NTNU
Norwegian University of
Science and Technology

Department of Economics

Examination paper for SØK3001 – Econometrics I

Academic contact during examination: Costanza Biavaschi
Phone: 73 59 16 54

Examination date:	04.12.2018
Examination time (from-to):	5 hours (09.00 -14.00)
Censorship date:	07.01.2019

Permitted examination support material: C
Formelsamling:
Knut Sydsæter, Arne Strøm og Peter Berck (2006): Matematisk formelsamling for økonomer, 4utg. Gyldendal akademiske.
Knut Sydsæter, Arne Strøm, og Peter Berck (2005): Economists' mathematical manual, Berlin.
Calculator:
Casio fx-82ES PLUS, Casio fx-82EX Citizen SR-270x, SR-270X College or HP 30S.

Language:	English
Number of pages (front page excluded):	6
Number of pages enclosed:	8

Informasjon om trykking av eksamensoppgave
Originalen er:

1-sidig □ **2-sidig** □

sort/hvit □ **farger** □

skal ha flervalgskjema □

Checked by:

Date Signature

1 Exercise 1 (40 points)

Standard economic theory predicts that the demand for children is influenced by the cost of raising children. Holding other things constant, a decrease in the cost of raising children should lead to an increase in the demand for children. Whittington, Alm, and Peters (1990) provide evidence for this relationship, exploiting the fact that between 1913 and 1984 the value of child tax benefits in the U.S. increased substantially relative to estimates of the cost of raising children. Whittington, Alm, and Peters (1990) claim a large positive effect of child tax benefits on fertility using time series methods. Their key conclusion is based on the following equation, estimated for the period 1913 to 1984:

$$\begin{aligned}\text{Fertility Rate}_t = & \beta_0 + \beta_1\text{Personal Exemption}_t + \beta_2\text{Male and Asset Income}_t + \\ & + \beta_3\text{Unemployment}_t + \beta_4\text{Infant Mortality}_t + \beta_5\text{Immigration}_t + \\ & + \beta_6\text{Female Wage}_t + \beta_6\text{Pill}_t + \beta_8\text{WW2}_t + \beta_9\text{Time Trend}_t + u_t.\end{aligned} \tag{1}$$

Here Fertility Rate_t measures the number of children born per 1,000 women; Personal Exemption_t is the dollar value of the personal tax exemption, that is, the dollar amount that a resident taxpayer is entitled to claim as a tax deduction in the presence of dependent children; Male and Asset Income_t is the dollar value of personal income per family, net of female earnings; Unemployment_t measures the share of people who are unemployed; Infant Mortality_t measures the number of children who die per 1,000 live births; Immigration_t measures the share of people who are foreign born; Female Wage_t is the dollar value of after-tax female wage; Pill_t is a dummy variable that equals one in years 1963-1984, when birth control became widely available; WW2_t is a dummy variable that equals one in years during which the US was in World War II and Time Trend_t is a time trend equal to one in 1913 and increasing by one unit each year.

In this exercise you are asked to revisit this question, discussing and interpreting the findings in Whittington, Alm, and Peters (1990) and the critique provided in Goda and Mumford (2010). All relevant results are reported in Table 1, on page 3.

(a) Present and perform a test that checks whether the use of the FGLS estimator in column (2) is empirically justified.

(b) Zhang et al. (1994) mention that there is a concern that some series in the Whittington et al. (1990)'s study may be non-stationary. Using results in Table 1, present and perform a test for non-stationarity in the fertility rate.

(c) Using and justifying your preferred specification, discuss the relationship between tax benefits and fertility.

In a recent paper, Mumford and Thomas (2016) address the same problem of the relationship between tax incentives and fertility using a different framework. The authors use a sample of women from the Panel Study of Income Dynamics (PSID) surveyed from 1985 to 2011 in the United States. The PSID is a longitudinal data set that began with a representative set of households in 1968 and followed these households, their descendants, and refresher samples. They restrict the sample to women between the ages of 20 and 44. The authors then exploit personal income tax changes that occur at the *U.S. State* level. The following equation represents the main linear regression specification of the paper:

$$\text{N. Children}_{ist} = \beta_0 + \beta_1 \text{Tax Subsidy}_{st} + \gamma X_{ist} + \tau_t + \eta_s + u_t, \qquad (2)$$

where N. Children$_{ist}$ measures the number of children born to individual i in state s at time t, Tax Subsidy$_{st}$ measures the value of the tax subsidy given in state s at time t, X_{ist} is a vector of relevant individual characteristics, τ_t represents year dummies and η_s represents state dummies.

(d) Explain what restrictions, if any, are imposed on the coefficients across states and over time in the estimation of equation 2.

(e) Compare now the model in equation (1) with the model in equation (2). Suppose that in both cases an OLS estimator is used for estimating the parameters of the two models. Discuss under which conditions each model is able to identify the causal impact of tax subsidies on fertility.

Table 1: Child Tax Benefits and Fertility

	OLS	FGLS	First Difference	First Difference	OLS	OLS
	Fertility Rate$_t$	Fertility Rate$_t$	ΔFertility Rate$_t$	ΔFertility Rate$_t$	$\hat{u}_t$ col(4)	$\hat{u}_t$ (col 1)
	(1)	(2)	(3)	(4)	(5)	(6)
Fertility Rate$_{t-1}$	-	-	-0.022	-	-	-
	-	-	(0.0260)	-	-	-
Personal Exemption$_t$	0.178	0.121	-	-0.084	-	-
	(0.0977)	(0.0446)	-	(0.042)	-	-
Male and Asset Income	0.0035	-0.0004	-	-0.003	-	-
	(0.0031)	(0.0027)	-	(0.002)	-	-
Unemployment	-68.12	-73.43	-	-20.985	-	-
	(25.818)	(34.20)	-	(31.280)	-	-
Infant Mortality	0.393	0.083	-	-0.042	-	-
	(0.321)	(0.255)	-	(0.315)	-	-
Immigration	964.13	774.24	-	68.878	-	-
	(329.44)	(311.31)	-	(119.073)	-	-
Female Wage	15.427	5.647	-	7.472	-	-
	(5.286)	(15.686)	-	(5.792)	-	-
Pill	-25.383	-10.856	-	-1.91	-	-
	(11.961)	(6.126)	-	(1.020)	-	-
WW II	-29.419	-17.223	-	5.138	-	-
	(8.057)	(4.989)	-	(3.377)	-	-
Time Trend	-0.843	-0.539	-	-	-	-
	(0.543)	(0.538)	-	-	-	-
$\hat{u}_{t-1}$	-	-	-	-	0.048	0.48
	-	-	-	-	(0.0322)	(0.0322)
Intercept	55.944	102.979	1.3049	-0.618	0.0499	0.078
	(25.831)	(24.666)	(2.5488)	(0.954)	(0.5837)	(0.5122)
Observations	72	71	71	71	68	70
R^2	0.829	0.916	0.829	0.203	0.829	0.829

In columns (1)-(2), the dependent variable is the fertility rate at time *t*, which measures the number of children born per 1,000 women; in columns (3)-(4) the dependent variable is the change in the fertility rate; in column (5) the dependent variables is the residuals from column (4), while in column (6) the dependent variable is the residuals from column (1). *Personal Exemption* is the dollar value of the personal tax exemption, that is the dollar amount that a resident taxpayer is entitled to claim as a tax deduction if one has dependent children; *Male and Asset Income* is the dollar value of personal income per family net of female earnings; *Unemployment* measures the share of people who are unemployed; *Infant Mortality* measures the number of children who die per 1,000 live births; *Immigration* measures the share of people who are foreign born; *Female Wage* is the dollar value of after tax female wage; *Pill* is a dummy variable that equals one in years 1963-1984; *WW2* is a dummy variable that equals one in years during which the US was in World War II and *Time Trend* is a time trend equal to one in 1913 and increasing by one unit each year.

2 Exercise 2 (40 points)

Class size is an extremely popular education reform among including students, parents, teachers, school administrators, and educationalists. With such broad appeal, reducing class size is also popular among policymakers. Intuitively, students in smaller classes should have better learning outcomes than students in larger classes-for example, the teacher can provide more individualized attention in smaller classes, and classroom discipline is easier with fewer students. At the same time, reducing class size is an expensive education policy.

In this exercise you are asked to revisit the evidence on the relationship between test scores and student achievement. The question is based on the findings in Angrist and Lavy (1999). Angrist and Lavy (1999) use data on test score from a national testing program administrated in Israeli primary schools at the end of the 1990-1991 and 1991-1992 academic years. The question below is based on their findings for the academic year 1991 and focuses on the results reported for reading skills of fourth graders. Data on class sizes came from administrative sources and were collected between March and June of the school year starting in September. The unit of observation is the class c in school s. Average Reading scores for each class were linked with data on school characteristics and class size from the administrative sources. Specifically, the linked class-level data sets include information on average test scores, scaled from 1 to 100, in each class c in school s (Test Score_{cs}), the class size of class c in school s (Class Size_{cs}), the fraction of students in the school who come from what is defined to be as disadvantaged background (Percent Disadvantaged_s), and beginning-of-the-year enrollment in the school s (Enrollment_s).

In other words, the authors estimate several versions of the following simplified model:

$$\text{Test Score}_{cs} = \beta_0 + \beta_1 \text{Class Size}_{cs} + \beta_2 \text{Percent Disadvantaged}_s + \beta_3 \text{Enrollment}_s + u_{cs}. \qquad (3)$$

(a) Consider first the model in column (1) and (2) of Table 2. Explain why the estimated effect of class size on student achievement changes between column (1) and column (2).

(b) Consider first the model in column (3) of Table 2. Discuss why this model might not pin down the true relationship between student achievement and class size.

(c) Angrist and Lavy complement the basic analysis reported in column (1) and (2) with an instrumental variable estimation. They derive their instrumental variable from a rule that governs classroom size in Israeli schools. The rule works in the following way: the classroom size assigned to class c in school s, denoted by f_{cs}, equals the number of students enrolled (Enrollment_s) if such number is less or equal 40; when the enrollment exceeds 40, however, classes are split in half. In other words:

$$f_{cs} = \begin{cases} \text{Enrollment}_s & \text{if Enrollment}_s \leq 40 \\ \frac{\text{Enrollment}_s}{2} & \text{if Enrollment}_s > 40 \end{cases}$$

Hence, for example, if the number of students in class c of school s in a given year is equal to 40, then the classroom size would be 40; whereas if however the number of students in that year is equal to 41, then the classroom size of the two separate classes would be 20 and 21. Note that, although f_{cs} is fixed within schools, in practice enrollment cohorts are not necessarily divided into classes of equal size (as shown in the example above).

Discuss the conditions needed for the use of the variable f_{cs}, the number of students enrolled in that grade.

(d) Consider now the results reported in column (5). A commentator suggests that these effects are not credible estimates of the causal impact of class size on student achievement because the regression omits indicator variables for each school in the sample. Explain whether you agree or disagree with this suggestion.

(e) Interpret all the coefficients on the class size variable. Are you concerned about the big changes in magnitude between the OLS and the IV estimates?

Table 2: The Impact of Class Size on Test Scores

	OLS	OLS	OLS	OLS	IV
	Test $Score_{cs}$	Test $Score_{cs}$	Test $Score_{cs}$	Class $Size_{cs}$	Test $Score_{cs}$
	(1)	(2)	(3)	(4)	(5)
f_{cs}	-	-	-	0.542	-
	-	-	-	(0.027)	-
Class $Size_{cs}$	0.221	-0.031	-0.025	-	-0.275
	(0.039)	(0.026)	(0.031)	-	(0.066)
Percent $Disadvantaged_s$	-	-0.35	-0.351	-0.053	-0.369
	-	(0.013)	(0.013)	(0.009)	(0.014)
$Enrollment_s$	-	-	-0.002	0.043	0.022
	-	-	(0.006)	(0.005)	(0.009)
$R2$	0.036	0.369	0.369	0.553	-
Observations	2,019	2,019	2,019	2,019	2,019

The dependent variables are reported on the top of each column. Test $Score_{cs}$ measures average test scores in each class c in school s, and varies from 1 to 100. Class $Size_{cs}$ measures the number of pupils in each class. Percent $Disadvantaged_s$ indicates the fraction of students in the school who come from a disadvantaged background. $Enrollment_s$ measures beginning-of-the-year enrollment in the school s. f_{cs} is defined in the text of the exercise.

3 Exercise 3 (20 points)

Josh Angrist and Steve Pischke in their popular book *Mostly Harmless Econometrics* seem to be strong supporters of linear probability models, preferring them, for their simplicity of interpretation, to limited dependent variable models such as Probit or Logit.

(a) Discuss whether you agree or disagree with their opinion.

(b) Consider a variable y_i which is binary and generated according to $y_i = \beta_0 + \beta_1 x_i + u_i$. Here u_i has a chi-squared distribution with 3 degrees of freedom. Explain how you would empirically compare the performance of the OLS estimator with that of a Probit estimator, when both are used to estimate β_0 and β_1.