category. Determine the total number of hours used.

3. Develop an integer programming model that will convert the billing amount to the total number of kilowatt hours used. Test your model for December 1999 using a bill of $59.12.