Roughly Outlined Solutions for SØK 3001 Fall Exam 2018

Costanza Biavaschi

SØK 3001 - Fall 2018

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 4.

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

Solution. This question asks to discuss issues of serial correlation. The estimation in column (2) employs a Feasible Generalized Least Squares estimator. This is motivated by a stated concern about serial correlation in the error term u_t in equation (1).

Consider the baseline model of equation (1). Now suppose that:

$$u_t = \rho u_{t-1} + e_t \quad \text{with } e_t \quad iid$$

The model in column (1) imposes $\rho = 0$. Under this condition and the additional assumption TS1-TS6 (should be explained), the OLS estimator would be unbiased and consistent.

In the presence of serial correlation (i.e. $\rho \neq 0$), the OLS is no longer BLUE, the OLS standard errors and test statistic are not valid even asymptotically.

It is always advisable to test for serial correlation (chapter 12.2a) - note, we have not discussed in class the Durbin Watson test, so a discussion of that test is not expected.

A test for serial correlation assumes that $E(e_t|u_{t-1}, u_{t-2}, \ldots) = 0$ and $V(e_t|u_{t-1}) = \sigma_e^2$. The hypothesis are:

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

Under the null hypothesis $\{u_t\}$ are weakly dependent so we could use a standard t-test. However, the error term is unobserved. Hence, in a first step we need to estimate the main model and obtain residuals, then we can regress the residuals on their one-period lag. The results are provided in column (8) and a t-test on $\hat{u}_{t-1}$ shows that serial correlation is present.

Finding serial correlation motivates the use of the FGLS.

(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.

Solution. This question tests a core concept in time series: unit roots. Relevant material is in chapter 11.3 and 10.5. Students should highlight the difference between trends and unit roots. It is extremely important not to confuse trending and highly persistent behaviors. A series can be trending but not highly persistent and can be highly persistent and trending. In addition, if the series are I(1) testing on a time trend would be invalid. In fact, suppose that:

$$\text{Fertility Rate}_t = \beta_0 + \beta_1 \text{Personal Exemption}_t + \ldots + \text{Fertility Rate}_{t-1} + u_t$$

For the sake of the argument, let's disregard all the observable characteristics.

$$E(\text{Fertility Rate}_t) = E(\text{Fertility Rate}_{1913})$$
$$V(\text{Fertility Rate}_t) = t\sigma_u^2$$

The variance of a random walk increases as a linear function of time. This process cannot be stationary and violates the key assumptions of the CLM.

In seminal work by Granger and Newbold on spurious regressions, the authors argue that annual macro series, like those used in this study, are almost always I(1); thus, regressions involving the levels will be misleading, suggesting relationships when there may be none.

A unit root test, rather than a test on the trend, seems necessary. This is a standard DF test. To perform it, a t-statistic on the lagged personal exemption can be constructed from column (6). The statistic is almost zero so we will reject the null and conclude that there is a unit root.

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

Solution. This tests ability to interpret the results. Given that no serial correlation is presented in column 5, as one can test using column 7, the preferred model is 4. Here the coefficient can be explained and its statistical significance should be discussed.

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:

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

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.

Solution. All coefficients are the same except we alow fertility to be permanently diferent across states and also we allow year shocks.

(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.

Solution. This tests knowledge of basic identifying assumptions. Students can list the assumptions of the OLS in a time series vs panel framework. Focus should be on discussing the conditional expectation of the error term.

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)
N	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:

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

(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).

Solution. Basic interpretation of simple versus multiple linear regression models/omitted variable bias. Column (1) shows the row correlation between the two variables. Part of the effect of Class Size goes through its relationship with socioeconomic status and enrollment. The MLR partials out the effect of this variables. In other word:

$E(\hat{\beta}_1) = \beta_1 + \beta_2\tilde{\delta}_1$,

where $\delta_1$is the partial correlation between class size and percentage of disadvantaged children. As soon as $\beta_2 \neq 0$ and $\delta_1 \neq 0$ the estimated coefficient on class size will not equal the population parameter.

(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.

Solution. This question tests the understanding of OLS assumptions and whether students can critically evaluate an empirical problem. The material is covered in chapter 15.

- Simultaneity bias: if a school provides smaller classes for its most "at-risk" students, the result would be higher achievement in the larger classes, again for reasons unrelated to class size.
- The presence of unobserved heterogeneity across schools: for example, in the US, where the biggest source of funding is local property tax revenue, schools in wealthier areas are more likely to have smaller classes and higher achievement due to students' more advantaged backgrounds, rather than being a casual effect of smaller classes.
- The presence of unobserved student ability: schools can base class size decisions, as well as the assignment of specific students to classes of different sizes, on prior student achievement.

(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:

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

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 the use of the variable N, the number of students enrolled in that grade.

Solution. This question requires students to think of validity and relevance of the instrument. Since f_{sc} is a deterministic function of Enrollment_s, and Enrollment_s is almost certainly related to pupil test scores for reasons other than effects of changing class size, the key identifying assumption is that any other effects of Enrollment_s on test scores are partialled out of the instrument by the term Enrollment_s.

To assess the plausibility of this assumption, it helps to consider why Enrollment_s is related to test scores in the first place. One reason could be that places with higher enrollment are inversely related to socioeconomic status. Also, better schools might face increased demand if parents selectively choose districts on the basis of school quality. On the other hand, more educated parents might try to avoid large-enrollment schools they perceive to be overcrowded. Any of these effects seem likely to be smooth, however; whereas the variation in test scores with enrollment has a rough up-and-down pattern that mirrors Maimonides' rule. Nevertheless, it remains an untestable identifying assumption that nonclass-size effects on test scores do not depend on enrollment except through the smooth functions included in Xs.

A final identifying assumption is that parents do not selectively exploit Maimonides' rule so as to place their children in schools with small classes. Selective manipulation could occur if more-educated parents successfully place children in schools with grade enrollments of 41-45, knowing that this will lead to smaller classes in a particular grade. In practice, however, there is no way to know whether a predicted enrollment of 41 will not decline to 38 by the time school starts, obviating the need for two small classes in the relevant grade. And even if there was a way to predict this accurately, we noted earlier that parents are not free to transfer children from one elementary school to another except by moving. Of course, parents who discover they got a bad draw in the enrollment lottery (e.g., enrollment of 38 instead of 41) might then elect to pull their kids out of the public school system entirely. Private elementary schooling is rare in Israel outside of the ultra-orthodox community. Nevertheless, for this reason, we define fsc as a function of September enrollment and not enrollment at the time testing was done, even though the latter is more highly correlated with class size.

(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.

Solution. This question requires students to understand the role of fixed effects/collinearity. In principle the commentator is correct. However, such fixed effects would be collinear with the enrollment variable. They would not however be collinear with the instrument as class size varies within school. An analysis similar to the one in column (5) can therefore not be performed with school fixed effects but one would not observe the direct impact of enrollment.

Such variable does not vary within school so it is already captured by the school fixed effects in column (2). Such concern would be valid if raised for column (1), assuming principals with longer experience make different choices in terms of pupils allocation to large/small classes.

(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?

Solution. This question tests some basic ability to interpret regression results. Interpretation is easy. Students need to point to the fact that the IV shows the expected positive relationship. However, the point estimates is also 10 times bigger than the OLS. Recall that

$$\hat{\beta}_1^{IV} = \beta_1 + \frac{Cov(f_{cs}, u)}{Cov(f_{cs}, \text{Class Size})}$$

If $Cov(f_{cs}, \text{Class Size})$ is close to zero, the bias in the IV estimates might be larger than in the OLS. A t-test on the first stage however seem to suggest that the instrument is sufficiently strong. It remains questionable whether it is sufficently exogenous....

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)
R^2	0.036	0.369	0.369	0.553	-
N	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.

Solution. This question requires students to understand the basics of LPM vs Probit/logit models. Marginal effects beyond 0-1 and heteroskedasticity inherent to the model are the main disadvantages of LPM. LDV models are on the other hand harder to interpret, as marginal effects are not directly estimated but need to be computed. They also rely on distributional assumptions about the process generating the error term. Not expected: in the case of normality, as the vast majority of the observations fall in the "mid region" of the distribution, the two methods will indeed give very similar results.

(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.

Solution. This question hints to the Monte Carlo studies that we have used in class (lectures, tutorials and term paper) to study the properties of the OLS estimator. A probit model builds the likelihood function under the hypothesis of normality of the error term. One way to show the behavior of these two estimators when normality is violated is to perform Monte Carlo simulations. We have used this tool several times during the course, e.g.

to discuss unbiasedness of the OLS under the CLM assumptions and when those assumptions were violated. Similarly here, one could use Monte Carlo simulations to study the behavior in small and large samples of $\hat{\beta}_1$.

Monte Carlo simulation is a method of analysis based on artificially recreating a random process with a computer, performing the estimation, observing the results and running the whole procedure many times.