# Advanced Econometrics

## Take-Home Exam #1, WS 2015/16
## December 18th, 2015 – January 8th, 2015 (16:00h)

Prof. Dr. Winfried Pohlmeier
Department of Economics
University of Konstanz

This exam contains 1 question which is worth 10 points.

Your solution should include answers (summary tables/figures and interpretations), as well as your Matlab code. Please hand in the answers as a hard copy in Sebastian Bayer's office (F309), and send your Matlab codes to sebastian.bayer@uni-konstanz.de in a zipped file with your student ID number as name.[1] Please document your answers and work steps in a way that makes it easy to understand what you did and why. Presentation of the results is a part of grading. Note that answers / interpretations (in the form of extensive comments) within the Matlab code will not be graded.

The hard copy **and** the Matlab code have to be delivered until January 8th, 2016, 16:00h. Solutions delivered after the deadline will not be graded. Do not forget to state your student ID number.

Policy with regard to academic dishonesty:
The grade for your take home exam will be a part of your overall grade of your course. You may program in groups up to three people and every group member can hand in the same Matlab code. Please indicate with whom you worked together. However, every student must formulate and hand in his or her work independently.

Please note that plagiarism is a serious offense! It can be avoided by simply citing the original source of ideas or material that are not your own. Any attempt to plagiarize will be marked with 0 points for all take home exams in this course.

Good Luck!

[1] For example, "691953.zip".

## Linear Regression Under Heteroscedasticity

Consider the Gaussian linear regression model with heteroscedasticity,

$$Y_i = X_i'\beta + \varepsilon_i, \quad \varepsilon \overset{\text{iid}}{\sim} N(0, \sigma_i^2). \tag{1}$$

You assume that the variance of the error term is given by

$$\sigma_i^2 = \exp\left(Z_i'\gamma\right), \tag{2}$$

where $Z_i = (1, Z_{i2}, \ldots, Z_{im})'$ is a $m \times 1$ vector of variance explaining variables including a constant term and $\gamma$ is a corresponding vector of unknown parameters.

Your goal is to estimate the wage equation

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{school} + \beta_2 \text{age} + \beta_3 \text{age}^2/100 + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma_i^2).$$

and you believe that the variables age and $\text{age}^2/100$ might influence the variance of the error term as in eq. (2).

a) Import the data set 'th1.csv' into Matlab which is a subset of the Jochmann & Pohlmeier (2004) study. You can find the variable names in the header column. 0.5 P

b) Initially, you assume your model to be homoscedastic. Report the OLS estimates of $\beta$, homoscedastic standard errors and, in addition, robust standard errors computed using four different heteroscedasticity robust variance-covariance estimators that have been proposed in the literature (see Hansen (2015), Section 4.11, for an overview):

$$\widehat{V}^{W}_{\widehat{\beta}} = (X'X)^{-1} X'D(e)X (X'X)^{-1} \tag{White}$$

$$\widehat{V}_{\widehat{\beta}} = \frac{n}{n-k} (X'X)^{-1} X'D(e)X (X'X)^{-1} \tag{Scaled White}$$

$$\tilde{V}_{\widehat{\beta}} = (X'X)^{-1} X'D(\tilde{e})X (X'X)^{-1} \tag{Andrews}$$

$$\bar{V}_{\widehat{\beta}} = (X'X)^{-1} X'D(\bar{e})X (X'X)^{-1} \tag{Horn-Horn-Duncan}$$

$e = (e_1, \ldots, e_n)'$ is a vector of residuals and $D(e)$ is a diagonal matrix with the squared values of $e_i$ in the diagonal elements. Furthermore, $\tilde{e}_i = (1 - h_{ii})^{-2} e_i$ and $\bar{e}_i = (1 - h_{ii})^{-1} e_i$ are rescaled residuals where $h_{ii}$ are the diagonal elements of the projection matrix $P = X(X'X)^{-1} X'$.

What do you conclude concerning the presence of heteroscedasticity and the four different variance-covariance estimators? 2.5 P

Next, you want to account for heteroscedasticity in the estimation process.

c) Derive the log likelihood function of the model. 0.5 P

d) Write a Matlab function for the log likelihood of the Gaussian regression model. The function should be structured as follows. 1.5 P

- Inputs: parameter vector $\theta = [\beta' \, \gamma']'$, vector of dependent data $Y$, regressor matrix $X$, instrument matrix $Z$ and an option to either return the individual negative log likelihood values or the sum of the negative log likelihoods.
- Output: $n \times 1$ vector of negative log likelihood values or the sum of the negative log likelihoods.

e) Use the fminunc command as well as your function to obtain the Maximum Likelihood estimate of the parameter vector, $\hat{\theta} = [\hat{\beta}' \, \hat{\gamma}']'$. Initialize the algorithm with the following starting values: $\beta^0 = \hat{\beta}_{\text{OLS}}$ and $\gamma^0$ is a vector of zeros. 1 P

f) Estimate the Gradient and the Hessian using the GradP and HessMp functions. Compute the variance-covariance matrix by the outer product of the gradient (OPG) and the sandwich form. Report coefficient estimates and standard errors for all parameters. Compare the ML estimates and standard errors of $\beta$ with the OLS estimates. 2 P

g) Test for the presence of heteroscedasticity in the data using a likelihood ratio test. 2 P

## References

Hansen, B. (2015). Econometrics. Available at: `http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf`.