Prof. Dr. Kai Carstensen
Prof. Dr. Uwe Jensen
Chair of Econometrics

Examination in Econometrics I

(Winter Term 2021/22)

Examination regulation

March 21, 2022, 12:00

Preliminary remarks:

1. Please read these instructions carefully!
2. Write your name and enrollment (matriculation) number on every sheet of paper!
3. Don't use a pencil!
4. The exam problems are listed on 3 pages. Check your exam for completeness!
5. **Round your solutions to 4 decimal places.**
6. For all tests use a significance level of 5%, if nothing else is specified.
7. You have 60 minutes in total to answer the exam questions.

Good luck!

Question 1 (23 points)

Consider the following model developed for analyzing individual emotional stability and its determinants.

$EmoStab$	= Emotional stability (0 = low, ..., 7 = high)
$Illness$	= Number of doctor visits in 3 months
$Illness2$	$= Illness^2/1000$
Age	= Age (in years)
$Age2$	$= Age^2/1000$
$LifeSat$	= Life satisfaction (0 = low, ..., 10 = high)
$NotEmpl$	= 1 for not employed
$LHhInc$	= Log of household income in euros
$LLabInc$	= Log of labor income in euros
$Height$	= Height (in cm)

It is known that the variation of emotional stability increases with increasing illness and decreases with increasing age.
Based on an individual cross-section data set, a random sample of size $N = 1274$, a LS estimation has led to the following results:

Variable	Coeff.	Robust std. err.
$Const$	1.1961	0.1003
$Illness$	-0.0382	0.0040
$Illness2$	0.5108	0.0940
Age	0.0213	0.0034
$Age2$	-0.1588	0.0325
$LifeSat$	0.1793	0.0060
$NotEmpl$	-0.1478	0.0251

1. Test the significance of the illness parameters separately.
2. Give reasons for using heteroskedasticity-robust standard errors in this example.
3. Shortly explain why measuring illness by number of doctor visits is a very rough approximation and what the resulting error has probably done with the illness parameters, as precisely as possible (sign of the effect).
4. Adding the variable $LHhInc$, another LS estimation with the same data has led to the following results:

Variable	Coeff.	Robust std. err.
$Const$	1.1163	0.1058
$Illness$	-0.0382	0.0040
$Illness2$	0.5111	0.0940
Age	0.0204	0.0034
$Age2$	-0.1417	0.0333
$LifeSat$	0.1791	0.0060
$NotEmpl$	-0.1114	0.0295
$LHhInc$	0.0079	0.0034

Shortly explain, as precisely as possible (sign of the effect), the change in the *NotEmpl* parameter from the first to the second table.

5. Using the estimated relation

$$EmoStab = \ldots + 0.0204\, Age - 0.0001417 Age^2 + \ldots,$$

derive mathematically whether the relation between age and emotional stability is monotonous for the individuals in the data set. Interpret your result.

6. Under which conditions can we interpret the age parameters as causal effect on emotional stability? Shortly explain why.

7. Can we interpret the $LifeSat$ parameter as causal effect on emotional stability? Shortly explain why (not).

8. A colleague is planning to use instrumental variables as remedies for the problems she has detected in the previous items. She proposes to use $LLabInc$ and/or $Height$ as instruments for $LifeSat$. Shortly discuss the advantages and disadvantages of this idea. Deal with the exogeneity and relevance of the instruments and the variance of the IV estimator.

   $Height$ probably exogenous but not highly correlated with $LifeSat$, i.e. not very relevant. High variance of IV estimator. Requires larger sample size than we have here for a satisfying result. (5 P.)

Question 2 (20 points)

1. Consider a linear regression model

$$y_i = \boldsymbol{x_i \beta} + \varepsilon_i$$

where the explanatory variables $\boldsymbol{x_i}$ are assumed to be orthogonal to the error term.

   (a) Set up the population orthogonality condition and the sample moment condition for GMM estimation with $g(\boldsymbol{w_i}, \boldsymbol{\theta_0}) = \boldsymbol{x}_i' \varepsilon_i$.

   (b) Show that in case of exact identification the GMM estimator is equivalent to the OLS estimator.

   (c) Find the asymptotic variance matrix $\mathbf{V}$ of the GMM estimator.

2. Given the population model $y = \beta_0 + \beta_1 x_1 x_2 + \beta_2 x_2^2$, assume that $E(x_1|x_2) = E(x_2|x_1) = 0$ and that x_1 and x_2 are independent. Derive the linear projection $L(y|1, x_2)$.

Question 3 (17 points)

Consider the linear regression model with scalar regressor x_i and parameters $\boldsymbol{\theta} = (\beta, \sigma^2)'$

$$y_i = x_i\beta + u_i, \qquad u_i \overset{iid}{\sim} N(0, \sigma^2).$$

Suppose you have a random sample of $i = 1, ..., N$ observations with $\sum_{i=1}^N x_i^2 = 40$ and $\sum_{i=1}^N y_i x_i = 28$. You want to test the hypothesis $H_0 : \beta = 0$ vs. $H_1 : \beta = 1$.

1. Write down the log likelihood function for observation i and for the full sample.
2. Derive the score with respect to $\boldsymbol{\theta} = (\beta, \sigma^2)$ for observation i.
3. Derive the Hessian $\boldsymbol{H_i(\theta)}$ and its conditional expectation $\boldsymbol{A(x_i, \theta)}$.
4. Derive the maximum likelihood estimators $\hat{\beta}$ and $\hat{\sigma}^2$.
5. Calculate the likelihood-ratio test statistic.
6. What is your test decision for the likelihood-ratio test? You can assume $\sigma^2 = 1$.