3-2014

Are We Really Going Green Part Three

Craig G. Harms
University of North Florida

Are We Really Going Green Part Three

Craig G. Harms

The University of North Florida

ABSTRACT

This paper is a second follow-up to "Green, Part One" and "Green Part Two." In Part One the topic of Interrupted Time Series (ITS) is used to develop a multiple regression model to predict the usage of electricity at the author's home. In the second paper a multiple regression model is developed to predict water usage in the home. The third paper is a follow-up paper after some of the data in the second paper was found to be spurious. This third paper not only cleans up the bad data, but looks at the proper usage of ITS versus a simple dummy variable. All three papers reveal our continued effort to save the planet by using less of its natural resources. The case is presented as a tutorial for a complex and real world student case.

INTRODUCTION

"Going Green" is a loyal cry from the conservation crowd. Yet all of us, even us evil capitalists are aware of the need to reduce our consumption of the limited natural resources of dear Earth. Even if we have three Chevy Camaros--all with V-8s--we can find other ways to appease Mother Nature. In the first paper twenty years of monthly electrical usage data was used in a multiple regression model to prove that our family does strive to reduce scarce resources. The model contained 19 independent variables and was extremely difficult for the students to analyze. In Part Two a less intense presentation is used to develop a case for use in a quantitative method classroom. In Part Two the reduction of water usage was studied using eight years of utility bills. A large reduction in electrical usage is expected by the addition of a well. After finding erroneous data in Part Two, the latest eight years of monthly utility bills are again used to show the reduction of water usage. The third paper also looks at the proper usage of ITS. Using updated, timely utility bills, the model attempts to quantify the reduction of water usage by the addition of a well and pump.

SUMMARY OF PART ONE

A Time Series model with monthly seasonal indexes is used to demonstrate the reduction of electrical usage. If the B-coefficient of the Time variable is negative, then electrical consumption is being reduced on a monthly basis. However, since the "Real World" nature of the problem included an addition of 700 square feet to the house, the addition of a baby, and the installation of a more efficient heat pump and A/C unit, the model became much more difficult. At the beginning of the study, the "Are We Going Green" was a resounding "NO!" To show that the three critical events had a great effect on the overall electrical usage, an ITS model is used--with three sets of two variables each. Once the effects of those three critical events are removed, it can be shown that the net overall electrical usage is in fact being reduced.

Students went "nuts." The model was just too complicated. Although they had studied both simple and multiple regression and time series with seasonal adjustment, they did not have seasonal adjustment in a multiple regression model and did not have six more independent variables pertaining to the three critical events. It was also extremely difficult for them to define the meaning of those six interrupted time series variables.

SUMMARY OF PART TWO

A Time Series model with monthly seasonal indexes is used to determine reduction of water usage. If the B-coefficient of the Time variable is negative, then water consumption is being reduced on a monthly basis. Unlike in the first paper where there are three significant events, in this paper there is only one. The critical event, the addition of a well and water pump to draw water from a shallow aquifer rather than the city line to water the grass, plants, and flowers should create a clear reduction of water usage and a "no brainer." However, since the "Real World" nature of the problem included the identification of a broken water meter within one month after the installation of the well and pump, spurious data entered the problem in the second paper.

RESEARCH GOAL OF PART THREE

The research goal of part three is to "clean up" the spurious data in part two by the addition of a simple dummy variable to identify the months when the water meter was not functioning. Again there is just one pair of ITS variables for the installation of the well and the pump. In the context of assigning the case, it is the students' assignment to defend or refute the professor's statement, "I am going green." Again real world glitches (meter breaking) need to be handled. This is definitely not a typical "classroom problem."

VERTICALLY INTEGRATED CASE APPROACH

Table One presents a list of tasks the students and faculty must accomplish during the six class assignment: The students required time to learn the topics and do the quantitative analysis. Class time was not entirely taken up by this case, rather portions of the six classes allowing students time to let the topic "sink in."

Table 1

Steps to Perform a Vertically Integrated Research Case

1) Statement of the model--the dependent variable (class #1)
2) Hypothesize the quantitative model and the relationship of each of the independent variables to the dependent variable (class #1)

*It should be noted that the available independent data is limited to what is recorded over the years. We could not play "Star Trek" and go back in time to collect data.

3) "Collect" the data and build the database. This includes going back to personal files of utility bills over the past eight years.

4) Make database available to students on-line for retrieval and use in an Excel program. (class #2)

5) Analyze the data: model validation & statistical tests.
   - a: time series with trend and seasonality (class #3)
   - b: multiple regression using interrupted time series (class #4)
   - c. addition of the dummy variable for the broken water meter (class #5)

6) Present student model and discuss results. (class #5)

7) Draw conclusions--review hypotheses. (class #6)
   - a. is the model valid?
   - b. when is ITS a proper methodology to use?

DEVELOPING A CAUSAL MODEL

Students are familiar with developing a causal model. They have worked with income, people, inflation, unemployment rates, and other econometric variables in Economics class. This model is a more personal model that encourages students to develop their own causal model with common (human/personal) variables.

As stated before, to not make this second and third case complicated, the students are given the model. In the process "good" independent variables are discussed. It is important to understand that if the presenter wants the reader to "buy into" the model, the independent variables must pass three tests:

1) The proposed independent variable must be logical. We must be able to sit back and say, "Yes, that makes sense to me."

2) The proposed independent variable must be quantifiable. We must be able to develop a number to represent the variable value.

3) The proposed independent variable must be obtainable. Beyond overcoming the proprietary problems in many corporate databases, we must be able to get our hands around the data in a timely manner and without spending an arm and a leg. In this case, there is no propriety, merely availability.

THE DEPENDENT VARIABLE -- UTILITY BILL IN DOLLARS

In my town water usage is charged on the monthly utility bill. Also included in that bill are the recycle charges, the garbage collection fees, and the sewer rates. The sewer charge is directly related to the amount of city water the customer uses. Therefore if the customer has a well, there will be no water charge and no sewer charge to water the lawn. From this point on "water" represents the charges for water and sewer. Water is charged as a fixed cost each month (small) plus a variable cost charge per 1,000 gallons used. There is no charge until 1,000 gallons are used. The customer simply gets less than 1,000 free for one month and then the charge appears on the next monthly bill. Therefore the bill is "lumpy." Although slightly affecting the seasonal indexes, the data does not seem to cause damage to the overall model or the interpretation of its meaning.

DISCUSSION ABOUT INDEPENDENT VARIABLES

There are three independent variables: Time, the Well and Pump, and finally a dummy variable for the months when the water meter did not function.. The latest 108 months (nine years) of data are retrieved, including the months with the broken meter. Using monthly seasonal indexes in a time series model is very easy. However, when the second independent variable is added, the well and pump, then a multiple regression model is required. The additional dummy variable for the broken meter is an easy, yet critical variable. Seasonality in a multiple regression model is a much more difficult process. At first glance merely calculating a trend model with the addition of the interrupted time series variables (one pair) for the one critical event may be sufficient to create a good model. However, in Florida, there is seasonality in water usage and it is deemed essential to forge ahead with the more difficult multiple regression model with seasonal indexes.

The first two models will be time series models. First, a simple time series model using only a trend component, hopefully with a negative B-coefficient and a second time series model that includes trend and seasonal components. Students have studied time series models in previous courses. The third model will include the variables in Table Two

IMPORTANCE OF USING INTERRUPTED TIME SERIES

Using a time series, even with the addition of seasonal indexes, most likely will include a large negative B-coefficient. A residual analysis might reveal large negative residuals for the observations after the well and pump are installed. Sometimes this might lead to wondering if the data is flawed. Determining a value for the effects of the well--the B-intercept coefficient of the well installation--could be interpreted as the dollar savings due to the well.

:

Table 2
Independent Variables and Hypothesis

Independent Variable #	Independent Variable Group	Hypothesis(increase/decrease)
1	Time	Decrease (going green)
2-13	Seasonality(months)	Sine Curve
14-15	11/6/2011 Well and Pump Installation	Big Decrease
16	Broken Meter	Big Decrease in recorded usage of water

A FIRST MODEL -- A SIMPLE TIME SERIES WITH TREND

The class is prepared to start the analysis process. The data, all 108 months, ending with August, 2013, is stored in a file for student access. Table Three presents the statistical analysis for the trend model.

Table 3

Trend Model of KWH Usage per Day

B-zero	$94.07612
B-one	$0.235112 / month
Standard Deviation	$31.52945
R-square	0.05219
F-statistic	5.8367
T-statistic (time)	2.4159

This model reveals surprisingly poor results including a weak R-square. The B(1) coefficient is not negative. It is important to define in English the meaning of the various numbers before we move to the next and more involved model. Quite possibly not doing a good job defining numbers contributed to the lack of understanding in "Going Green, Part One."

The B(0) of $94.07 represents a lot of costs: garbage, sewer, recycling, and the fixed component of the water charge. The B(1) represents the increase or decrease in the cost and/or the increase in usage of water per month plus any increases in other costs that are part of the utility bill. Although there have been some cost increases in garbage collection and recycle fees, these have been very small as compared to the increases in the cost of water per thousand gallons. Therefore the positive value of $0.23 per month, which is in itself a very small number, reflects two things: 1) the increase in the usage of water, and 2) the increase in the cost of that water. It is virtually impossible to use actual water used as the dependent variable because of the lumpy billing and small numbers of the 1000s of gallons of water used. Students need to learn that sometimes the "perfect" variable is not easily obtained and compromises must be made if you want any model.

ACCOUNTING FOR SEASONALITY

Seasonality in a time series model is straight forward. Seasonality in a multiple regression is far more complicated and takes several steps. The steps and results are described below. The results for both the Time Series model with seasonal indexes and the Multiple Regression model with seasonal indexes are presented in Table Four.

Step 1: Develop a matrix of 11 dummy variables--September through July--for this case. There is no August dummy variable. Thus when the computer calculates the B-coefficients for each of the eleven months, that figure is the difference between the particular month and August. For example, the July index from the computer program is -12.824. This means that the use-of-water cost coefficient in July is 12.824 (per month) less than in August. Thus to calculate the additive seasonal indexes so that we can talk about the usage in July relative to July, more steps are required.

Step 2: Add up the values of the eleven B-coefficients, September through July. For this case, that summation is -85.535315. The 12th month, August, is given a seasonal index value of 0.0. Divide this summation by the 12 months. This average is -7.1279429.

Step 3: Subtract -7.1279429 from each of the twelve seasonal indexes (11 from the computer output plus August) and add -7.1279429 to the B-zero value from the computer (102.0044). to become 94.87646. The August index becomes +7.1279429.

Table Four presents the numerical results of the above three steps and the statistical measures from the computer output.

The two models are very similar. The difference must be in the actual calculation of the equations. The measures of performance show improvement in the Seasonal model versus the Trend model. However we have not looked at the key independent variables in this study: the well and water pump and the broken water meter. Comparing the statistics between the Trend model and the Seasonal model, the B-zero and B-one coefficients barely changed. That is expected. The R-square for the seasonal model has improved four fold. However for a model using aggregate data, the measures are still very weak. That should easily justify the need for inclusion of the ITS variables and finally the inclusion of the dummy variable for the broken meter.

Table 4

A Second Model that Considers Trend and Seasonality

Measure or Variable	Trend & Seasonal Model Multiple Regression Model	Trend & Seasonal Time Series Model
B-zero	$94.87646	$94.07612
B-one	$0.220426 /month	$0.235112 /month
Standard Deviation	30.38248	28.76653
R-square	0.211226	0.211025
adjusted R-square	0.111591	
F-statistic	2.120	
Sum of Squares Regression	23483.57	23461.21
Sum of Squares Error	87694.03	87716.39
Sum of Squares Total	111177.6	111177.6
	Seasonal Index	Seasonal Index

September	-4.82405	-4.74327
October	+4.52442	+4.590504
November	+1.15733	+1.208725
December	-3.98645	-3.94972
January	-9.46906	-9.44705
February	-17.28176	-17.2744
March	-18.07886	-18.0862
April	+1.09964	+1.07761
May	+28.82477	+28.78805
June	+16.60213	+16.55072
July	-5.69306	-5.76217
August	+7.12794	+7.047162

The numeric values of the seasonal indexes seem to "bounce around a lot." This could be because of the lumpy billing. It could just be because of the rain cycles in Florida. At least the Winter months show the lowest amount of watering. This is very reasonable. The next step to account for the critical event--installation of the well and the water pump presented in Table Two.

USING INTERRUPTED TIME SERIES -- GENERAL MODEL STATEMENT

Each critical event requires the addition of two independent variables. The basic model is developed from the model used in Coleman and Wiggins.

Dependent Variable = B(0) + [B(1) * TREND] + [B(2, 3, ...13) * DUMMY]
+ [B(14) * SHIFT] + [(B15) TRCHGE] (1)

Where: Dep. Var is the dollars of the utility bill

TREND is for the time series variable.

DUMMY is for the matrix of 11 dummy variables for the months or the seasonality + August calculated from output regression information. (a total of 12 variables)

SHIFT is the change in usage due to the critical event. This could be thought of as the immediate decrease in the city water usage, or another B(0) for the model. SHIFT = 0 for all pre-event months and event month. SHIFT = 1 for all post-event months.

TRCHGE is the change in the trend component due to the effects of the critical event. TRCHGE = 0 for all pre-event months and the even month;

TRCHGE = 1 for the first post-event month;
TRCHGE = 2 for the second post-event month, etc.

DUMMY/METER =1 for the months when the meter was not working
=0 for months when the meter was working properly

During the next phase of the research, the critical events are tested. The students hope to confirm their hypotheses about the critical event as presented in Table Two. Table Five presents two model results. The first model (left) is with the inclusion of the ITS variable pair. the second model (right) contains a dummy variable for the months the water meter was not functioning properly--zero water usage was recorded.

The measures of performance are much improved. Let’s look at each model separately. The model that includes the ITS pair of variables for the installation of the well (shift) and the change in the monthly water usage (trend change) showed a healthy improvement in the adjusted R-square (0.385) versus the adjusted R-square without accounting for the well installation (0.112). The B(0)[shift] of -$67.59 states that once the well was installed, the decrease in the water/sewer charge was about $67.59 less per month--a very healthy savings.

Table 5

Taking Into Account the Installation of the Well

Measure or Variable	Trend & Seasonality Multiple Regression Model Well and Pump Installation	Trend & Seasonality MultipleRegression Model +Dummy for Broken Meter
B-zero	$79.987502	$80.439848
B-one	$0.689311 /month	$0.677592 /month
Standard Deviation	25.26193	22.22019
R-square	0.466175	0.591431
Adj R-square	0.385815	0.524816
F-statistic	5.801035	8.878
Sum of Squares Regression	51828.25	65753.83
Sum of Squares Error	59349.35	45423.77
Sum of Squares Total	111177.6	111177.6
B(0) [shift]	-67.5943	+0.416026
B(1) [change]	+1.158599	-2.27458
	Broken Meter	-68.3367
Month	Seasonal Index	Seasonal Index

September	-6.687282	-2.71454
October	+2.06360	+6.42954
November	-1.90111	-4.73494
December	-0.26074	-9.87664
January	-6.46978	-15.31104
February	-15.00878	-15.48234
March	-16.53218	-16.31138
April	+1.91993	+2.98558
May	+28.91871	+30.76901
June	+15.96971	+18.594655
July	-7.05488	--3.65524
August	+5.04282	+9.21706

Regrettably, some of that supposed savings was due to the fact that the water meter stopped working and thus there was little or no water usage recorded. As soon as the city realized this problem, they installed a new meter. At the time, I did not know that, first, the meter was broken, and second when they installed the new meter. Only after the bills greatly increased did I review the monthly bills and find the spurious billings.

Once that is discovered, a dummy variable is added for the months when the meter was not working. That model (right side, above) shows a large improvement in the measures of performance (adjusted R-square 0.385 versus 0.525). Therefore taking into account the months when the meter was not working, we would want to use this model to predict the next month's utility bill.

However, before using the model in Table Five to predict the utility bill, it is important to look at the actual Beta-coefficients of the models. The B(0) and B(1) of the basic model are similar and make sense. The seasonal indexes are somewhat different, but overall they make sense--more watering in the Spring and Summer and less in the Winter. There is a huge change in the "Shift" variable--the variable that indicates how much lower the bill is because of the installation of the well and pump. With the addition of the dummy variable for the broken meter, the shift variable changed from -$67.59 (a savings of $67.59 per month) to +$0.41. That makes no sense at all!

The students did not understand. I was puzzled. We decided to go back to the definition of an interrupted time series variable pair and when it is proper to use them and when it is better to use a single dummy variable. ITS deals with critical events that change the setting of the model. After some bantering back and forth, we decided to change the definition for the proper use of an ITS pair to the following: Using ITS is proper when a critical event occurs and <u>that critical event changes how you live your life forever</u>.

We reviewed the variables that have been used in all three papers. Did the critical event change your life forever or did it just occur and then you went on? The addition of 700 square feet to the house was the first critical event in the first paper. Did that change your life forever? Yes, those 700 extra square feet were part of the master bedroom, the kid's room, and the office. The second variable was the birth of the child. Did that change our lives forever? Oh, yes! The usage of electricity increased greatly. The third critical event was the installation of a more efficient A/C/Heat Pump unit. Did that change the way we lived our lives? No, we continued to use the house as we had before, but it did not change anything about living our lives. A logical conclusion is that the first two critical events were properly handled with ITS pairs. However, the third critical event would more properly be represented in the model with just a single dummy variable--it did not change how we lived our lives.

Let's look at the critical event in this third paper. When we installed a well and pump, did we change how we lived our lives forever? No, we still watered the grass and the flowers using the well water, but did not change when we watered or how long the water stayed on. Therefore, ITS may not have been properly used in this model. Rather, a simple dummy variable should represent all the months when we used the well and pump, just as a dummy variable was included to show the months when the meter was not working. Hence we have one final model to develop as shown in Table Six and the results shown in Table Seven.

Table 6

Independent Variables and Hypotheses

Independent Variable #	Independent Variable Group	Hypothesis(increase/decrease)
1	Time	Decrease (going green)
2-13	Seasonality(months)	Sine Curve
14*	11/6/2011 Well and Pump Installation	Big Decrease
15	Broken Meter	Big Decrease in recorded usage of water

*dropping the "change" variable

Table 7

Removing the ITS Variable Pair

Measure or Variable	Trend & Seasonality Multiple Regression Model Well and Pump Installation no ITS	Trend & Seasonality MultipleRegression Model +Dummy for Broken Meter ITS model
B-zero	$81.05054	$80.439848
B-one	$0.664078 /month	$0.677592 /month
Standard Deviation	22.66374	22.22019
R-square	0.570336	0.591431
Adj R-square	0.505655	0.524816
F-statistic	8.817729	8.878
Sum of Squares Regression	63408.59	65753.83
Sum of Squares Error	47769.01	45423.77

Sum of Squares Total	111177.6	111177.6
B(0) [shift]	-31.402	+0.416026
B(1) [change]	removed	-2.27458
Broken Meter	-50.9657	-68.3367

Month	Seasonal Index	Seasonal Index
September	-3.58506	-2.71454
October	+5.31973	+6.42954
November	-4.15386	-4.73494
December	-6.29757	-9.87664
January	-12.17846	-15.31104
February	-14.77196	-15.48234
March	-16.01266	-16.31138
April	+2.72215	+2.98558
May	+30.00363	+30.76901
June	+17.33733	+18.594655
July	-5.40456	--3.65524
August	+6.97584	+9.21706

Looking at the B-coefficients of the model with two dummy variables is very pleasing. They can be explained as follows: The addition of the well and pump will save us about $31.40 per month forever. We saved approximately $50.96 per month for each month that the water meter did not work properly. Although the bottom line measures of performance for the model with two dummy variables are slightly worse than the model with the one ITS pair, the explanation of the B-coefficients is vastly more logical!

USING THE MODEL

The final step is to use the final model with the two dummy variables to predict the utility bill for the next month, September 2013 using all three models.

Well Model:

$$= \{[81.05054 + 0.664078 * 109] - 3.58506\} + (-31.402 * 1) + (-50.9657 * 0) \quad (2)$$

$$\$122.03 = \{153.435\} - 31.402 - 0 \quad (3)$$

This forecast for the monthly utility bill is not as low as wished, but spending the dollars to install the well and the pump will be repaid in four years.

CONCLUSION

The time had finally come to wrap it all up. As a class we had come a long way. The goal of learning was reinforced by the process of discovery. When everyone is involved in building the model, meaningful learning becomes a more pleasant experience. The interaction of the students, the professor, and the database itself made the project interesting and engaging. We are not spoon fed a sterilized, meaningless, laboratory tested batch of data from a textbook.

Somehow, we all are invested in the project and the enthusiasm for the subject infected the entire class. The class performed the analysis and learned that *real* data does not always come easy or have simple conclusions. The hypotheses are not always validated which is very frustrating. Finally, even though it is challenging material, the class managed to have fun with it! Here are a few reasons why this case is such an effective method of learning:

1) The data is real and timely.
2) The situation is realistic and not just a "classroom exercise."
3) Students are encouraged and expected to interact throughout the case.
4) The computer is used extensively
5) Sophisticated models are developed using the computer.
6) Many steps are needed to reach a conclusion.
7) There is not one clean, final answer, thus reinforcing the "real" idea of the case.
8) The students enjoyed the realistic and far-reaching discussions.
9) When it makes sense, it sinks in!

REFERENCES

Andrews, Robert L., "A Data Collection and Analysis Project for a Statistics or Quality Management Course," Proceedings of the Southeast Decision Science Institute, 1995, page 127.

Coleman, B. Jay and Wiggins, C. Donald, "Using Interrupted Time Series Analysis to Access an Allegedly Damaging Event, " Litigation Economics Digest, 1986.

Davis, Arthur and Harms, Craig, "Are We Really Going Green? The Development of a Student Case Using Multiple Regression. Proceedings of the Southeast Decision Science Institute, 2009.

Harms, Craig, "Using Interrupted Time Series Analysis and Multiple Regression to Predict Daily Prices of a Closed End Mutual Fund," Proceedings of Southeast INFORMS, October, 2002.

Harms, Craig, "Are We Really Going Green, Part Two," Proceedings of Southeast Decision Sciences Institute, 2013.

McDowall, D. R., McCleary, E. E., and Hay, R. A. Jr., Interrupted Time Series Analysis, Sage University Paper Series; Quantitative Applications in Research Methods, Beverly Hills, Sage Publications, 1986.

Neter, J, Wasserman, W. , and Kutner, M. H., Applied Linear Statistical Models, 3rd edition, Richard D. Irwin, Burr-Ridge Illinois, 1990.

ABOUT THE AUTHOR

Craig Harms received has Ph.D. from The Ohio State University in 1984. He has taught at Miami (Ohio) University from 1976-1980 and is currently at The University of North Florida in Jacksonville, Florida (1980-present). Currently using the 6th edition of his major book, *The Swift Shoe Company*, Dr. Harms is working on his 7th and last edition. The book has over 100 adoptions around the world including Notre Dame University and The University of Capetown, South Africa. His latest book, *Technical Analysis in the Stock Market*, 2011, is used in the quantitative methods course and is sold in the open market as a self help to conservative stock trading.