Written Examination in Econometrics (B2)[1]

Spring 2017

2017-03-24 13.00-17.00 Venue: Ekonomikum A153.

Lars Forsberg, Department of Statistics, Uppsala University

Allowed means of assistance:

1. Pen or **pencil** (recommended) and eraser
2. **Calculators**,
   (a) 'programmable' calculator, e.g. calculator with graphing functions is OK.
   (b) Calculators with blue-tooth are not allowed.
   (c) Calculators with access to internet are not allowed.
   (d) Calcuators with which it is possible to send and recieve messages of any kind are not allowed.
3. **Physical (paper) dictionary** (no electronic dictionary allowed).
   (a) Dictionary must contain *no notes* of any kind.
   (b) Each student must have his/her own dictionary. It is not allowed for students to pass a dictionary between them.
4. **Ruler.**
5. Collection of formulae and Statistical Tables named '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams*', that the student brings to the exam location.

[1]Disclaimer: The narrative in this exam is designed to make the exam tasks to become a bit more 'interesting' by vaguely resemble a potential real life situation for a statistician working as a consultant in the private or public sector.

The names, persons, places and events depicted in this exam are purely fictional. All and any potential resemblance or similarities with real persons, places, events or situations is completely coincidental.

No implications whatsoever should be made and no conclusions whatsoever should be drawn regarding any potential political, philosophical or religious views or preferences on the behalf of any person or legal entity involved in the creation of this document.

6. Please note that a collection of critical values for the Student´s t, Normal, Chi-square and F-distributions is given in the Appendix of the '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams*'.
7. Also note that the '*Test template*', that should be used when performing tests, is given in the '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams'.*

That is:

1. NO BOOK (except paper-dictionary) is allowed.
2. NO (student-written) notes are allowed.
3. NO other document than the one 'Collection of Formulae and Statistical Tables for Time Series Exam' is allowed.

Instructions: **Please note the following:**

1. Start with reading through the instructions!
2. Make sure you **follow** the instructions!
3. Start with reading through the exam.
4. You may write your solutions in Swedish or English.
5. Total score is **100** points
   (a) If you want the ECTS grades, please indicate that on the cover page!
   (b) For each task the maximum number of points is given within parenthesis, e.g. (16p in total).
   (c) For each subtask the number of points is given within parenthesis, e.g. (2p)
6. All solutions must be on separate sheets. No solutions on the questionnaire! (If so, they will be disregarded.)
7. Make sure your solutions are: easy to read and easy to understand, that is:
   (a) For each task that you solve, please start with a new sheet: after Task 1, start with a blank sheet for Task 2, etc.

(b) Write the *task number* at the top of each page, in the

.................................**MIDDLE OF THE PAGE!!!**..

Like:

....**TASK 1**...

- if you write it in the upper left corner, the staple will cover it, and there is no for way for the examinator to know if the text of that sheet belongs to the previous sub-task or what it is. The Examinators will not make any 'qualified guesses' of what is being displayed on any given page. It is the responsibility of the student to make sure that every task and sub-task is easily identifiable.

(c) If you continue a sub-task on the next sheet of paper - indicate that at the top of the page - IN THE MIDDLE OF THE PAGE, like, for example:

....'Task 1B (cont.)'..

(d) Please separate each subtask A, B etc with a horizontal line across the sheet
_ _
if they are on the same sheet of paper - that way it will be easy for the examinator to actually see where one subtask ends and next begins.

(e) For examinator readability, it is highly recommended that you use a pencil, (and not a pen), which will allow you to erase and rewrite if you make a mistake. Crossed-over text and corrections using 'tipp-ex' will just cause blurriness and confusion to the examinator.

(f) For examinator readability: Write clearly, that is, letters, mathematical/statistical symbols and numbers should be easy recognizable!! Do not underestimate the correlation between readability and points scored, that is, when readability goes to zero, points scored also goes to zero, no matter your intentions or wheather *you* can read it or not.

(g) Also note that everything that you write will be taken at 'face value'. That is, for example, if you write β_1 the examinator will take that as a β_1 even though you may claim that it is given from the context it should be clear that you meant something else, like β_3. Thus, given this example, writing β_1, and that is not correct in that specific formula or statement, this will lead to subtraction of points, even if you will claim that it is just a typo, and that in another task or subtask, it is clear that you understand the issue.

(h) Please put the sheets in **order**, that is first Task 1, and then Task 2 etc...

8. Please keep the questionaire.

9. Do well!

Task 1

(12 points in total) Consider the following single linear regression

$$Y_i = \beta_1 + \beta_2 X_i + u_i.$$

A) (6p) Do the following:

1. Draw a Figure representing the Population Regression Function (PRF), draw the regression line, mark out what is displayed on the axes.
2. Mark out what distance is represented by β_1.
3. Mark out what distance is represented by β_2.
4. Mark out an arbitrary observation Y_i, given this observation, mark out the *conditional expected value* given the corresponding X_i, that is, mark out exactly where in the Figure this conditional expected value is 'located',
5. Write down a formula for the conditional expected value of Y.
6. Indicate in the Figure what distance that is represented by u_i.

B) (6p) Do the following:

1. In a SEPARATE FIGURE from the one in Sub-task A, draw a Figure representing the corresponding *Sample* Regression Function (SRF) for the model above, draw the sample regression line. Mark out what is displayed on the axes.
2. Mark out what distance is represented by $\widehat{\beta_1}$.
3. Mark out what distance is represented by $\widehat{\beta_2}$.
4. Mark out an arbitrary observation Y_i, and given this observation, mark out the *estimated conditional expected* value given the corresponding X_i, that is, mark out exactly where in the Figure this estimated conditional expected value is 'located'.
5. Write down a formula for the estimated conditional expected value of Y_i given that value of X_i.
6. Indicate in the Figure what distance that is represented by $\widehat{u}_i$.

Task 2

(15 points in total)

Important: All outputs nessecary to solve the sub-tasks in Tasks 2, 3 and 4 are given in an Appendix in the very back of the exam. The information needed is given in a number of Figures. Some of these figures may be relevant, some may not be relevant to solve different subtasks in this exam, it is part of the examination that you are able to know what output to use when.

Now, you are hired as a statistian to examine the effect of tax rates on entrepreneurship in Swedish municipalities. You collect a cross section of data (sampled at one point in time) for the year of 2015 for the 291 municipalities in Sweden and you have the following variables:

- $NewEnt_i$ is the number of new enterprises per 1000 inhabitants in municipality i, from now on, referred to only 'Number of new enterprises', that is, for the remainder of this exam, the words 'per 1000 inhabitants in municipality i' will be excluded when referring to this variable.
- $TaxRate_i$ is the tax rate in municipality i in percentage
- Age_i is the average age of the inhabitants of municipality i
- $Univ_i$ is the percentage of inhabitants in municipality i with a university degree, which is considered as a proxy for level of education in the municipality, will henceforth be referred to as 'level of education'
- $InHab_i$ is the number of inhabitants (in thousands) of municipality i

First, consider the model

$$NewEnt_i = \beta_0 + \beta_1 TaxRate_i + u_i. \tag{1}$$

A) (3p) Interpret β_1 for model (1).

B) (3p) Use relevant output to interpret the *estimated* β_1, that is interpret $\widehat{\beta_1}$ for the estimated model (1).

C) (3p) Interpret the coefficient of determination for this model, use the actual estimate from relevant output to do so.

D) (6p) One Mr Tax S. Ucks claims that taxes are very bad for companies, (let alone *high* taxes), and he claims that a one percentage point increase in the tax rate in a municipality will decrease the number of new enterprises by at least one. Perform a test to test the null hypothesis that β_1 is smaller than or equal to -1 against the alternative it is bigger than -1. Use the significance level 5%. Fully outline the test according to the test template.

Does the test support his claim?

Task 3

(34 points in total) Again, for relevant Eviews output - see Appendix.

Consider the model

$$NewEnt_i = \beta_0 + \beta_1 TaxRate_i + \beta_2 Age_i + \beta_3 Univ_i + u_i. \qquad (2)$$

You present this model at a meeting arranged by a group of important people working in a building called Rosebath in Stockholm, but before you have the chance to comment on it any more than just presenting the theoretical model, a Ms A. Duh Bowl stands up and with great emphasis claims that the most important variable is the number of inhabitants, and the fact that this variable is not included in (2) makes the model useless. (Her opinion is that, in order to have an increase in the number of new enterprises (and thus economic growth), the number of inhabitants should be increased no matter their educational background or age.)

****

A) (6p) Formulate a model that, over and above the variables in model (2) also includes the number of inhabitants. Use relevant Eviews output to perform a t-test (repeat: *t-test*) of the null hypothesis that, given that tax rate, Age and level of education are taken into account, number of inhabitants have no effect on the number of new enterprises, against the alternative that it has some effect. Perform the test of 5% significance level. Fully document the test according to the test template.

Given the result of the test, what would be your reply to Ms Bowl?

****

B) (6p) Now, Ms Bowl claims that the reason for the result in the above task is that model (2) is no good to begin with (and therefore the effect of number of inhabitants is polluted and somehow 'relatively margarinilized from a dairy point of view'[2]). Instead of arguing with her: Perform a test of model (2) . Use significance level 1% and fully document the test as outlined in the test template.

****

Now, Ms Bowl takes a time out and travels to Thailand, where she cannot be reached for three weeks. Her collegue Mr A. Y. Giveman reluctlantly admits that it is possible that it is an alternative fact that an increase in number of inhabitants alone will increase the number of new enterprises, or more generally, that is might be an even bigger alternative fact that a population increase *in of itself* (not considering age or level of education), would guarantee

[2] It is not very clear what this means, and the student is highly recommended not to spend any time or thought energy whatsoever contemplating this statement.

increased economic growth, which is a statement that may or may not have been given in various instances.[3] In any event, he claims that tax rate is the *only* thing that is important in this context.[4] So, he claims that model (1) is the correct one. And he argues that age and level of education should *not* be added to that model. That is, in statistical terms, he thinks that age and level of education does *not* add any explanatory power whatsoever.

In the coffee break, he asks you to secretely perform a test for that, and he also wants you to promise not to tell anyone about the result unless he gives the go-ahead. Of course, being a statistician with integrity and always striving to reveal the truth, you firmly reject his proposal!

C) (6p) So, now, for the world to see, perform a test to test the null hypothesis that, over and above tax rate and tax rate alone, age and level of education does *not* add any explanatory power whatsoever, against the alternative that they explain at least some of the variation in the dependent variable.

- Write out any model(s) that you need to specify to perform this test, name them in an intuitive way, (which should relate to the statistics you use to perform the test).
- Use significance level 1 percent.
- Follow the test templete to fully document the test.

****

Now, Mr Giveman says that he hired a statistician (Mr C.O. Rupt) that estimated model (1) and that this guy had performed a test of the model, and that the test showed that indeed that tax rate caused an increase in new enterprises.

D) (3p) Now, cleary Mr Giveman is mislead by Mr Rupt, and Mr Giveman needs a crash course in basic regression analysis. Explain to him why it is not possible, from a statistic point of view (without relating to economic theory), to claim that Tax Rate *causes* a change in new enterprises.

E) (3p) Some economists claim that, given the overall economic situation in Sweden, it is very likely that the tax rates in the municipalities must be raised in the coming years. Now,

[3] On an unrelated note: Rumour has it that Mr Giveman was later severly scolded for this statement by his boss Mr S. Leafóne in front of all his collegues. Before he resigned to spend more time with his family, he made a public statement claiming that he did not express himself very clearly, or that what he said was taken out of context, or that he was simply misunderstood, or that dark forces had misrepresented his statement in alternative media, or that he potentially suffers from narkolepsi and actually was unknowingly talking in his sleep (having a surrealistic nightmare), or any combination thereof, but most likely the statement was a fake news to begin with.

[4] He thinks that the higher the tax rate, the more enterprises will be started, but we do not pursue to test this claim for the time being.

assuming that the economic theory that states that Tax Rate changes *causes* changes in number of new enterprises is indeed correct, you decide to give Mr Giveman a concrete example of the effect of the number of new enterprises given a tax change. Using the estimation output from estimation of model (2) - tell him what would be the expected change in the number of new enterprises, if the tax rate was increased by a whopping 6 (six) percentage points, that is for example from 34% to 40%. For simplicity - round to the nearest integer.

****

Now, Mr Giveman calls Mr Rupt and it is clear that Mr Giveman is very upset, but all of a sudden, he calms down, smiles, hangs up and come back to you and now he seem to have completely forgotten what you just talked about and he now claims that the reason that model (2) is a better model than model (1) is that it has more explanatory variables than model (1) .

F) (4p) Use a relevant measure of explanatory power to compare the two models. Explain why you use this specific measure and not any other measure. Given this measure and this measure alone, which of the two models is better from a statistical point of view?

****

Mr Giveman calls Mr Rupt again and comes back claiming that Mr Rupt performed a test of normality of the error term (of model (2)) and he rejected normality (see Eviews ouput in Figure 5), and thus, all the inference that has been done regarding model (2) is incorrect.

G) (4p) Now, even if this would be correct that the error term is *not* normally distributed, (you do not need to confirm his claim by doing this normality test) is this a problem with respect to the inference regarding model (2)? If so, how? If not, why?

****

Being relentless, Mr Giveman consults another statistician, Mr S. Nake, who claims that according to the output from the estimation of the model (2) there are indications of autocorrelation, that is, he claims that the error term is correlated over time, and thus the inference from model (2) is complete nonsense.

H) (2p) Is this possible? If so, what are the consequences regarding inference? If not, how come?

Task 4

(21 points in total) Again, for relevant Eviews output - see Appendix.

Having lost your patience with Mr Giveman and his yes-men, you simply walk away. But then you bump into Mr M. LadyMountain. He is a well educated man. In his younger years, he actually took a very good econometrics class at one of the top universities in Sweden. He recently came back from a trip to USA where he, among other things, visited Texas and studied the cattle industry.

Thinking back on his experiences visiting the ranches, visualizing in his mind the scenic landscape with all the cows and bulls, he wonders if your model (2) might suffer from multicollinearity.

A) (4p) Use relevant output to comment on this.
****

Since it was quite a few years since he took the course, he does not really remember the consequences of multicollinearity.

B) (4p) Without acknowleding or denying that there is a problem with multicollinearity in your model, explain what the consequences of multicollinearity are. Is OLS still BLUE? What will be the (potential) implications of the inference regarding, say t-tests?

****

Being very thankful for your input, he suggests that the two of you should go hot-air-balloning next weekend.
****

Also, he visited the south of Florida, studying alligator farming, and he was very impressed with the success of this industry. When thinking back on this, he wonders if there there might be a problem with heteroscedasticity in your model (2).

C) (4p) Again, it was a long time since he took the course, so

- Explain the (potential) problem if there was an issue with heteroscedasticity, what happens with the OLS-point estimates? Are they unbiased?
- What happens (if anything) with the OLS estimates of the variance of the estimators? What will be the (potential) implications of the inference regarding, say t-tests?

****

You consult the literature, and you come across research performed by Ms K. R. Odille, and she claims that, for this kind of model, u_i has expected value zero and variance

$$\sigma_i^2 = \sigma^2 TaxRate_i. \tag{3}$$

D) (5p) Suggest a way to test this claim. State every steps in how to perform such a test, that is, what model(s) if any, would need to be estimated. (Note that you cannot do the test, there is no relevant output, just outline exactly how you would perform such test if you had the data). That is

- Write down any model(s) if any, you would need to estimate
- What numerical quantities would you need to calculate?
- State the (both the) hypotheses of the test
- State the estimator/statistic that you would use
- State the assumptions of the test
- State the test-statistics you would use, with degrees of freedom (if applicable), cleary define all the quantities involved.
- Draw a figure of the distribution under the null of the test statistic and
- State the rejection rule.

E) (4p) Assuming that her claim is indeed correct, suggest a transform of model (2) to remedy the situation. That is,

1. Perform the transformation, that is, write down the tranformed model.
2. Show that the error term in the transformed model indeed has constant variance.

Task 5

(18 points in total)

Consider the following model

$$Y_i = \beta X_i + u_i$$

A) (6p) Derive the OLS estimator for β. State any assumptions that you make as you make them.

B) (6p) Derive the expected value of the OLS estimator. State any assumptions, if any, that you make as you make them.

C) (6p) Under the assumption of $Var(u_i) = \sigma_i^2$ but $Cov(u_i, u_j) = 0$, derive the variance of the OLS estimator. State the assumptions that you make when you make them, not just state them at the top of the derivation, state explitly when you make use of them.

Appendix - Eviews output

Dependent Variable: NEWENT
Method: Least Squares
Date: 03/21/17 Time: 17:18
Sample: 1 291
Included observations: 291

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	49.82891	3.685143	13.52157	0.0000
TAXRATE	-1.198231	0.111905	-10.70753	0.0000

R-squared	0.284035	Mean dependent var	10.39588
Adjusted R-squared	0.281558	S.D. dependent var	2.679382
S.E. of regression	2.271070	Akaike info criterion	4.485228
Sum squared resid	1490.592	Schwarz criterion	4.510474
Log likelihood	-650.6007	Hannan-Quinn criter.	4.495342
F-statistic	114.6513	Durbin-Watson stat	1.982084
Prob(F-statistic)	0.000000		

Figure 1: Eviews output

Dependent Variable: NEWENT
Method: Least Squares
Date: 03/21/17 Time: 18:48
Sample: 1 291
Included observations: 291

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	25.46318	4.578276	5.561741	0.0000
TAXRATE	-0.835683	0.128198	-6.518672	0.0000
AGE	0.212659	0.062815	3.385470	0.0008
UNIV	0.172709	0.021946	7.869775	0.0000

R-squared	0.411517	Mean dependent var	10.39588
Adjusted R-squared	0.405365	S.D. dependent var	2.679382
S.E. of regression	2.066140	Akaike info criterion	4.302891
Sum squared resid	1225.184	Schwarz criterion	4.353384
Log likelihood	-622.0707	Hannan-Quinn criter.	4.323119
F-statistic	66.89814	Durbin-Watson stat	1.835916
Prob(F-statistic)	0.000000		

Figure 2: Eviews ouput

Dependent Variable: NEWENT
Method: Least Squares
Date: 03/21/17 Time: 18:58
Sample: 1 291
Included observations: 291

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	47.12746	3.817073	12.34649	0.0000
TAXRATE	-1.121099	0.115348	-9.719301	0.0000
INHAB	0.004732	0.001935	2.445858	0.0150

R-squared	0.298605	Mean dependent var	10.39588
Adjusted R-squared	0.293734	S.D. dependent var	2.679382
S.E. of regression	2.251743	Akaike info criterion	4.471542
Sum squared resid	1460.260	Schwarz criterion	4.509411
Log likelihood	-647.6094	Hannan-Quinn criter.	4.486713
F-statistic	61.30501	Durbin-Watson stat	1.968110
Prob(F-statistic)	0.000000		

Figure 3: Eviews output

Dependent Variable: NEWENT
Method: Least Squares
Date: 03/21/17 Time: 19:01
Sample: 1 291
Included observations: 291

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	25.37174	4.591476	5.525835	0.0000
TAXRATE	-0.836579	0.128412	-6.514814	0.0000
AGE	0.215946	0.063505	3.400457	0.0008
UNIV	0.170193	0.022960	7.412489	0.0000
INHAB	0.000737	0.001945	0.378733	0.7052

R-squared	0.411812	Mean dependent var	10.39588
Adjusted R-squared	0.403585	S.D. dependent var	2.679382
S.E. of regression	2.069230	Akaike info criterion	4.309263
Sum squared resid	1224.570	Schwarz criterion	4.372378
Log likelihood	-621.9978	Hannan-Quinn criter.	4.334547
F-statistic	50.05972	Durbin-Watson stat	1.836821
Prob(F-statistic)	0.000000		

Figure 4: Eviews ouput

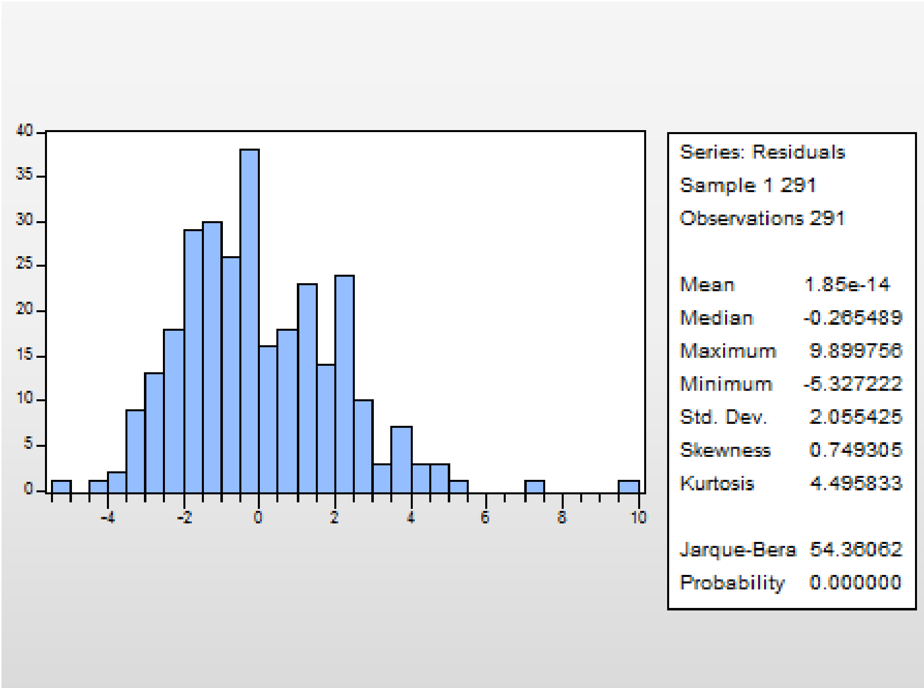


Figure 5: Eviews output with respect to estimation of the model $NewEnt_i = \beta_0 + \beta_1 TaxRate_i + \beta_3 Age + \beta_4 Univ + u_i$

Correlation					
	NEWENT	TAXRATE	AGE	UNIV	INHAB
NEWENT	1.000000	-0.532950	-0.307332	0.565107	0.261811
TAXRATE	-0.532950	1.000000	0.541781	-0.558413	-0.273397
AGE	-0.307332	0.541781	1.000000	-0.640348	-0.388331
UNIV	0.565107	-0.558413	-0.640348	1.000000	0.460343
INHAB	0.261811	-0.273397	-0.388331	0.460343	1.000000

Figure 6: Eviews output

Dependent Variable: TAXRATE
Method: Least Squares
Date: 03/22/17 Time: 13:04
Sample: 1 291
Included observations: 291

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	27.81551	1.319822	21.07519	0.0000
AGE	0.141651	0.027640	5.124893	0.0000
UNIV	-0.056078	0.009531	-5.883927	0.0000

R-squared	0.369339	Mean dependent var	32.90938
Adjusted R-squared	0.364959	S.D. dependent var	1.191737
S.E. of regression	0.949688	Akaike info criterion	2.744890
Sum squared resid	259.7495	Schwarz criterion	2.782759
Log likelihood	-396.3815	Hannan-Quinn criter.	2.760061
F-statistic	84.33187	Durbin-Watson stat	2.027139
Prob(F-statistic)	0.000000		

Dependent Variable: AGE
Method: Least Squares
Date: 03/22/17 Time: 13:05
Sample: 1 291
Included observations: 291

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	27.07528	3.987447	6.790128	0.0000
UNIV	-0.169264	0.018009	-9.398580	0.0000
TAXRATE	0.590003	0.115125	5.124893	0.0000

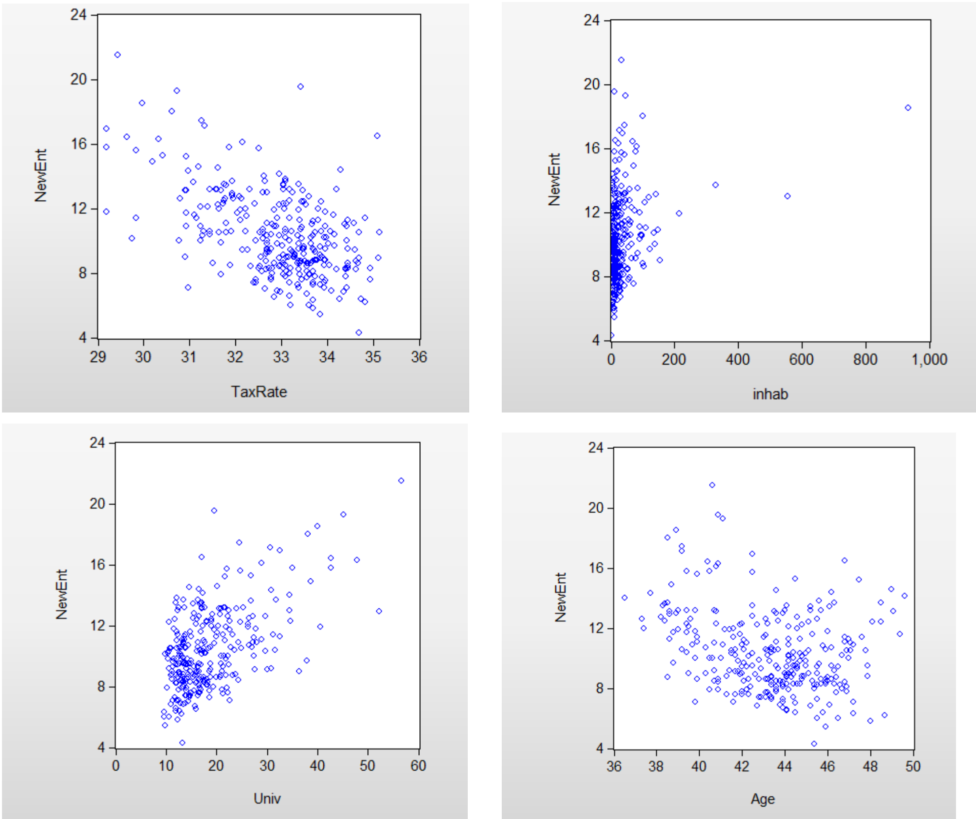
R-squared	0.459351	Mean dependent var	43.33780
Adjusted R-squared	0.455597	S.D. dependent var	2.626867
S.E. of regression	1.938199	Akaike info criterion	4.171652
Sum squared resid	1081.906	Schwarz criterion	4.209521
Log likelihood	-603.9753	Hannan-Quinn criter.	4.186822
F-statistic	122.3466	Durbin-Watson stat	2.082791
Prob(F-statistic)	0.000000		

Dependent Variable: UNIV
Method: Least Squares
Date: 03/22/17 Time: 13:05
Sample: 1 291
Included observations: 291

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	141.7071	9.021601	15.70753	0.0000
TAXRATE	-1.913600	0.325225	-5.883927	0.0000
AGE	-1.386719	0.147546	-9.398580	0.0000

R-squared	0.473354	Mean dependent var	18.63436
Adjusted R-squared	0.469697	S.D. dependent var	7.618140
S.E. of regression	5.547675	Akaike info criterion	6.274891
Sum squared resid	8863.690	Schwarz criterion	6.312760
Log likelihood	-909.9966	Hannan-Quinn criter.	6.290061
F-statistic	129.4285	Durbin-Watson stat	2.047702
Prob(F-statistic)	0.000000		

Figure 7: Eviews output


NewEnt
TaxRate
24
20
16
12
8
4
29 30 31 32 33 34 35 36
NewEnt
inhab
24
20
16
12
8
4
0 200 400 600 800 1,000
NewEnt
Univ
24
20
16
12
8
4
0 10 20 30 40 50 60
NewEnt
Age
24
20
16
12
8
4
36 38 40 42 44 46 48 50