Advanced Econometrics, S3

2012–2013

I Mid-Term Exam 12.12.2012

All four questions are worth six points. Write your solutions in English, Finnish or Swedish. Please, return the question paper with your solutions.

1. Consider the following linear regression model

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i \\ &= \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, N. \end{aligned}$$

Assume that the Gauss-Markov assumptions are valid and the error term ε_i follows a normal distribution.

(a) Derive the ordinary least squares (OLS) estimator $\widehat{\boldsymbol{\beta}}$ for the parameter vector $\boldsymbol{\beta}$. (2pts)

(b) Show that under the Gauss-Markov assumptions the OLS estimator is an unbiased estimator of $\boldsymbol{\beta}$. What does the unbiasedness of the estimator mean? (1pt)

(c) Explain how one can test the hypothesis $\beta_3 + \beta_4 = 1$. Similarly, how can one test the hypothesis $\beta_2 = \beta_3 = 0$? (2pts)

(d) The OLS estimator is also a consistent estimator (under, e.g., the Gauss-Markov assumptions). What does the consistency of the estimator mean? (1pt)

2. Consider the following linear regression model

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \\ &= \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i. \end{aligned}$$

Unless stated otherwise, the usual four OLS assumptions hold:

1. $E(\varepsilon_i | \boldsymbol{x}_i) = 0, \quad i = 1, \ldots, N.$

2. $(\boldsymbol{x}_i, y_i), i = 1, \ldots, N$ are independent and identically distributed (IID) draws from their joint distribution.

3. Large outliers in $\boldsymbol{x}_i$ and y_i are unlikely (fourth moments are finite).

4. $\boldsymbol{X} = [\boldsymbol{x}_1^{'} \quad \ldots \quad \boldsymbol{x}_N^{'}]^{'}$ has a full column rank (no exact multicollinearity).

Are the following statements TRUE or FALSE? Explain *briefly but accurately*. The correct answer without the correct explanation yields no points (1pt/question).

(a) In a two-tail t-test the null hypothesis $H_0 : \beta_1 = 1$ is rejected at the 1% level of significance. Unity is included in the 95% confidence interval of β_1.

(b) Assume that $E(x_{i1}\varepsilon_i) \neq 0$, $\text{Corr}(x_{i2}, \varepsilon_i) = 0$ and $\text{Corr}(x_{i1}, x_{i2}) \neq 0$, the parameters of the model can be estimated consistently by the ordinary least squares.

(c) Assume that there is conditional heteroskedasticity in the error term (i.e. $E(\varepsilon_i^2|\boldsymbol{x}_i) = \sigma_i^2$ $(\sigma_i^2 \neq \sigma_j^2, i \neq j)$. In that case, the OLS estimator is an inconsistent estimator of $\boldsymbol{\beta}$.

(d) To test the hypothesis $\beta_2 = 0$ with the t-test, the critical values coming from the standard normal distribution can be used when the number of observations N is large.

(e) Let us assume $E(\varepsilon_i|x_{i2}) > 0$, but $E(\varepsilon_i|x_{i1}) = 0$. When estimating the model by the two-stage least squares (TSLS) using three instruments, the p-value of the overidentifying restrictions J test statistic is 0.005. This means that the parameters have been estimated consistently.

(f) Assume that x_{i1} is an endogenous variable, x_{i2} is an exogenous variable and there is one instrumental variable. The R^2 (coefficient of determination) in the first-stage regression of the TSLS is very large, say $R^2 > 0.5$, and the instrument z_{i1} is also highly significant predictor in the first-stage model. Can we conclude that the instrument z_{i1} is (relatively) strong?

3. Let us assume that the researcher estimates a linear regression model

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \\ &= \boldsymbol{x}_i^{'}\boldsymbol{\beta} + \varepsilon_i. \end{aligned}$$

In this question, we consider the estimation results of the simulated data ($N = 100, K = 3$) given below.

(a) Here are the estimation results (R regression output) based on the OLS estimator with the OLS (homoskedasticity-only) covariance matrix:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.99686    0.65305  -1.526   0.1301
x_2         -0.06025    0.05111  -1.179   0.2413
x_3          0.10334    0.04726   2.186   0.0312
---------------------------------------------------------
Residual standard error: 3.247 on 97 degrees of freedom
Multiple R-squared: 0.05978,    Adjusted R-squared: 0.0404
F-statistic: 3.084 on 2 and 97 DF,  p-value: 0.0503
```

Interpret (*briefly but accurately*) the estimation results. Is there statistically significant explanatory power in the explanatory variables x_{i2} and x_{i3} (denoted by x_2 and x_3, respectively) at the 5% significance level? How well the model can explain variation in y_i? (2pts)

(b) It turns out that there is evidence of remaining conditional heteroskedasticity in the residuals of the estimated model. The researcher tries to take this effect into account assuming that the asymptotic OLS assumptions are valid leading to the asymptotic distribution result for the OLS estimator

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} N\Big(\mathbf{0}, \boldsymbol{W}\Big),$$

where $\boldsymbol{W} = \boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_V\boldsymbol{\Sigma}_{XX}^{-1}$. The asymptotic covariance matrix can be consistently estimated by

$$\widehat{\boldsymbol{W}} = \frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1},$$

where $\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i' \xrightarrow{p} \boldsymbol{\Sigma}_V$, $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\boldsymbol{x}_i' \xrightarrow{p} \boldsymbol{\Sigma}_{XX}$, and $\boldsymbol{\Sigma}_{XX}$ and $\boldsymbol{\Sigma}_V$ are finite and positive definite. Alternatively, the asymptotic distribution result can be written

$$\frac{1}{\sqrt{\boldsymbol{W}}}\sqrt{N}\Big(\widehat{\beta} - \beta\Big) = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{1}{N}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_V\boldsymbol{\Sigma}_{XX}^{-1}}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}_K).$$

The researcher wants to test a hypothesis $\beta_3 = 0$ in the above-mentioned model. Explain how he/she can construct a heteroskedasticity robust t-test in this case. What is the asymptotic distribution of the test statistic under the null hypothesis? (3pts)

(Hint: You can first specify the hypothesis within the general linear hypothesis framework. Then, you can use the linear transformation of the multivariate normal distribution for the above-mentioned asymptotic distribution result, under the null hypothesis, to construct the test statistic.)

(c) In the following estimation results, the heteroskedasticity-consistent standard errors constructed using the estimator $\widehat{\boldsymbol{W}}$ are reported with the estimated parameter coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.99686	0.472888	-2.1080	0.03761
x_2	-0.06025	0.070375	-0.8562	0.39401
x_3	0.10334	0.069249	1.4922	0.13888

Are the conclusions on the statistical significance of the explanatory variables the same as above in the case of the OLS standard errors? In particular, what is your conclusion on the hypothesis considered in (b) in this case? (1pt)

4. Let us consider the regression model where the dependent variable y_i is binary. In other words, $y_i = 0$ or $y_i = 1$ for all i. Explanatory variables are included in the vector $\boldsymbol{x}_i$. Answer *briefly but accurately* to the following questions.

(a) Give two reasons why the linear regression model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, N,$$

is not really an appropriate model for the binary variable y_i, where $\boldsymbol{x}_i$ is K-dimensional vector and $\boldsymbol{\beta} = [\beta_1 \quad \ldots \quad \beta_{K-1} \quad \beta_K]'$ is K-dimensional parameter vector. (1pt)

(b) An alternative model to the linear regression model in this context is the logit model. Explain the logit model (among other things, give the model formulation and explain how this model differs from the linear regression model). (2pts)

(c) Logit model is typically estimated by the method of maximum likelihood (ML). Explain briefly the basic idea of the ML estimation method. Furthermore, explain how to construct the log-likelihood function in the logit model. (2pts)

(d) Explain how we can test the hypothesis of $\beta_{K-1} = \beta_K$ (the last two coefficients in $\boldsymbol{\beta}$) in the logit model by using the likelihood ratio (LR) test. (1pt)