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

Examination in Econometrics I

(Winter Term 2021/22)

Examination regulation

February 7, 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!

Problem 1 (24 points)

Consider the following model developed for understanding the effect of local immigrant shares on natives' incomes in Germany, where 'local' stands for the same region and the same sector for natives and immigrants.

$LnInc$	=	Log gross hourly wages of German natives
Exp	=	Job experience of German natives (in years)
$Exp2$	=	Exp^2
$Educ$	=	Education of German natives (in years)
$LISls$	=	Local share of low-skilled immigrants
$LISls5$	=	Local share of low-skilled immigrants, 5 years before
$LIShs$	=	Local share of high-skilled immigrants
$Dist$	=	Distance between Germany and immigrants' home country (in km)

It is well-known that immigrants tend to go to regions with relatively high incomes and low unemployment. Furthermore, it is known that the variation of incomes increases with increasing education and increasing age. Finally, it is known that the correlation between $LISls$ and $LISls5$ is 0.99.

Based on an individual cross-section data set, a random sample of size $N = 1790$, a LS estimation has led to the following results:

Variable	Coeff.	Robust std. err.
$Const$	-0.124	0.1380
Exp	0.343	0.0381
$Exp2$	-0.085	0.0171
$Educ$	0.180	0.0181
$LISls$	-0.106	0.0642

1. Test the significance of the experience parameters separately.
2. Assuming that the estimation with centered data has yielded the same parameters, perform a Wald test for checking the joint significance of the experience parameters $(\beta_1, \beta_2)'$. Use
$$\widehat{Avar(\boldsymbol{\beta}_{(1,2)})} = \begin{pmatrix} 0.00145 & 0.00064 \\ 0.00064 & 0.00029 \end{pmatrix}.$$
3. Give reasons for using heteroskedasticity-robust standard errors in this example.
4. Shortly explain why measuring education in years is a very rough approximation and what the resulting error has probably done with the education parameter.

5. Adding the variable $LIShs$, another LS estimation with the same data has led to the following results:

Variable	Coeff.	Robust std. err.
$Const$	-0.126	0.1382
Exp	0.346	0.0383
$Exp2$	-0.087	0.0177
$Educ$	0.175	0.0189
$LISls$	-0.156	0.0509
$LIShs$	0.221	0.0995

   Shortly explain the change in the $LISls$ parameter from the first to the second table.

6. Can we interpret the LIS parameters as causal effects of local immigrant shares on natives' incomes? Can we expect to have obtained unbiased parameter estimates? Shortly explain why (not).

7. A colleague is planning to use instrumental variables as remedies for the problems he has detected in the previous item. He proposes to use $LISls5$ and/or $Dist$ as instruments for $LISls$. 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.

<u>Problem 2</u> (19 points)

Consider the exponential regression model where $E(y_i|x_i) = \mu(x_i) = e^{x_i\theta}$ and the distribution of y_i conditional on x_i is the exponential distribution with pdf

$$f(y_i|x_i, \theta) = \frac{1}{\mu(x_i)} \exp\left(-\frac{y_i}{\mu(x_i)}\right).$$

Remember that this distribution has the property that $Var(y_i|x_i) = [E(y_i|x_i)]^2$.

1. Write down the conditional log likelihood function for observation i and for the full sample.

2. Derive the score with respect to θ for observation i. Show directly that the score has conditional mean zero.

3. Find the Hessian with respect to θ.

4. Show directly that the conditional information matrix equality holds.

5. Find the asymptotic standard errors of $\hat{\theta}$.

Problem 3 (17 points)

Consider the following model:

$$y = \beta_0 + \beta_1 x_1 x_2 + u$$

where $E(u|x_1, x_2) = 0$. Further denote $E(x_1) = \mu_1$ and $E(x_2) = \mu_2$ as well as $Var(x_1) = \sigma^2$ and $E(x_1^3) = \mu_1^3 + 3\mu_1\sigma^2$. Additionally, x_1 and x_2 are independent.

1. Derive the partial effect and the average partial effect of x_1 on y.
2. Derive the linear projection $L(y|1, x_1)$ in terms of population moments of x_1. Compare to the result in 1.
3. Now imagine having a random sample from the true model. Can you consistently estimate β_0 and β_1 with OLS from the original model and/or the linear projection derived in 2.?