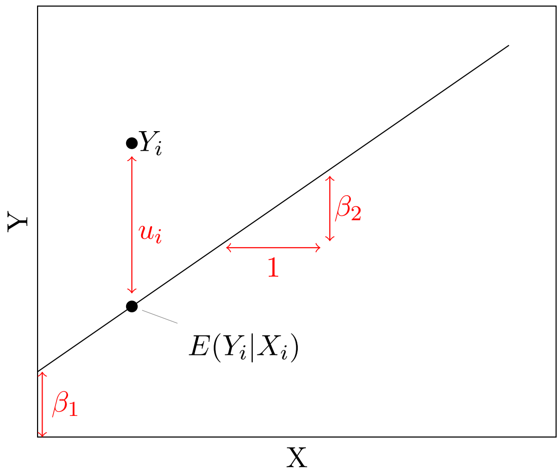
Written Examination in Econometrics
2017-03-24
Solutions

Task 1

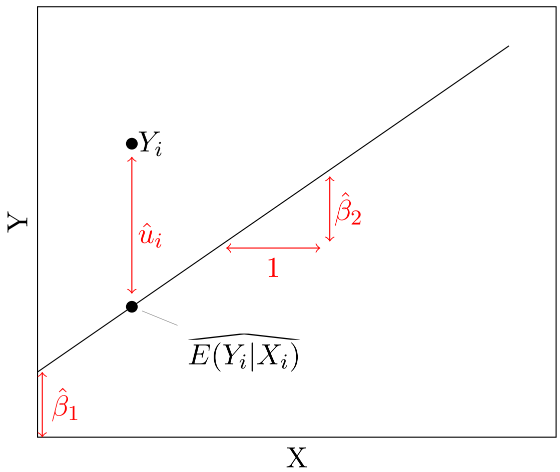
A.

$$\text{PRF} : E(Y_i|X_i) = \beta_1 + \beta_2 X_i$$



B.

$$\text{SRF} : \widehat{E(Y_i|X_i)} = \hat{\beta}_1 + \hat{\beta}_2 X_i$$



Task 2

The model of interest:

$$NewEnt_i = \beta_0 + \beta_1 TaxRate_i + u_i$$

A.

In the model of interest, β_1 is the expected change in the conditional mean of number of new enterprises, given a one percentage point increase in Tax Rate.

B.

From Figure (1) we can see that the estimated expected change in the conditional mean is 1.2 fewer new enterprises, given a one percentage point increase in Tax Rate.

C.

From the coefficient of determination, R^2, we can see that this model explains about 28 percent of the variation in the number of new enterprises.

D.

We are testing this statement with an one-sided t-test.

1. Hypotheses

$H0 : \beta_1 \leq -1$
$H1 : \beta_1 > -1$

2. Significance level

$\alpha = 0.05$

3. Estimator(s)/Statistics

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

4. Assumptions

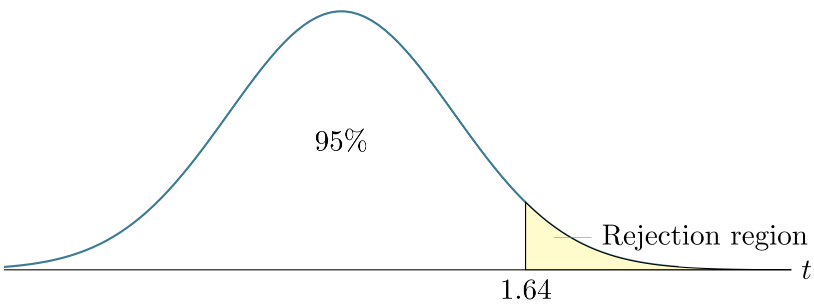
$u_i \sim NID(0, \sigma^2)$
$Cov(u_i, X_i) = 0$

5. Test statistic

$$t_{obs} = \frac{\hat{\beta}_1 - \beta_1^{H_0}}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{n-k}$$

6. Rejection rule and figure

Reject if $t_{obs} > t_{291-1,0.05} \approx 1.64$



7. Calculations and results

$$\begin{aligned} t_{obs} &= \frac{\hat{\beta}_1 - \beta_1^{H_0}}{\hat{\sigma}_{\hat{\beta}_1}} \\ &= \frac{-1.198... - (-1)}{0.111...} \\ &= -1.771... \\ &< 1.64 \rightarrow H_0 \text{ is not rejected} \end{aligned}$$

8. Conclusion

We cannot reject the null hypothesis at the 5 percent significance level. We do not find evidence that a one percentage point increase in tax rate decreases the number of new enterprises by less than one.

Task 3

The model of interest:

$$NewEnt_i = \beta_0 + \beta_1 TaxRate_i + \beta_2 Age_i + \beta_3 Univ_i + \beta_4 InHab_i + u_i$$

A.

We are testing the statement with a two-sided t-test.

1. Hypotheses

$H_0 : \beta_4 = 0$
$H_1 : \beta_4 \neq 0$

2. Significance level

$\alpha = 0.05$

3. Estimator(s)/Statistics

$\hat{\beta}_4$

4. Assumptions

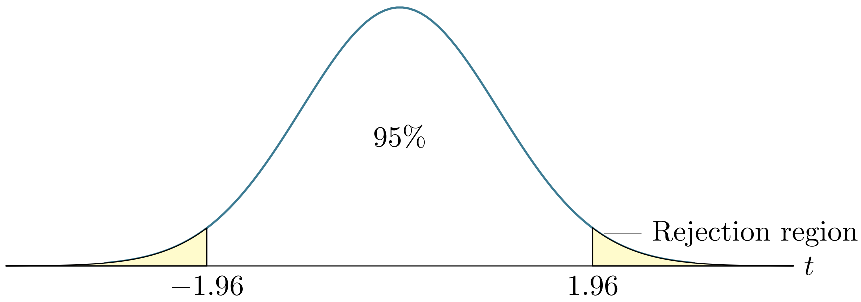
$u_i \sim NID(0, \sigma^2)$
$Cov(u_i, X_i) = 0$
No exact multicollinearity

5. Test statistic

$$t_{obs} = \frac{\hat{\beta}_4 - \beta_4^{H_0}}{\hat{\sigma}_{\hat{\beta}_4}} \sim t_{n-k}$$

6. Rejection rule and figure

Reject if
$t_{obs} > t_{291-5,0.025} \approx 1.96$
or if
$t_{obs} < -t_{291-5,0.025} \approx -1.96$



7. Calculations and results

$$\begin{aligned} t_{obs} &= \frac{\hat{\beta}_4 - \beta_4^{H_0}}{\hat{\sigma}_{\hat{\beta}_4}} \\ &= \frac{0.0007... - 0}{0.0019...} \\ &= -0.378... \end{aligned}$$

Since $-1.96 < t_{obs} < 1.96$, we are in the acceptance region and H_0 is not rejected.

8. Conclusion

We cannot reject the null hypothesis at the 5 percent significance level. We can therefore answer Ms Bowl that we do not find evidence enough to state that the number of inhabitants have an effect on the number of new enterprises.

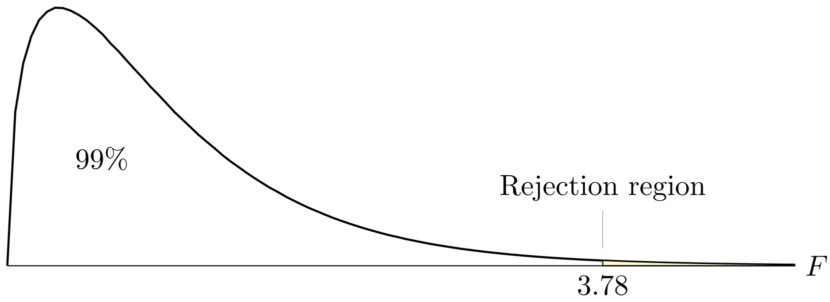
B.

We are testing the model with the F-test. That is, we compare the two models

$$\begin{aligned} \text{Unrestricted: } NewEnt_i &= \beta_0 + \beta_1 TaxRate_i + \beta_2 Age_i + \beta_3 Univ_i + u_i \\ \text{Restricted: } NewEnt_i &= \beta_0 + u_i \end{aligned}$$

1. Hypotheses

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0,$
$H_1 :$ At least one $\beta_j \neq 0,$ for $j = 1, 2, 3$



2. Significance level

$\alpha = 0.01$

3. Estimator(s)/Statistics

$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$
R^2_{UR}, R^2_R

4. Assumptions

$u_i \sim NID(0, \sigma^2)$
$Cov(u_i, X_i) = 0$
No exact multicollinearity

5. Test statistic

$$F_{obs} = \frac{(R_{UR} - R_R)/m}{(1 - R_{UR})/(n-k)} \sim F_{m,(n-k)}$$

where
R^2_{UR} = coefficient of determination for the unrestricted model
R^2_R = coefficient of determination for the restricted model
n = number of observations
m = number of resrictions
k = number of parameters in the unrestricted model

6. Rejection rule and figure

Reject if
$F_{obs} > F_{3,291-4,0.01} = 3.78$

7. Calculations and results

$$R^2_{UR} = 0.411517$$
$$R^2_R = 0$$
$$n = 291$$
$$m = 3$$
$$k = 4$$

$$F_{obs} = \frac{(0.4115... - 0)/3}{(1 - 0.4115)/(291 - 4)}$$
$$= 66.89...$$

Since $F_{obs} > 3.78$, we are in the rejection region and H_0 is therefore rejected.

8. Conclusion

At the 1 percent significance level, we reject the null that the model does not explain any of the variation in the number of new enterprises.

C.

We are now to compare two models, that is, another F-test.

$$\text{Unrestricted: } NewEnt_i = \beta_0 + \beta_1 TaxRate_i + \beta_2 Age_i + \beta_3 Univ_i + u_i$$
$$\text{Restricted: } NewEnt_i = \beta_0 + \beta_1 TaxRate_i + u_i$$

1. Hypotheses

$H_0 : \beta_2 = \beta_3 = 0,$
$H_1 :$ At least one $\beta_j \neq 0,$ for $j = 2, 3$

2. Significance level

$\alpha = 0.01$

3. Estimator(s)/Statistics

$\hat{\beta}_2, \hat{\beta}_3$
R^2_{UR}, R^2_R

4. Assumptions

$u_i \sim NID(0, \sigma^2)$
$Cov(u_i, X_i) = 0$
No exact multicollinearity

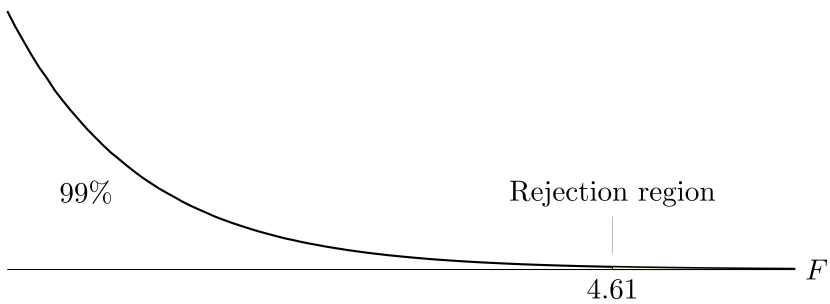
5. Test statistic

$$F_{obs} = \frac{(R_{UR} - R_R)/m}{(1 - R_{UR})/(n-k)} \sim F_{m,(n-k)}$$

where
R^2_{UR} = coefficient of determination for the unrestricted model
R^2_R = coefficient of determination for the restricted model
n = number of observations
m = number of resrictions
k = number of parameters in the unrestricted model

6. Rejection rule and figure

Reject if
$F_{obs} > F_{2,291-4,0.01} = 4.61$



7. Calculations and results

$$R^2_{UR} = 0.411517$$
$$R^2_R = 0.284035$$
$$n = 291$$
$$m = 2$$
$$k = 4$$

$$F_{obs} = \frac{(0.4115... - 0.284...)/2}{(1 - 0.4115)/(291 - 4)}$$
$$= 31.08...$$

Since $F_{obs} > 4.61$, we are in the rejection region and H_0 is rejected.

8. Conclusion

At the 1 percent significance level, we reject the null that age and level of education do not add any explanatory power to the model.

D.

Neither correlation nor regression can indicate the causation. Already when stating the model we must assume (or know) the causal relationship between the included variables, and the intention with the regression analysis is then to describe the dependence of a variable, Y, on an independent variable X.

It is not possible from an *estimated* model to see in what direction the causality goes.

E.

From the estimated model

$$\widehat{NewEnt_i} = 25.5 - 0.84 TaxRate_i + 0.21 Age_i + 0.17 Univ_i$$

we are to calculate the expected change in the number of new enterprises given a 6 percentage points increase in the Tax Rate.

$$\widehat{E(NewEnt_{i,TaxRate=x})} = 25.5 - 0.84 * x + 0.21 Age_i + 0.17 Univ_i$$
$$\widehat{E(NewEnt_{i,TaxRate=x-6})} = 25.5 - 0.84 * (x - 6) + 0.21 Age_i + 0.17 Univ_i$$

and the difference is calculated as

$$\widehat{E(NewEnt_{i,TaxRate=x})} - \widehat{E(NewEnt_{i,TaxRate=x-6})} = -0.84 * x - (-0.84) * (x - 6)$$
$$= -0.84 * x + 0.84 * (x - 6)$$
$$= -5.014$$

Or, we can see from the output that the estimated expected change in new enterprises given a one percentage ponint increase in Tax Rate, everything else held constant, is -0.84. A 6 percentage point increase can therfore be calculated as

$$-0.84 * 6 \approx -5.0$$

So, the estimated expected change is about 5 new companies less in a municipality, when the tax rate is increased by 6 percentage points, everything else is held constant.

F.

The explanatory power of the models can be compared with the adjusted R^2. Looking at the outputs in Figure (1) and (2), it is clear that model 2 has the highest R^2_{adj}.

Model 2 explains 40.5 percent of the variation in the number of new enterprises, taking the number of explanatory variables into account (to be compared with 28.2 percent for model 1).

G.

The assumption of normally distributed error terms can be relaxed in the case of large samples. When the sample size is large, the estimators will be normally distributed (from the central limit theorem) and inference will be valid, also in the case of non-normal errors.

In this case the sample size of 291 can be considered large enough for inference to be valid.

H.

Cross-sectional data is used in this case (we have a sample from one time point) and therefore autocorrelation over time is not an issue in this case.

The OLS estimators are still unbiased and consistent in the case of autocorrelation, but not BLUE. The variance of the estimators will not be correctly calculated. We would under- or overestimate the true variance and inference would therefore be misleading.

Task 4

A.

Looking at the correlation matrix in Figure (6), the correlation between the regressors doesn't seem problematic.

To examine the multiple regression model for multicollinearity an examination of the auxillary regressions in Figure (7) is needed. It doesn't seem like any of the three explanatory variables to a large extent can be explained by the others (the R^2 is low for all the regressions). That is, there is no indication of severe multicollinearity in model 2.

B.

The estimators are still unbiased in the case of multicollinearity. The variance is increased (or inflated) but the estimators are still BLUE. Severe multicollinearity could lead to too small t-statistics and parameters not beeing significantly different from zero.

C.

If there would be a problem of heteroscedasticity in the model, the estimators would still be unbiased. However, they would no longer have minimum variance, that is they would not longer be BLUE.

In the case of heteroscedasticity the estimated variance would be under- or overestimated. When overestimating the variance, the model would seem worse than it actually is. If the variance on the other hand is underestimated, the model would seem better than it actually is.

D.

There are several tests that could be used to detect heteroscedasticity (Park's, Gleiser, GQ-test, BPG-test or White's test). In this case we use the Breusch-Pagan test to test the claim that

$$\sigma_i^2 = \sigma^2 |TaxRate_i|$$

We start by running the regression

$$NewEnt_i = \beta_0 + \beta_1 TaxRate_i + \beta_2 Age_i + \beta_3 Univ_i + u_i$$

to obtain the residuals. We then run the auxiliary regression

$$\hat{u}_i^2 = \alpha_0 + \alpha_1 TaxRate_i + e_i$$

1. Hypotheses

$H_0 : \alpha_1 = 0,$
$H_1 : \alpha_1 \neq 0$

2. Significance level

$\alpha = 0.05$

3. Estimator(s)/Statistics

R^2 from the auxilary regression

4. Assumptions

$u_i \sim NID(0, \sigma_i^2)$

5. Test statistic

$LM = nR^2 \sim \chi^2_{k-1}$

6. Rejection rule

Reject if
$nR^2 > \chi^2_{1,0.05} = 3.8$

E.

Assuming that

$$\sigma_i^2 = \sigma^2 |TaxRate_i|$$

we are to transform

$$NewEnt_i = \beta_0 + \beta_1 TaxRate_i + \beta_2 Age_i + \beta_3 Univ_i + u_i$$

to remedy the situation. Dividing both sides with $\sqrt{|TaxRate_i|}$ gives

$$\frac{NewEnt_i}{\sqrt{|TaxRate_i|}} = \frac{\beta_0}{\sqrt{|TaxRate_i|}} + \frac{\beta_1 TaxRate_i}{\sqrt{|TaxRate_i|}} + \frac{\beta_2 Age_i}{\sqrt{|TaxRate_i|}} + \frac{\beta_3 Univ_i}{\sqrt{|TaxRate_i|}} + \frac{u_i}{\sqrt{|TaxRate_i|}}$$

Rewriting this equation as

$$NewEnt_i^* = \beta_0^* + \beta_1^* TaxRate_i + \beta_2^* Age_i + \beta_3^* Univ_i + u_i^*$$

then

$$\begin{aligned} Var(u_i^*) &= Var\left(\frac{u_i}{\sqrt{|TaxRate_i|}}\right) \\ &= \frac{Var(u_i)}{|TaxRate_i|} \\ &= \frac{\sigma^2 |TaxRate_i|}{|TaxRate_i|} \\ &= \sigma^2 \end{aligned}$$

that is, the error term in the transformed model has constant variance.

Task 5

A.

We have the following model

$$Y_i = \beta X_i + u_i$$

and are to derive the OLS estimator of β. The OLS estimator is the value of β that minimizes the sum of squared residuals. We find the minimum by differentiating w.r.t. β and put the derivative equal to zero

$$\begin{aligned}
\min_{\beta} \sum_{i=1}^{n} \hat{u}_i^2 &= \frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \beta} \\
&= \sum_{i=1}^{n} \frac{\partial \hat{u}_i^2}{\partial \beta} \\
&= \sum_{i=1}^{n} \frac{\partial (Y_i - \hat{\beta} X_i)^2}{\partial \beta} \\
&= -2 \sum_{i=1}^{n} (Y_i - \hat{\beta} X_i) X_i \\
&= -2 \sum_{i=1}^{n} (Y_i X_i - \hat{\beta} X_i^2) \\
&\text{put equal to zero to find minimum} \\
&\rightarrow \sum_{i=1}^{n} Y_i X_i - \hat{\beta} \sum_{i=1}^{n} X_i^2 = 0
\end{aligned}$$

The estimated slope parameter can then be written as

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2}$$

B.

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2}$$

$$= \frac{\sum_{i=1}^{n} (\beta X_i + u_i) X_i}{\sum_{i=1}^{n} X_i^2}$$

$$= \frac{\beta \sum_{i=1}^{n} X_i^2 + u_i X_i}{\sum_{i=1}^{n} X_i^2}$$

$$= \beta + \frac{\sum_{i=1}^{n} u_i X_i}{\sum_{i=1}^{n} X_i^2}$$

The expected value of β is then

$$E(\hat{\beta}) = E\left(\beta + \frac{\sum_{i=1}^{n} u_i X_i}{\sum_{i=1}^{n} X_i^2}\right)$$

If X is assumed to be fixed, or $Cov(X_i, u_i) = 0$

$$E(\hat{\beta}) = E(\beta) + \frac{\sum_{i=1}^{n} E(u_i) X_i}{\sum_{i=1}^{n} X_i^2}$$

With the assumption $E(u_i) = 0$, and since β is a constant,

$$E(\hat{\beta}) = \beta + 0$$

and it is shown that β is unbiased.

C.

$$
\begin{aligned}
Var(\hat{\beta}) &= E\left(\hat{\beta} - E(\hat{\beta})\right)^2 \\
&= E\left[\left(\beta + \frac{\sum_{i=1}^{n} u_i X_i}{\sum_{i=1}^{n} X_i^2} - \beta\right)^2\right] \\
&= E\left[\frac{\left(\sum_{i=1}^{n} u_i X_i\right)^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2}\right]
\end{aligned}
$$

Since

$$\left(\sum_{i=1}^{n} u_i X_i\right)^2 = \sum_{i=1}^{n} u_i^2 X_i^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} u_i u_j X_j X_i$$

$$Var(\hat{\beta}) = E\left[\frac{\sum_{i=2}^{n} u_i^2 X_i^2 + 2\sum_{j=1}^{i-1}\sum_{i=1}^{n} u_i u_j X_j X_i}{\left(\sum_{i=1}^{n} X_i^2\right)^2}\right]$$

And if X is assumed to be fixed, or $Cov(X_i, u_i) = 0$,

$$Var(\hat{\beta}) = \frac{\sum_{i=2}^{n} E(u_i^2) X_i^2 + 2\sum_{j=1}^{i-1}\sum_{i=1}^{n} E(u_i u_j) X_j X_i}{\left(\sum_{i=1}^{n} X_i^2\right)^2}$$

If we ssume that $E(u_i) = 0$, $Var(u_i) = \sigma_i^2$ and $Cov(u_i, u_j) = 0\ \forall\ i \neq j$, then $E(u_i^2) = \sigma_i^2$ and $E(u_i u_j) = 0$ such that

$$
\begin{aligned}
Var(\hat{\beta}) &= \frac{\sum_{i=2}^{n} \sigma_i^2 X_i^2 + 0}{\left(\sum_{i=1}^{n} X_i^2\right)^2} \\
&= \frac{\sum_{i=2}^{n} \sigma_i^2 X_i^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2}
\end{aligned}
$$