EXAM: S3: Advanced Econometrics, part I
11th of December 2013

FOR EACH QUESTION, WRITE YOUR ANSWER ON A SEPARATE SHEET AND WRITE YOUR NAME AND STUDENT ID ON EACH OF THE FOUR SHEETS.
For students of Aalto & Hanken: This is also the final exam for Econometrics I: Basics /Econometrics I. Also, remember to write down which exam you participate: "The MID-exam" or "The FINAL-exam".

You might find these following results helpful in answering Questions 1-4 below:

- If Gauss-Markov assumptions hold, $rank(\mathbf{R}) = J$, and the null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ is true, then

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'\left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{Js^2} \sim F_{J,N-K}$$

- Derivatives: Let $\mathbf{A}$ be a $N \times K$ matrix, $\mathbf{B}$ be a symmetric $K \times K$ matrix, and $\mathbf{x}$ be a $K \times 1$ vector, then

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}' \text{ and } \frac{\partial \mathbf{x}'\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{B}\mathbf{x}.$$

1. (a) (3 p.) Write the multiple linear regression model using matrix notation.
   (i) Explain your notation in detail. (Do not forget to explain what the elements are in the matrix of regressors! Assume that you have K regressors and N observations.)
   (ii) Explain what is random/non-random and observed/unobserved in the regression model.
   (b) (3 p.) Derive the OLS estimator for the regression coefficient vector $\boldsymbol{\beta}$ using the residual sum of squares function $S(\boldsymbol{\beta})$.

2. (a) (2 p.) What are the Gauss-Markov assumptions in a multiple linear regression model?
   (b) (2 p.) Show that the OLS estimator $\hat{\boldsymbol{\beta}}$ is unbiased and $\text{Cov}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, when Gauss-Markov assumptions hold.
   (c) (2 p.) Consider the multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i2} + \beta_2 x_{i3} + \beta_3 x_{i4} + \varepsilon_i,$$

   for $i = 1, ..., N$ and $f_{\boldsymbol{\varepsilon}|\mathbf{X}}(\varepsilon|\mathbf{X})$ is $\text{N}_N(0, \sigma^2\mathbf{I}_N)$ and the Gauss-Markov assumptions hold. Explain how to test the hypothesis, $H_0 : \beta_2 - 3\beta_3 = 1$ and $\beta_1 = 0$, using the F-test and explain how to select the elements in $\mathbf{R}$ and $\mathbf{r}$ when writing this hypothesis in the general linear hypothesis form $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

3. (a) (2 p.) If the errors in the linear regression model are heteroskedastic or autocorrelated or both, but other regression assumptions hold, what are then the asymptotic properties of the OLS estimator? Especially, is the asymptotic covariance matrix of the OLS estimator the same as in 2.(b)?
   (b) (2 p.) How do you estimate a multiple linear regression model, when the errors follow $\text{N}_N(0, \sigma^2\boldsymbol{\Psi})$ distribution and $\boldsymbol{\Psi}$ is a known positive definite matrix.
   (c) (2 p.) Write the latent variable presentation for both the Probit and Logit models and explain how to derive the Probit and Logit models from the latent variable presentation. Can you estimate the parameters of the Probit and Logit models with OLS? If not, describe briefly how?

4. (a) (2 p.) Explain briefly the basic idea of the IV regression model. What are the two essential assumptions for a valid instrumental variable in the IV regression?
   (b) (3 p.) Based on the conditional homoskedasticity assumption, the two-stage least squares (TSLS) estimator is asymptotically efficient among the class of IV estimators. Explain the TSLS estimator and how the assumptions behind the valid instruments (see (a)) are related to these two stages.
   (c) (1 p.) Can one test the two essential assumptions for instrumental validity? If yes, explain briefly how.