

**Take home exam;
Due: Wednesday September 24**

Exercise 1

- (a) The increasing/decreasing reputation by 1 unit change tuition by 3,985.20 dollars. Each student decrease tuition by 0.20 dollars. If university is private - it's increase tuition by 8406.79 dollars. If university is liberal arts college - it's decrease tuition by 416.38. And if university is religious it's decrease tuition by 2376.51 dollar. All coefficient have an expected sign.

(b) $\hat{Cost} = 7311.17 + 3985.20 \cdot 4.5 - 0.20 \cdot 1500 + 8406.79 \cdot 1 - 416.38 \cdot 1 - 2376.51 \cdot 0 = 32935$

- (c) If out college become to the public, it get -0.5 points in rank and 10 000 more students. The new tuition is:

$$\hat{Cost} = 7311.17 + 3985.20 \cdot 4.0 - 0.20 \cdot 11500 + 8406.79 \cdot 0 - 416.38 \cdot 1 - 2376.51 \cdot 0 = 20535.6$$

The difference tuition is 12 400 dollars. It's very substantial. Is student study at the college 4 years he can save 49600 dollars. Student can by a small flat in Moscow region by this money.

- (d) It's increased because Size and Dlibart had a negative sign before the coefficient.
- (e) It's possible that the university president and chief academic officer have personal reasons for rising a tuition. This can cause serious damage to the validity of the regression.

Exercise 2

- (a) We can calculate t statistic

$$t_n = \frac{\beta_N}{SE_N} = \frac{-12.694}{0.377} = -3.99 \quad t_s = \frac{\beta_S}{SE_S} = \frac{1.397}{0.229} = 6.100437$$

We tested at 5% significance level that coefficients are equal to zero. We have to reject it because all |t-statistics| > 1.96

R^2 shows that our regression fits the data better than 50% but worse than 100% because it is 62.1%.

- (b) Our regression is $-5.869 \cdot N + 0.738 \cdot S + 0.055 \cdot \text{Educ} + 0.046 = 0.239973$. Thus the regression underpredicts. If Brazil doubles Educ then the regression gives us 0.432473.

- (c) $R_{inc} = -0.063 * N + 0.719 * S + 0.044 * Educ - 0.068$ when $OECD = 0$
 $R_{inc} = -0.063 * N + 0.719 * S + 0.044 * Educ - 0.068 + 0.381 + -8.038 * N - 0.430 * S + 0.003 * Educ$ when $OECD = 1$ That is equal to:

$$R_{inc} = -8.101 * N + 0.289 * S + 0.047 * Educ + 0.313$$

Then we can find from appendix in introduction to econometric that critical value on 5% for A statistic with 4 parameters is 2.37. According to the task F-statistic is 6.76 for hypothesis that all coefficients with OECD are equal to zero. We have to reject it due to the fact that 6.76 is bigger than 2.37.

- (d) $t - statistic = \frac{0.241}{0.055} = 4.381818$, so we have to reject the hypothesis that coefficient on OECD is close to zero because 4.38 is bigger than 1.96 - critical value on 5% value.

Confidence intervals are $0.241 \pm 1.96 * 0.055$

[0.1332, 0.3488]

- (e) F-statistic is 1.05. The critical value is 2.6. So we cannot reject the hypothesis that coefficients of joint OECD are equal to zero. The difference between previous example is the fact that we didn't count coefficient on OECD. That shows that joint events are very unlikely in our data.

Exercise 3

- (a) Obviously, the farm status = false, and education decrease probability for live woman for her own. It's true for all race. Also for white woman the probability of independent live increase during the age, but after the one point this probability is start to decrease it cause the square of age in the regression. The white women in the south are more likely live for their own then non-white women. An increase in expected family earnings and family composition increase the probability of females living on their own
- (b) 1) The signification interval for age with 5% signification level is (for whites women) is:
 $-0.275 - 1.96 \cdot 0.037, -0.275 + 1.96 \cdot 0.037 = -0.34752, -0.20248$
 Zero isn't in this signification interval, hence age is significantly different from zero with 5% signification level.
- 2) The signification interval for age with 5% signification level is (for nonwhites women) is:
 $\{0.084 - 1.96 \cdot 0.068, 0.084 + 1.96 \cdot 0.068\} = \{-0.04928, 0.21728\}$.
 Zero is in this signification interval, hence age isn't significantly different from zero with 5% signification level.

- 3) The signification interval for age squared with 5% signification level is (for whites women) is: $\{0.00463 - 1.96 \cdot 0.00044, 0.00463 + 1.96 \cdot 0.00044\} = \{0.0037676, 0.0054924\}$
Zero isn't in this signification interval, hence age squared is significantly different from zero with 5% signification level.
- 4) The signification interval for age squared with 5% signification level is (for nonwhites women) is:
 $\{0.00021 - 1.96 \cdot 0.00081, 0.00021 + 1.96 \cdot 0.00081\} = \{-0.0013776, 0.0017976\}$.
Zero is in this signification interval, hence age squared isn't significantly different from zero with 5% signification level.
- (c) The F-statistic is greater then 3.00 for white women. That's why we reject this null hypothesis with 5% signification level.
And for nonwhite women F-statistic is less then 3.00. That's why we accept this null hypothesis with 5% signification level.
- (d) For white women: 0.90
For nonwhite women: 0.88
If we mix white and nonwhite women we need have the same coefficients in the regression. And it's reduces the accuracy of forecasting.
- (e) It's decrease to 0.81

Exercise 4

- a) We have next constraints for valid using IV method:
- 1) Correlation with parameters: $\text{corr}(Z_i, X_i) \neq 0$
 - 2) Exogeneity: $\text{corr}(Z_i, u_i) = 0$
- Obviously, the temperature one month ahead in previous year is correlated with current month. But it's doesn't correlated with current residuals, such us the hurricane, storms, rains and etc.
- b) We can use first-stage F-statistic. If $F < 10$ - we have a weak instrument.
- c) The F-statistic > 10 , hence we have strong instrument. I assume that TSLS estimator calculate by software and S.E. is correct. Hence, the significant interval with 5% signification level is:
 $\{1.07 - 1.96 \cdot 0.06, 1.07 + 1.96 \cdot 0.06\} = \{0.9524, 1.1876\}$
One is in significant interval, hence we accept this hypothesis with 5% signification level.

Exercise 5

- a) Lets calculate signification interval with 5% signification level for σ_{t-1}^2 :
 $\{0.53 - 1.96 \cdot 0.15, 0.53 + 1.96 \cdot 0.15\} = \{0.236, 0.824\}$
Zero isn't in interval, that's why we reject hypothesis $\sigma_{t-1}^2 = 0$ at the 5% signification level.
Lets calculate signification interval with 5% signification level for u_{t-1}^2 :
 $\{0.27 - 1.96 \cdot 0.11, 0.27 + 1.96 \cdot 0.11\} = \{0.0544, 0.4856\}$
Zero isn't in interval, that's why we reject hypothesis $u_{t-1}^2 = 0$ at the 5% signification level.
Both σ_{t-1}^2 and u_{t-1}^2 are statistical significant on 5% signification level.
- b)
- c) All the models do the regression by the same variables but with different coefficients. Moreover regression 2 assumes that error u_t is exact. GARCH model assumes that ν_t is a random variable with $N(0,1)$ distribution. I personally think that due to the fact that as regressors we use lags we have to deal with GARCH model and OLS will give us wrong results.
- d) According to a) the variance of the error terms in the Phillips Curve for Canada is statistically significant. Moreover we know that it has autoregressive character (depends on previous values of itself). So there will be volatility clustering in this regression.
- e) We can see that residuals depends on self previous values and so we can observe volatility clustering. Also we see that the process fluctuates across zero, so the mean is obviously close to zero. Also we can see crisis e.g. on 91 period there is big positive crisis.