Written Exam for the B.Sc. or M.Sc. in Economics Summer 2015

Økonometri B/Econometrics B

Take-home exam

May 27-28, 2015

This exam consists of 9 pages in total.

Please note that the language used in your exam paper must correspond to the language of the title for which you registered during exam registration. That is, if you registered for the English title of the course, you must write your exam paper in English. Likewise, if you registered for the Danish title of the course or if you registered for the English title which was followed by 'eksamen på dansk' in brackets, you must write your exam paper in Danish.

If you are in doubt about which title you registered for, please see the print of your exam registration from the students' self-service system.

Focus on Exam Cheating
In case of presumed exam cheating, which is observed by either the examination registration of the respective study programmes, the invigilation or the course lecturer, the Head of Studies will make a preliminary inquiry into the matter, requesting a statement from the course lecturer and possibly the invigilation, too. Furthermore, the Head of Studies will interview the student. If the Head of Studies finds that there are reasonable grounds to suspect exam cheating, the issue will be reported to the Rector. In the course of the study and during examinations, the student is expected to conform to the rules and regulations governing academic integrity. Academic dishonesty includes falsification, plagiarism, failure to disclose information, and any other kind of misrepresentation of the student's own performance and results or assisting another student herewith. For example failure to indicate sources in written assignments is regarded as failure to disclose information. Attempts to cheat at examinations are dealt with in the same manner as exam cheating which has been carried through. In case of exam cheating, the following sanctions may be imposed by the Rector:

1. A warning
2. Expulsion from the examination
3. Suspension from the University for at limited period or permanent expulsion

The Faculty of Social Sciences
The Study and Examination Office
October 2006

Practical instructions for the take-home exam

Read entire exam before you respond. Answer every question in each problem. The exam consists of six problems in total.

The exam can be answered in groups of a **maximum of 4 students**. Hand-in a single report for the entire group **and specify each group member's contribution to the report**.

You must submit a comprehensive report with relevant tables and figures. The front page of the report must use the template available at `https://eksamen.ku.dk/`. Fill in the exam numbers of all group members on the front page. The second page of the template must specify which paragraphs and/or sections of the report is answered by which group member. This page may not contain other information.

Prepare one STATA do-file generating all tables and figures that appear in your report. The program must produce tables and figures in the same order as they appear in the report. Comments should clearly indicate which table or figure appearing in the report is being produced. Make sure that the do-file can be executed without any errors. The do-file must include the exam numbers of all group members.

The report must not exceed 15 (normal) pages. This includes the main text, tables and figures in the report, but not the front page and the list summarizing each group member's contribution to the report.

According to the Study Handbook for Economics[1], a normal page is defined as a text document with the following attributes:

- A4 format
- Font size set to 12
- Line spacing set to 1.5
- Margins (left/right/top/bottom) of at least 2.5 cm

The exam ends **May 28 at 16.00**. The report and the STATA do-file must be uploaded electronically no later than 16.00.

Uploading your report

Each group must hand-in only one report in total.

One student hands in the report by uploading it to University of Copenhagen's Digital Exam system and then adding the rest of the group members to the hand-in. Go to the website `https://eksamen.ku.dk/` and click on 'Log in as student'. Use your regular KU login and password to enter Digital Exam. Click on 'Econometrics B' in your assignments. On the page 'Information about the hand-in', you must add all other group members to the handed-in answer (if you are in a group). Click on 'Add member' and follow the instructions on Digital Exam to invite your fellow group members. Group members will be added to the handed-in answer as soon as they **accept** your invitation.

[1] `http://www.econ.ku.dk/polit/studieordningen/2008_BA_august_2014__endelig_.pdf`

Next, go to 'Upload hand-in' to upload your files. Each group must upload two files:

1. The report itself must be uploaded as a PDF file. The filename must start with the letter R followed by the exam numbers of all members of the group separated by _ ("underscore").
2. The STATA do-file must be uploaded as a file in plain text format (.txt). The filename must start with the letter P followed by the exam numbers of all members of the group separated by _ ("underscore").

Use the same combination of exam numbers for both files.

> Example: A group of four members with exam numbers 72, 82, 87 and 174 will submit the following files:
>
> 1. R_72_82_87_174.pdf
> 2. P_72_82_87_174.txt

If needed, a free PDF converter is available at `www.pdf995.com`.

If you have problems accessing the Digital Exam system at the deadline of the take-home exam or if you have difficulties with the upload function you must e-mail your answer to `samf-fak@samf.ku.dk` by 16.30. Handing in your exam answer by e-mail requires that you describe the problems and provide screen dumps that document this.

Access to data

For the take-home exam, there are several data sets available on the Digital Exam website (`https://eksamen.ku.dk/`). Follow the instructions below to pick the correct data set for your group:

1. Determine the **lowest** number among the exam numbers of the group members. Use the **last** digit of the **lowest** exam number as your "group number".

   > Example: A group of four members with exam numbers 72, 82, 87 and 174 will have "2" as the last digit of the lowest exam number.

2. Download the STATA file groupdataX.dta from the Digital Exam website, where X is equal to the group number.

   > Example: The group from before downloads groupdata2.dta from the Digital Exam website.

3. Download the data to your computer.
4. Open the data in STATA and execute the **describe** command to ensure the data appears operative.

If you have trouble selecting or opening the data, you can contact Rasmus Jørgensen on telephone 3532 3075 during the period 10.00 to 12.00 on May 27.

After this, no additional help will be provided for the exam.

Introduction to the assignment:

"Children and their Mom's Labor Supply"

An understanding of the relationship between fertility and labor supply is important for a number of reasons. First, the link between childbearing and labor supply might partly explain the postwar increase in women's labor-force participation rates if having fewer children causes an increase in labor-force attachment. Second, fertility-induced changes in female labor supply has important implications for many other phenomena, including marriage, fertility, divorce, child care, the distribution of family earnings and male-female wage differentials.

Not surprisingly, given the wide and long-standing interest in the link between child bearing and labor supply, hundreds of empirical studies report estimates of this relationship. The vast majority of these studies find a negative correlation between fertility (or family size) and female labor supply. As noted by many researchers, however, the interpretation of these correlations remain unclear.

In a survey of the 'Economics of the Family', Willis (1987) writes "... it has proven difficult to find enough well-measured exogenous variables to permit cause and effect relationships to be extracted from correlations among factors such as the delay of marriage, decline of childbearing, growth of divorce, and increased female labor force participation.[2] Browning (1992) expresses similar views: "... although we have a number of robust correlations, there are very few credible inferences that can be drawn from them."[3]

Skepticism regarding the causal interpretation of associations between fertility and labor supply arises in part from the fact that there are strong theoretical reasons to believe that fertility and labor supply are jointly determined. In fact, labor economists often consider the relationship between work hours and fertility, while demographers and others aim to characterize the impact of wages on fertility. Since fertility variables cannot be both dependent and exogenous at the same time, it seems unlikely that either sort of regression has a causal interpretation.

In a seminal article, Angrist and Evans (1998) use parental preferences for a mixed sibling-sex composition to construct instrumental variables (IV) estimates of the causal effect of fertility on labor supply.[4] Their identification strategy exploits the fact that parents with two kids of the same sex are significantly and substantially more likely to go on and have an additional child. Because sex mix is virtually randomly assigned, a dummy for whether the sex of the second child matches the sex of the first child provides a plausible instrument for further childbearing among women with at least two children.

To investigate the link between fertility and labor supply, we will consider data on American women in 1990, containing detailed information about their labor and family characteristics.

[2]Robert J. Willis (1987): "What Have We Learned from the Economics of the Family?", *American Economic Review*, 77(2), pp. 68-81.

[3]Martin Browning (1992): "Children and Household Economic Behavior", *Journal of Economic Literature*, 30(3), pp. 1434–75.

[4]Joshua D. Angrist and William N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", *American Economic Review*, 88(3), pp. 450-477.

Documentation of the data

The data set consists of 337,500 observations from a representative sample of women aged 21–35 with at least two children in the United States in 1990[5]. For each individual, we observe selected information related to their family and work situation. The following variables are available:

Table 1: List of variables

Variable name	Description
WORKM	Dummy, indicating if the mother works or not
WEEKM	Weeks worked during the year
HOURM	Hours worked per week
LINCM	Labor income in US dollars
FAMINC	Family (joint) income in US dollars
AGEM	Age of mother
AGEM1STKID	Age of mother at first birth
NKIDS	Number of kids born to mother
MOREKIDS	Dummy, indicating if mother had more than two children
BOY1STKID	Dummy, indicating if firstborn is a boy
BOY2NDKID	Dummy, indicating if secondborn is a boy
TWOBOYS	Dummy, indicating if both firstborn and secondborn are boys
TWOGIRLS	Dummy, indicating if both firstborn and secondborn are girls
SAMESEX	Dummy, indicating if firstborn and secondborn have same sex
BLACK	Dummy, indicating mother's race (reference category: White)
HISP	Dummy, indicating mother's race (reference category: White)
OTHER	Dummy, indicating mother's race (reference category: White)
LOWEDU	Dummy, indicating if mother's education is less than high school
HSEDU	Dummy, indicating if mother is a high-school graduate
HIGHEDU	Dummy, indicating if mother's education is more than high school

[5]The data used for this exam is an edited version of the original data used by Angrist and Evans (1998).

Problem 1 (10%):

1. Describe the variables listed in table 1. Provide one or more tables that present relevant characteristics for each variable. Include a short discussion of the table(s).

2. Provide a table that compares mothers with $SAMESEX = 1$ to mothers with $SAMESEX = 0$ in terms of relevant characteristics. Comment briefly on the table.

Problem 2 (20%):

1. Consider the regression model in (1), where $y_i = \{workm_i, weekm_i, hourm_i, lincm_i, log(faminc_i)\}$ and u_i is an unobserved error term:

$$y_i = \beta_0 + \beta_1 morekids_i + \beta_2 boy1stkid_i + \beta_3 boy2ndkid_i + \beta_4 agem_i + \beta_5 agem1stkid_i + \beta_6 black_i + \beta_7 hisp_i + \beta_8 other_i + u_i \quad (1)$$

   (a) What is the interpretation of β_1? What is the expected sign of β_1?

   (b) What is the interpretation of β_4 and β_5?

2. Estimate the parameters of model (1) by OLS using work status ($workm$), weeks worked ($weekm$), hours worked ($hourm$), labor income ($lincm$) and log family income ($log(faminc)$) as dependent variables. Report your estimates in a table with your preferred choice of standard errors or t statistics. Based on your OLS estimation results, comment on the relationship between having more kids and the mothers' labor market outcomes.

3. Does the ordering of children's gender matter for the labor market outcomes of mothers' with two children of opposite sex? That is, do mothers with a firstborn girl and secondborn boy have different labor market outcomes compared to mothers with a firstborn boy and a second born girl? Explain how this hypothesis can be tested on the basis of model (1) and OLS estimation. Write up the null hypothesis and specify the alternative that you are testing against. Explain what test statistic is used and why. What is your conclusion?

4. *morekids* is likely to be an endogenous variable in model (1). Provide an example of why and how *morekids* may be an endogenous variable and discuss how the OLS estimates of the parameters in model (1) may be biased using your example.

Problem 3 (20%):

You may disregard any heteroskedasticity of the error term in your answer to Problem 3.

1. Discuss the conditions that must be satisfied for *samesex* to a relevant and valid instrument for *morekids* in model (1). Are the conditions likely to be satisfied in this case? Present empirical evidence as needed to support your answer.

2. Estimate model (1) by IV using *samesex* as an instrument for *morekids*. Report your results in a table and discuss how they compare to your OLS results from Problem 2.

3. An advantage of the *samesex*–instrument is that it can be decomposed into two separate instruments: *twoboys* and *twogirls*. Estimate model (1) without the $boy2ndkid_i$ dummy variable using *twoboys* and *twogirls* as instruments for *morekids*. Comment briefly on your results and relate them to earlier findings. Why do you have to exclude $boy2ndkid_i$ when using the two separate instruments?

4. Explain what is meant by overidentifying restrictions and perform a test for it.

Problem 4 (20%):

Assuming that relevant and valid instruments are available, the asymptotic variance of the IV estimator in the case of homoskedastic errors may be estimated using:

$$Var\left(\widehat{\beta}\right) = \widehat{\sigma}^2 \left(\widehat{X}'\widehat{X}\right)^{-1} \tag{2}$$

where $\widehat{\sigma}^2 = (n-k)^{-1}\sum_{i=1}^{n}\widehat{u}_i^2$ and $\widehat{u}_i$ is the IV residual. The matrix $\widehat{X}$ includes the predicted endogenous variables from the first stage regression(s). Specifically, it is a $n\times(k+1)$ matrix defined as $\widehat{X} = Z(Z'Z)^{-1}Z'X$. The asymptotic standard error of $\widehat{\beta}_j$ is just the square root of the jth diagonal element of the matrix expression in (2).

In the presence of heteroskedasticity, the variance matrix estimator in (2) is no longer valid. It is, however, possible to construct an estimator that is robust towards arbitrary heteroskedasticity. An estimator of the asymptotic variance of the IV estimator in the case of heteroskedastic errors is:

$$Var\left(\widehat{\beta}\right) = \frac{N}{N-K}\left(\widehat{X}'\widehat{X}\right)^{-1}\left(\sum_{i=1}^{n}\widehat{u}_i^2\widehat{x}_i'\widehat{x}_i\right)\left(\widehat{X}'\widehat{X}\right)^{-1} \tag{3}$$

As before, the square roots of the diagonal elements of the matrix expression in (3) are the heteroskedasticity-robust standard errors for the IV estimator.

To implement heteroskedasticity-robust inference for the IV estimator in STATA, one should use the **robust**–option in connection with the **ivregress** procedure. E.g., the following STATA code will calculate heteroskedasticity-robust standard errors and t statistics along with the usual IV parameter estimates:

```
ivregress 2sls y (x=z) ..., small robust
```

In this problem, you are first asked to compare the two estimators in (2) and (3) in a simulation experiment (also known as a Monte Carlo experiment). Afterwards, you are asked to examine whether your results from Problem 3 are sensitive towards heteroskedasticity or not.

For the first part, consider the following data generating process (DGP):

$$y_i = \delta_0 + \delta_1 x_i + u_i \tag{4}$$

$$x_i = \chi_{1i} + \chi_{2i} \tag{5}$$

$$u_i = \chi_{2i} + \rho\chi_{1i}^2 + \eta_i \tag{6}$$

$$z_i = \chi_{1i} + \epsilon_i \tag{7}$$

$$\delta_0 = 1, \delta_1 = 2, \rho = 5 \tag{8}$$

$$\chi_{1i} \sim N(0,1),\ \chi_{2i} \sim N(0,1),\ \eta_i \sim N(0,1), \epsilon \sim N(0,1) \tag{9}$$

1. Based on the DGP defined in equations (4) through (9), discuss if and why:
   - x is an endogenous variable
   - u is a heteroskedastic error term
   - z is a relevant and valid instrumental variable for x

2. Implement the simulation experiment of the estimators (2) and (3) in STATA. Pick a seed number of your choice and write it explicitly in the main text of your report. Draw samples of size 5,000 and replicate the experiment 10,000 times (that is, $n =$ 5,000 and $R =$ 10,000). For each replication, calculate the IV parameter estimates and the IV standard errors according to (2) and (3) and store these calculations for later analysis.

   Hint: The following STATA code may be helpful for storing the parameter estimates and standard errors from the **ivregress** procedure:

```
ivregress 2sls y (x=z) ..., small robust
return scalar deltaiv=_b[x]
return scalar seiv=_se[x]
```

   Once your simulation experiment is completed, you should have 10,000 IV estimates of δ_1, 10,000 estimates of the conventional IV standard errors, and 10,000 estimates of the heteroskedasticity-robust IV standard errors.

   Calculate averages for the two methods outlined in (2) and (3) using all 10,000 estimated standard errors and compare the averages to the calculated standard deviation of the 10,000 IV parameter estimates. Discuss why this is a relevant comparison and what conclusions can be drawn from it.

   Based on this particular experiment, in what direction are the conventional IV standard errors, as defined in (2), biased? And how will the bias in the conventional standard errors affect the statistical test of the hypothesis $H_0 : \delta_1 = 0$?

3. Using your preferred IV specification from Problem 3, are you earlier conclusions sensitive to heteroskedasticity? Calculate the heteroskedasticity-robust standard errors according to (3) and compare your results to earlier findings.

Problem 5 (20%):

It is often claimed that the labor supply of less educated women is more sensitive to the presence of children than the labor supply of more educated women. In this problem, you are asked to investigate this hypothesis.

1. Divide the data into three sub-samples: One sample for women with less than a high-school education, one sample with high-school graduates, and finally, one sample for women with more than a high-school education.

   Estimate model (1) by IV for each of the three sub-samples using your preferred IV specification from Problem 3. Is the impact of more children on labor market outcomes different for women with different levels of education? Include a short discussion of your empirical results.

2. In order to test for differences in parameters across groups, it is often convenient to allow parameters to differ across groups within the same regression model. Estimate an extended version of model (1) where the effect of *morekids* is allowed to vary across the three educational groups defined above. Calculate IV estimates using your preferred instruments from Problem 3. How many endogenous variables does the extended model include? Motivate your choice of instruments. Perform a test of the hypothesis that fertility has a different effect on labor supply for women with different educational attainment.

3. Compare your IV estimates from Problems 5.1 and 5.2. Discuss the assumptions underlying the two estimation approaches. Which specification is the preferred one?

Problem 6 (10%):

Produce a table summarizing your results regarding the relationship between fertility and labor market outcomes based on your answers to Problems 2–5. Comment briefly on the table and on how the various models relate to each other.

Which specification is your preferred econometric model and why? What do you conclude about fertility and labor supply from the econometric analysis?