Hands on Workshop

Do you know when your data is lying to you? The HOW of Regression Analysis with Quantitative and Qualitative Variables

Steven C. Myers, Ph.D.  
Department of Economics, College of Business Administration  
The University of Akron

# Abstract

A qualitative indicator such as a binary variable, D, may describe a population difference, such as ‘male’ and ‘female,’ or ‘before’ and ‘after’ some event. This workshop will walk you through the regression analysis of whether an outcome variable, Y, is influenced by a qualitative binary event, D. Using only 14 years of data on Y, you will learn that what seems very simple actually takes 8 different regressions and many Wald statistical tests to reveal that the best conclusion is a complexity of model specification and statistical inference. The take away is no matter how sophisticated the technique and how good the data, there is no substitute for thinking your way through a problem. Blindly following technique alone is a bad practice that leads you to make huge mistakes.

You will learn the value of articulating a problem, preparing data, exploring the data and the importance of the data generating process. You will experience interpreting the results and inferring the validity of the results and drawing a conclusion. You will learn a rather comprehensive set of techniques in a very simple example. Most importantly, you will learn that the roadmap for similar problem solutions is not guided by the techniques as much as the critical application of human thought.

# Business Problem

An entity, such as a business or government, has a metric of great interest to their operation. This success metric, denoted simply as Y, is tracked and recorded once during each period of time. Y could be output produced, number of people served, number of items sold, state domestic product, US gross domestic product, or any other success metric reasonably related to the entity’s operation. This is the environment of our business problem.

The entity has implemented a policy change 7 periods ago and now wonders whether it “paid off” in higher values of the success metrics. To test this “pay off” we have 14 periods of data, seven years before and seven years after the policy change. These time periods, for example, can be months, quarters or years.

# Articulation of the problem

The first step in any applied analytic approach is to articulate the problem needing to be solved. Only by starting out with a clear statement of the problem can you hope to do all that is possible to solve the problem. The problem to be solved is this: Did a policy or procedure implemented during the seventh period cause a change in the outcome metric of interest? From this articulation various hypothesis will emerge to statistically test the problem.

## Data Acquisition

A data series of 14 periods of data are collected on a metric of interest. The measure of this variable is shown below and is called Y. We have 7 periods of data before and 7 periods of data after the change in the policy or procedure. A means procedure is shown to verify that you typed the data in correctly.

Data Y;

input Y @@;

datalines;

12.35 13.71 16.00 17.94 20.76 21.11 24.63

27.56 32.88 35.16 39.26 44.28 47.27 51.55;

run;

proc means data=y mean std maxdec=4;

title 'The correct mean of y is 28.890 and the Std Dev is 12.9719'; run; title;

# Preparing the data for Analysis

## Data cleaning

Preparation of the data should involve much work exploring and cleaning the data, but in this illustration we accept the data as high quality and accurate. In practice one should never skip this step. Doing so will lead to peril.

## Data Transformations for Analysis

Our business problem suggests a course of action since the hypothesis The first question to explore is what is the pattern of the data and how does it trend over time? Are there any obvious characteristics to that data and its trend? Are their significant perturbations in the data or is it fairly smooth? These and other questions will be addressed below, but it become obvious that we need to create some variables first as shown in the next code block.

Data work.trdata;

/\* Problem is to explain the trend in variable Y.\*/

/\* H0: An intervention that begins in T=8 has no effect on the trend line.\*/

/\* H1: An intervention at T=8 changes the trend line.\*/

/\* Alternative problem: \*/

/\* The actual equation is simply nonlinear in variables such as y = T TSQ.\*/

set Y;

T=\_N\_; /\* 1. create time variable. \*/

TSQ = T\*T; /\* 2. and time-squared value. \*/

D=0; if T>=8 then D=1; /\* 3. Create binary variable for the intervention. \*/

DT = D\*T; /\* 4. create interaction of D and T. \*/

run;

Four variables are created to aid the analysis, T which is the linear time measure, TSQ which will allow for a quadratic bend, D which is the intervention or treatment variable and DT which allows that the influence of time, T, may vary with D.

## Model selection and specification

Second is we must resolve how we will solve the problem. Some analytic approaches start with the idea that the truth is in the data and what is revealed is true, but the premise of this paper is that data does lie to you. A common problem is that like any tortured prisoner is if you beat the data long enough it will confess. Of course the tormentor will stop the beating when their biases are confirmed. And if all you want to do is to confirm your expected solution, then why do the data work at all?

# We will select both quantative and qualitative variables for this exercise

Y and T and TSQ are quantitative and

D is qualitative

DT and TSQ are interaction variables,

# Next section

Linear line throughout – no obvious change by D

Nonparametric fitting of the trend – oops it is quadratic

Examine linear line by time-slice

Will the quadratic shown in the time slices or not. Hit its linear.

# A visual Exploratory approach using PROC SGPLOT

Regression is a statistical procedure that can be a highly visual one in the few parameter case. This is especially true in this paper since we are using a small dataset and few explanatory variables. It is a small data set with much meaning. We suspect that an intervention may have had an effect and are interested in examining whether that effect has meaning or is just an artifact of the particular sample. In part one, we will take a visual approach and in part two we will apply a statistical inference approach to solve our problem whether the intervention has a statistical effect.

Make sure you include this code before continuing:

ods graphics on / noborder width=5in;

%let xref = %str(xaxis values=(1 to 14 by 1); refline 7.5 /

axis=x label="<-- Policy change" labelloc=inside labelpos=min ;);

It makes sure you graphics are on and xref is defined as a common setting for all X axis in the following graphs.

So what does the trend look like? Is it linear before and after the intervention? Figure 1 suggests that the regression line is possibly linear, but that at least five residuals are large enough to be outside the 95% confidence interval. The points from the scatter also look like they may be quadratic.

title1 'Model 1: Y follows a Linear Trend.';

title2 'PROC SGPLOT with REG Statement.';

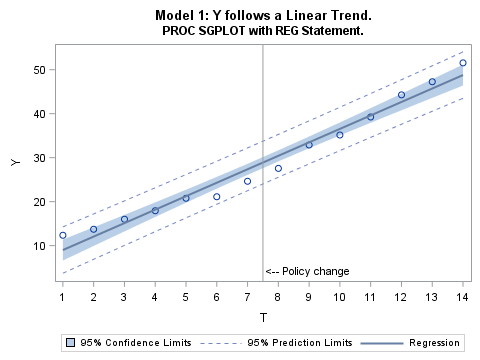
PROC SGPLOT data=trdata ;

reg x=T y=Y / CLM CLI ;

&xref;

run;

Figure : Explore the trend of the metric Y



The pattern of residuals (vertical differences between the actual observations and the line plotted in Figure 1) suggest a “U” or “V” pattern as at low and high values of T the residuals are positive while in the middle of the series, the residuals are negative.

Change the last code to add degree=2 as an option to the reg command. The default is degree=1 and plots the regression of Y on X, while a change to degree=2 plots Y as a quadratic in T. The code is:

title1 'Model 2: Y follows a Quadratic Trend.';

title2 'PROC SGPLOT with REG Statement.';

PROC SGPLOT data=trdata ;

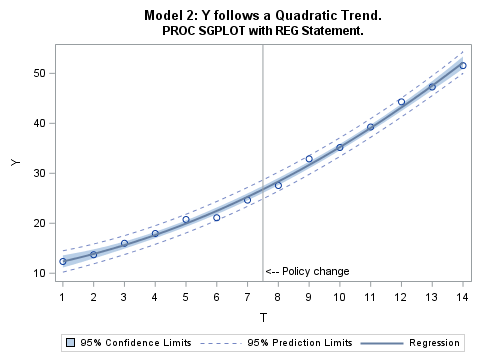
reg x=T y=Y / degree=2 CLM CLI ;

&xref;

run;

The result of this change is shown in Figure 2 and rather clearly suggests that the trend in Y may be quadratic and have little to do with the intervention, D.

Figure : Y as a quadratic function of T



What if we do not impose either a linear or quadratic form on the trend in Y, but rather use a non-parametric technique invoked by the plot command loess to trace out the most obvious pattern. Loess calculates a large number of regressions by using a neighborhood sample of the points at each observation to calculate a weighted linear or cubic regression of degree 1 or 2 on the neighborhood where the weights are greatest for the nearest neighbors.[[1]](#footnote-1)

To examine the exact path we again use PROC SGPLOT to trace out the points, but this time with a LOESS statement. The LOESS is a non-parametric regression called local regression that will trace out the scatter points by running a regression only among the nearest neighbors to each point. That way the many regressions run are not sensitive to points far away from the point of interest. In the loess command we can choose cubic or linear and each with a degree of 1 or two. You can run all 4 combinations and see that in this case the lines look the same for each combination.

We leave the linear reg plot (changing the code from last time to degree=1) and cause it to be somewhat transparent. The loess plot is listed next and is drawn over the reg plot. There are many commands that will deal with how the plots look. The code for this investigation is:

title1 'Nonparametric Local Regression LOESS Model.';

title2 'Tracing out the points with LOESS and comparing to the Linear Trend.';

PROC SGPLOT data=trData;

reg x=T y=Y / degree=1 CLM CLI CLMTRANSPARENCY=.5;

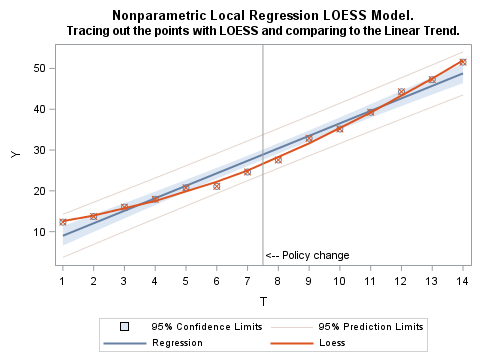
loess x=T y=Y /interpolation=linear degree=2;

&xref;

run;

We can see in Figure 3, the loess plot seems to trace out a fairly definite quadratic-looking relationship and in this case the scatter seems to show points closer to the quadratic than the linear. This is the first lie in this data, but we have to finish the analysis to know that!

Figure : What does a non-parametric trend reveal?



Since our interest is before and after the midpoint of the data, that is did the intervention, D, effect the trend, we turn again to linear regression and apply it separately by group, before (D=0) and after (D=1). For this visual we modify the code to overlay the before and after regressions on top of the full-sample linear regression. The group=D option on the second reg plot instructs the plot to plot first the observations with D=0 and then the last observations with D=1. The code now looks like:

title1 'Model 4: Structural Break with Linear Trend by Group=D';

title2 'Separate linear regressions before and after policy change';

PROC SGPLOT data=trdata;

reg x=T y=Y / CLM CLI CLMTRANSPARENCY=.5;

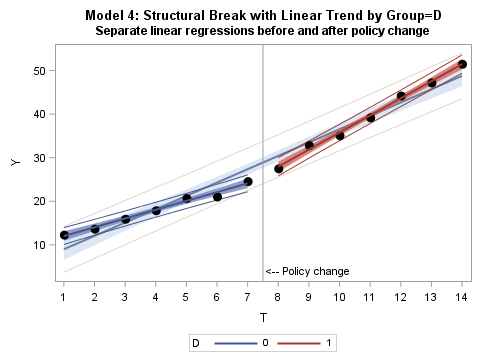
reg x=T y=Y / CLM CLI CLMTRANSPARENCY=.25 group=D

markerattrs=(symbol=circlefilled color=black size=10px);

&xref;

run;

Figure : What does the regression look like before and after the intervention?



It seems obvious from Figure 4 that there is a different trend line before (blue line) and after (red line). This suggests that there is strong evidence that the trend is consistent with a structural break and that D has a strong influence.

The local regression in Figure 3 traced out a strong looking quadratic relationship and that was said to be the first lie in the data. To see why we can visualize the local regression of the last 7 periods separately from the first 7 periods to get a feel if the separate periods (groups) are more linear such as in Figure 4 or more likely due to a natural quadratic-like relationship as in Figure 2. The following code will overlay the full-sample linear trend with the local regressions in the before and after time-slice of the full sample:

title1 'Local regression by group=D';

title2 'Separate LOESS regressions before and after policy change';

PROC SGPLOT data=trData;

reg x=T y=Y / CLM CLI CLMTRANSPARENCY=.5;

loess x=T y=Y / group=D interpolation=linear degree=1

markerattrs=(symbol=circlefilled color=black size=10px)

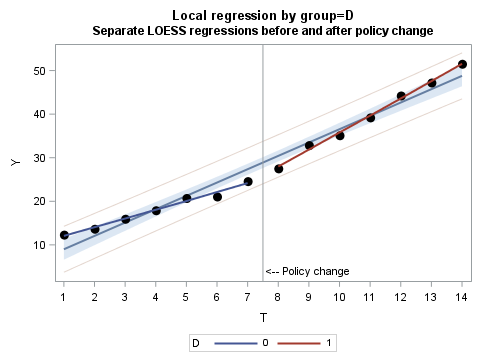
CLMTRANSPARENCY=.25;

&xref;

run;

Our result in Figure 5 is that the quadratic relationship visualized over the entire sample, seemingly vanishes when each group is analyzed separately. But of course we have only beat the data a little, so far, and this confession seems pretty weak. Like much analysis we have dug into the data, but have not yet confirmed our assertions using statistical inference. We turn to that next.

Figure : What do the separate trends by D look like?



To summarize this section, we wondered whether there was a break in the trend of Y due to an intervention indicated by D. Because of large residuals in Figure 1 and the strong looking fit in Figure 2 we could have thought that the trend was natural following a quadratic pattern. However, when regressing separately by group (in Figure 4) we see the possibility of completely separate linear trends before and after the intervention. Figure 5 suggests that the separate trends are actually linear and not just artifacts of an overall quadratic trend since the quadratic pattern fails to visually emerge when separated by group. Our data so far seems to suggest that the intervention did have an effect. The effect illustrated is called a structural break. Now to quantify that.

# Reg

Linear,

linear with D (Hasty, D is insig)

test d=dt over linear

test dt=0 over t d

test d=dt=0

# A statistical approach using PROC REG

## Hasty Regression

I use the term hasty regression to describe when a researcher quickly runs a regression with minimal or no articulation of the problem, virtually no data cleaning and preparation, a lack of understanding of the data generating process (DGP) and no critical though on the appropriate model specification. Who would do this you might ask? Nearly all of us.

Hasty regression is running a model pulled out of the air without regard to the theoretical problem under study and without rigor in the model specification stage, all the while giving no regard to whether the data is clean and ready for analysis.[[2]](#footnote-2)

Mistakes, both of commission and omission, made in the articulation of the problem and the specification and selection of the model may prove fatal to your analysis, but do not announce themselves as such. Indeed, the poorly conceived problem and the misspecified model can lead us to what can only be described as a failure.

More on this in a bit.

Researchers use dummy variables to ascertain whether a point or many such points are deviations from the overall trend in a linear regression model. Our dummy variable is D and defined as D=0 for before the intervention and D=1 after the intervention. And we ask our first statistical inference question: Is D significant in a linear regression of Y on T? To test that end we run two models (but only need the second).

ods graphics on;

proc reg data=work.trdata;

model\_1: model y = t;

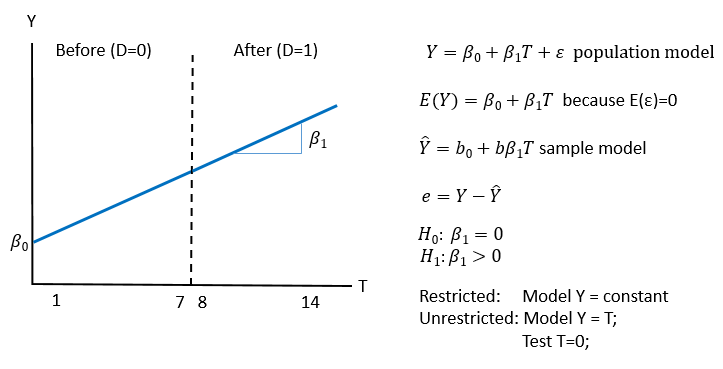
model\_2: model y = t d;

title2 'Full Sample, T=1,..., 14';

run;

Model 1 is illustrated in the following figure.

Figure : The linear model without regard to the intervention



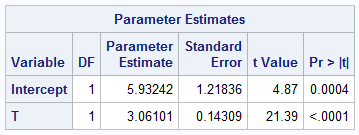
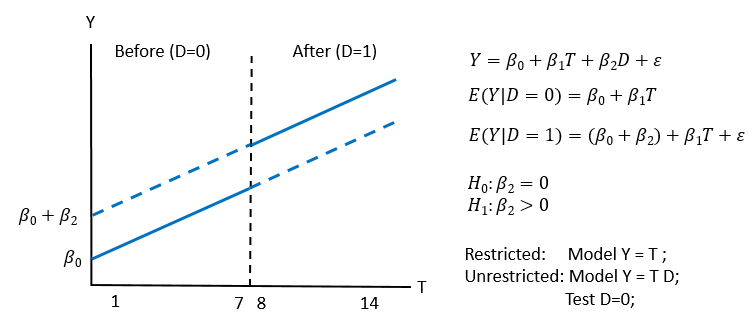
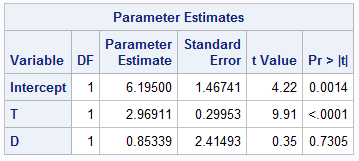
The following table shows the parameters estimated for Model\_1. Using labels for models and tests in your PROC REG specifications helps avoid confusion when sifting through the copious output. Model\_1 has an adjusted R-square of 0.9723 and a RMSE (root mean Squared Error) of 0.9744 showing that this is a pretty good representation of the trend of Y. One might be tempted to say that with such low p-values that this ‘proves’ that the regression is linear, but this proves nothing. The t-values and the F statistic for the model show that the variable T is statistically significance on a null hypothesis of a model where Y = random error. This is hardly confirmation where the ‘bar’ is T is literally better than nothing. If you were to think that these results prove anything, then that would be the second lie in this data.

Table : Model 1 Linear results

The next model can tell us something about D. The null hypothesis here is the previous model Y = T and the alternative is that Y = T D can explain significantly more of the variation in Y. This can be illustrated as

Figure : Test of an intercept difference holding slopes the same (Model 2)



So we run the model and find the following results with an adjusted R-square of 0.9701 (worse than the previous model) and a RMSE of 0.9747 (worse than the previous model). Further the parameter estimate of D is found to be not statistically different from zero in this model. That is, the value of 0.85339 cannot be differentiated from zero.

**So at this point many researchers may conclude that D has no effect on Y and stop their questioning. This is hasty regression at its finest (read sarcasm).**

Table : Model 3 linear and dummy variable.

By saying that D is unimportant in explaining Y as a general statement then this is the next lie from our data. For as we will show, D is extremely important.

# Lack of proof

The statistician’s dilemma is he or she can never know truth. What can be known is an estimate based on an estimator (formula) and a specific sample data set. The statistician wants to test a hypothesis on an unknown population estimator, by using a known sample estimate. Type I errors occur when we reject the null hypothesis but in reality the hull hypothesis is false while type II errors have us failing to reject the null hypothesis when indeed the null hypothesis is true.

We can never prove, we can only falsify on the bases of this estimator and this dataset. The best we can do is use the most powerful test based on our best estimators. In a controlled experiment we can replicate the experiment and test over and over, but in economics and in the testing of events that have historically occurred, there is often no replication possible, the history we have is all we have.

The statistical community has called into disuse the declaration of models being significant or not significant based on a crossed threshold by our test. The tables below do show a breakdown by threshold, but in this case this is for convenience and the exact p-values in this example are universally extremely small (very close to 0.00) or quite large (over 0.30).

# Dealing with indecision

Evidence! I need more evidence! I am a fan of mystery and thrillers and enjoy the twists and turns of a problem where typically the protagonist is up to their ears in a mystery and every step frustrates more, because it seems that they will never get to the end. But of course they do and we will as well. In the words of Fenton Hardy, the fictional father of the Hardy Boys, a series I read religiously a half century ago, we must “leave no stone unturned.”

So what to do? Obviously time to overturn some stones. That is to dig deeper, to go beyond the obvious as in the hasty regression above.

Our grand hypothesis is about D, or rather about what D is measuring. Did the intervention have an impact on the outcome variable as measured by the Y variable? Two points: First it is easy to think we are seeking truth from data, but the manner as I write this paper is that data will lie to you. We need to first think of and seek an overall explanation of why D would affect Y and then measure every effect we can think of. Falling short of that means the data may still lead us astray. Second, tests of statistical hypothesis have three parts, (1) the null hypothesis which we try to reject, (2) the alternative hypothesis and (3) the maintained hypothesis that is not subject to test. The latter is indicative that what other variables are in a model and equally important what is not in a model affects the restricted/unrestricted tests. Consider these models each with the same test. While the null and alternative tests are the same, the maintained hypotheses are different, meaning we may find a failure to reject Test A and an ability to reject Test B.

Model A: Model Y = T D;

Test A: Test D=0;

Model B: Model Y = T D DT;

Test B: Test D=0;

Analysts need not only to develop a modeling and estimation strategy, but a must have a testing strategy to answer the overall question. In this simple case I argue that it takes 9 separate tests to answer the one grand hypothesis. No single equation answers the question *until we have rejected all other competitors*.

# Statistical significance AND economic significance

Two questions to further contemplate:

(1) have you found statistical significance or sufficient evidence to draw an empirical conclusion.

(2) Do the findings make economic (or theoretical or financial, or …) sense.

Something can be statistically significant, but have little or no economic significance. Something can be economically significant without being statistically significant.

# Regression Modeling

Figure : Test of changing intercept and changing slope by the intervention, D.

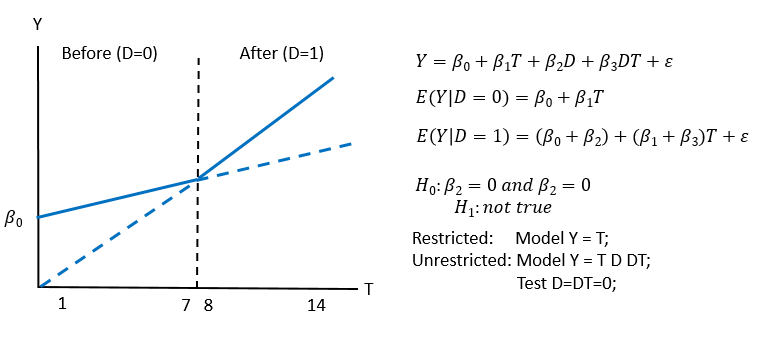
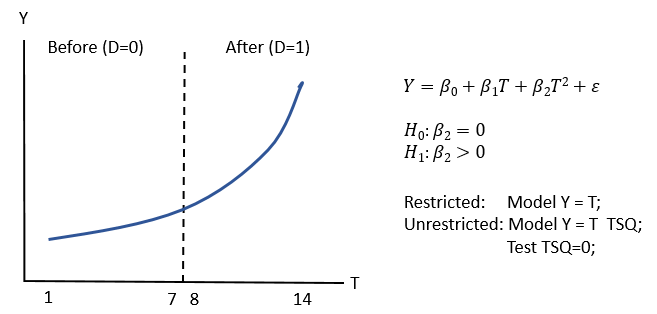


Figure : Alternate model, test of a quadratic form



Title1 'Statistical models';

proc reg data=work.trdata;

model\_1: model Y = T ;

model\_2: model Y = T TSQ ;

model\_3: model Y = T D ;

model\_4: model Y = T D DT ;

title2 'Full Sample, T=1,..., 14';

run;

proc reg data=work.trdata;

model\_5: model Y = T ;

model\_6: model Y = T TSQ ;

title2 'Partial Sample, T=1,...,7 Before';

where D=0;;

run;

proc reg data=work.trdata;

model\_7: model Y = T ;

model\_8: model Y = T TSQ ;

title2 'Partial Sample, T=8,..., 14 After';

where D=1;

run;

quit;

Every adjusted R-square term is very high and close to 1.

Table 3: Summary of all eight statistical models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Sample | Model name | What we learn by each model |  |
| 1 | Full n=14 | Linear | Y is trending upward linearly. |  |
| 2 | Full n=14 | Quadratic | Y is better described as trending upward quadratically. |  |
| 3 | Full n=14 | Linear intervention | Based on the linear model, D has no apparent effect. |  |
| 4 | Full n=14 | Fully interactive | Based on the linear model, D (through D and DT) has a large effect. | Model 4 seems to be slighly better than the Model 2. |
| 5 | Before n=7 | Linear for D=0 | Before, Linear model is trending upward at about 2 points a period. |  |
| 6 | Before n=7 | Quadratic for D=0 | The quadratic model fails in the before period. |  |
| 7 | Before n=7 | Linear for D=1 | After, Linear model is trending upward at almost 4 points a period. |  |
| 8 | Before n=7 | Quadratic for D=1 | The quadratic model fails in the after period. | Models 6 and 8 reject model 2. |

Table 4: Full Sample Statistical Models

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Sample defined as years 1 to 14** | | | | | | |  |
|  | **(1)** |  | **(2)** |  | **(3)** |  | **(4)** |  |
| **constant** | 5.93 | \*\*\* | 11.12 | \*\*\* | 6.2 | \*\*\* | 10.01 | \*\*\* |
|  | (4.87) |  | (14.96) |  | (4.22) |  | (18.25) |  |
|  |  |  |  |  |  |  |  |  |
| **T** | 3.06 | \*\*\* | 1.12 | \*\*\* | 2.97 | \*\*\* | 2.01 | \*\*\* |
|  | (21.39) |  | (4.9) |  | (9.91) |  | (16.42) |  |
|  |  |  |  |  |  |  |  |  |
| **TSQ** |  |  | 0.13 | \*\*\* |  |  |  |  |
|  |  |  | (8.77) |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| **D** |  |  |  |  | 0.85 |  | -13.47 | \*\*\* |
|  |  |  |  |  | (0.35) |  | (-9.12) |  |
|  |  |  |  |  |  |  |  |  |
| **DT** |  |  |  |  |  |  | 1.91 | \*\*\* |
|  |  |  |  |  |  |  | (11.01) |  |
|  |  |  |  |  |  |  |  |  |
| **n** | 14 |  | 14 |  | 14 |  | 14 |  |
| **adj R sq** | 0.972 |  | 0.996 |  | 0.970 |  | 0.998 |  |
| **F** | 457.6 |  | 1713.4 |  | 212.2 |  | 1717.1 |  |
| **root MSE** | 2.16 |  | 0.80 |  | 2.24 |  | 0.65 |  |
| **DW** | 0.44 |  | 2.07 |  | 0.46 |  | 3.30 |  |
| note: All regressions estimated with OLS using the SAS REG procedure.  t-stats in parentheses.  \*\*\* significant at the .01 level  \*\* significant at the .05 level  \* significant at the .10 level | | | | | | | | |

Table 5: Partial Sample statistical models: Before and After Regressions

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Sample year 1 to 7** | | | | **Sample year 8 to 14** | | | |
|  | **(5)** |  | **(6)** |  | **(7)** |  | **(8)** |  |
| **constant** | 10.01 | \*\*\* | 10.42 | \*\*\* | -3.45 | \*\* | -3.19 |  |
|  | (17.13) |  | (9.87) |  | (-2.42) |  | (-0.03) |  |
|  |  |  |  |  |  |  |  |  |
| **T** | 2.01 | \*\*\* | 1.74 | \*\* | 3.92 | \*\*\* | 3.87 | \*\* |
|  | (19.04) |  | (2.88) |  | (30.76) |  | (2.13) |  |
|  |  |  |  |  |  |  |  |  |
| **TSQ** |  |  | 0.034 |  |  |  | 0.002 |  |
|  |  |  | (0.46) |  |  |  | (0.03) |  |
|  |  |  |  |  |  |  |  |  |
| **n** | 7 |  | 7 |  | 7 |  | 7 |  |
| **adj R sq** | 0.980 |  | 0.976 |  | 0.994 |  | 0.992 |  |
| **F** | 293.5 |  | 123.7 |  | 946.1 |  | 378.5 |  |
| **root MSE** | 0.62 |  | 0.68 |  | 0.68 |  | 0.75 |  |
|  |  |  |  |  |  |  |  |  |
| note: All regressions estimated with OLS using the SAS REG procedure.  t-stats in parentheses.  \*\*\* significant at the .01 level  \*\* significant at the .05 level  \* significant at the .10 level | | | | | | | | |

Table 6: Use of automatic model selection in PROC REG

|  |  |
| --- | --- |
| Regression selection process | winning model |
| Selection=adjRsq | T D DT |
| Selection=Stepwise | T TSQ |
| Selection=Forward | T TSQ D DT |
| Selection=Backward | T D DT |
| Selection=maxR | T TSQ D DT |
| Selection=minR | T TSQ D DT |
| Selection=CP | T D DT |

# Non-nested hypothesis testing: Quadratic versus fully interactive model

In the above visual analysis and in the regression-statistical-inference analysis lead to the conclusion that the model of a structural break and the implication that D has a large influence is better if only slightly better than the quadratic model and the implication that D has no effect.

Let us one look beyond the structural break “win” and ask if the two models are actually statistically different? To do so requires a test called a non-nested hypothesis test, because the models are different and one does not fit inside (or nest within) the other. That is, there is no model we can run (either 2 or 4) that will allow linear restrictions on the parameters of one to reveal the other.

## J-test – the variance encompassing test

Do the residuals from the quadratic model have anything to contribute to the ignorance from the fully interactive model?

No

Do the residuals form the fully interactive model have anything to do to the ignorance of the quadratic model?

Yes

Therefore based on the J-test methodology the fully interactive model is acceptable and the quadratic model is not acceptable.

## F-test – the mean encompassing test

When combining a super model of variable from the fully interactive model and the quadratic model and holding constant the variables in common, is there a winner?

Yes, the variable unique to the quadratic model fails to reject the null hypothesis of zero effect. So the quadratic model is not acceptable. However the unique variables to the fully interactive model do contribute to the explanatory power over and above the quadratic model meaning the fully interactive model is acceptable.

## J-test and F-test together – the complete encompassing test

The J and F tests of non-nested hypothesis both agree that the only model acceptable is the fully interactive model therefore the intervention did have an effect.

Title1 'Non-nested hypothesis - J-test';

Proc reg data=trdata;

model\_2: model Y = T TSQ;

output out=Mquad p=Yquadhat;

run;

Proc reg data=trdata;

model\_4: model Y = T D DT;

output out=Minter p=Yinterhat;

run;

Proc reg data=mquad;

model\_4A: model Y = T D DT Yquadhat;

run;

Proc reg data=Minter;

model\_2A: model Y = T TSQ Yinterhat;

run;

Title1 'Non-nested hypothesis test - super model, F-test';

Proc reg data=trdata;

model\_2A: model Y = T TSQ ;

model\_3A: model Y = T TSQ D ;

model\_4A: model Y = T TSQ D DT ;

quad: test tsq = 0;

interactive: test d =dt=0;

run;

Table 7: Non-nested Hypotheses Testing



# Conclusion

The focus of this paper is to support the hands on workshop on testing whether an intervention leads to a …

# References

# ACKNOWLEDGMENTS

I wish to thank the section chair, Carl Nord, for accepting my paper and for his patience in my completion of it. And my thanks to Jessica Chen, the Academic Chair, for her support as well. I am grateful to Kirk Paul Lafler and Josh Horstman for their encouragement to begin presenting at SAS conferences in general.

# Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Steven C. Myers

[myers@uakron.edu](mailto:myers@uakron.edu)

<https://www.linkedin.com/in/stevencmyers/>

<https://econdatascience.com>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

1. Footnote needed here for more on loess [↑](#footnote-ref-1)
2. The data in this paper are assumed clean and if they were not then a whole another layer of concern emerges. Fortunately, that worry is outside the purpose of this paper. [↑](#footnote-ref-2)