Written Exam for the B.Sc. or M.Sc. in Economics Summer 2018

Applied Econometric Policy Evaluation

SUGGESTED ANSWERS

June 13, 2018

Problem 1 (10%):

1. The data set is a panel data set including observations for 12,000 individuals who are all observed for 24 months starting at the time where they get unemployed, i.e. the data set includes 288,000 observations in total. In the data set nobody returns from employment to unemployment.

   Table 1 shows summary statstics for the entire sample.

   Table 1: Summary statistics for the sample

VARIABLES	(1) N	(2) mean	(3) p50	(4) sd	(5) min	(6) max
age	288,000	33.49	33	9.191	18	49
yob	288,000	1,980	1,980	9.191	1,964	1,995
mob	288,000	6.440	6	3.468	1	12
dob	288,000	15.69	16	8.822	1	31
educ	288,000	2.640	3	1.179	1	4
jobcourse	288,000	0.190	0	0.393	0	1
employed	288,000	0.851	1	0.356	0	1
eligible	288,000	0.494	0	0.500	0	1

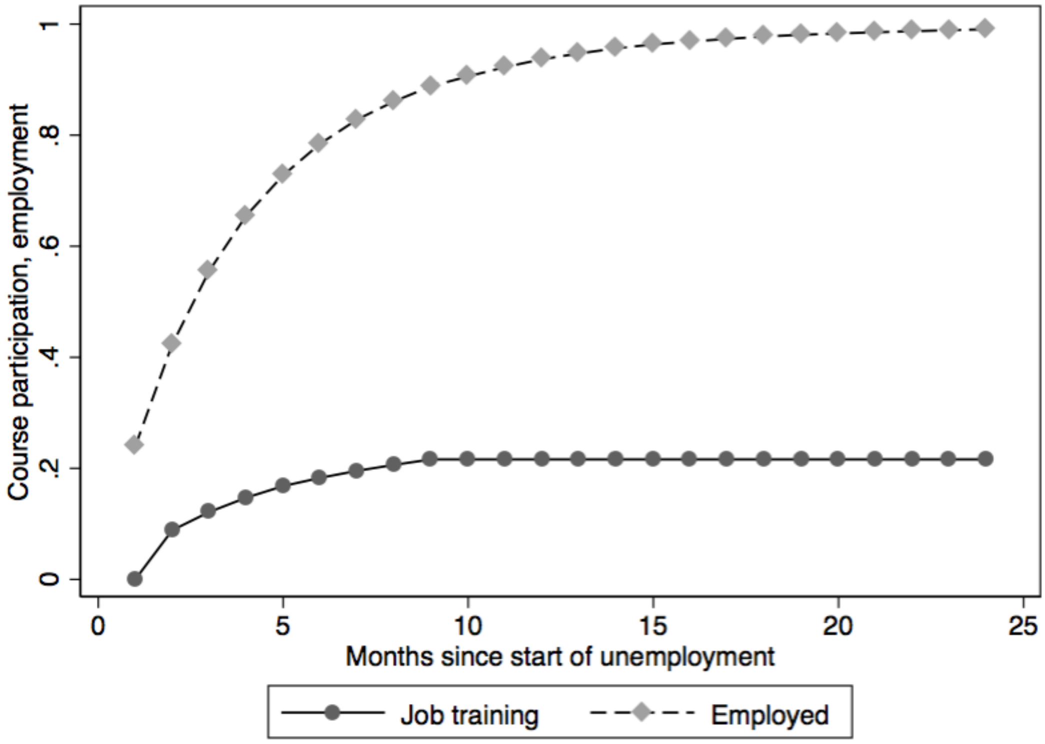
   The average age of the individuals in the sample is 33 years. On average, 83 percent are employed during the 24 months following the beginning of the unemployment spell. About 50 percent are eligible for being enrolled in the job search course.

   Figure 1 shows how enrolment in the job search course and employment ar are related to time elapsed since the unemployment spell began. Enrolment into the course is rising until eight months after the beginning of the unemployment event after which further enrolment stops. Employment increases as time elapses but it increases the most in the early phase where course participation is increasing. The raw data thus do not rule out that the course could be driving part of the increase in employment.

2. (a) Figure 1.2, Panel (A), presents the relationship between employment status and time since the beginning of the unemployment spell for people who are enrolled in the course and people who are not. Somewhat counter intuitively it shows that people who do not enrol in the course return to work sooner. This could be because the course prevents people from actually spending time looking for at job or it could be the result of self selection where the most able individuals choose not to attend the course.

   (b) In Figure 1.2, Panel (B), is presented the relationship between employment status and time since the beginning of the unemployment spell for people who are eligible and for people who are ineligible. This diagram shows that people who are eligible to participate in the course on average return to work sooner.

Figure 1. Unemployment, enrolment in the job search course and employment



Given that eligibility is based on quasi random assignment this appears to suggest that there might in fact be a positive effect of the course on the chances of finding a job.

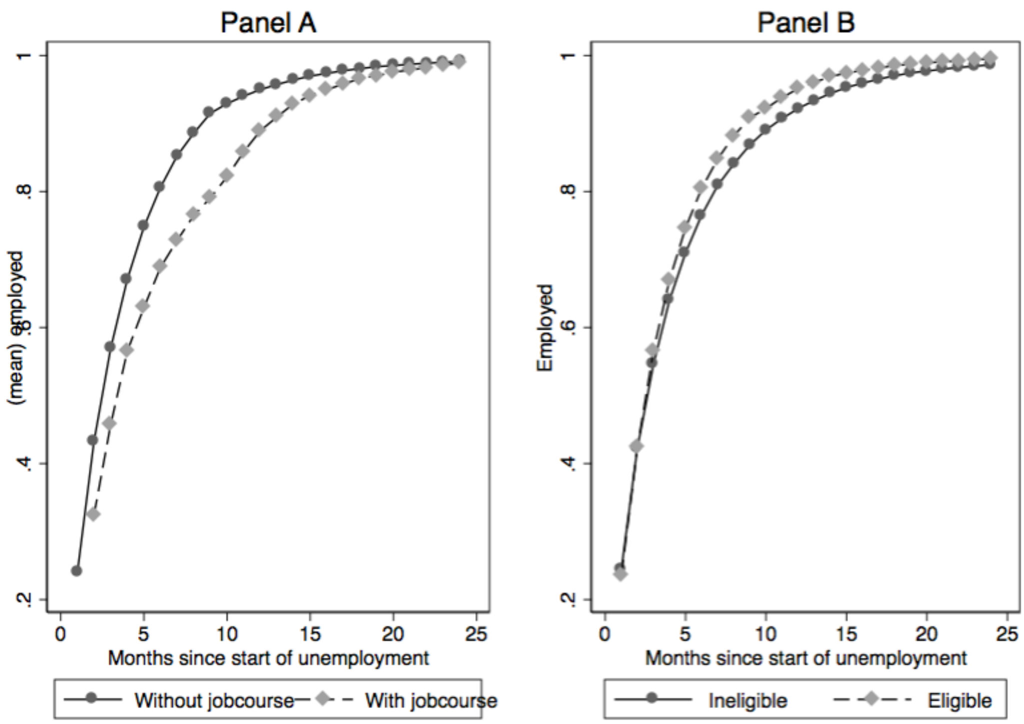
Problem 2 (20%):

In table 2 regression results applying to problem 2 are presented. The table only reports parameter estimates for the variables $jobcourse$, $eligible$, and $educ2 - educ4$. The variables age, $age2$ and $month2$-$month24$ have also been included in the regressions, but parameter estimates are not reported due to space considerations. All regressions are linear probability models, and robust standard errors are called for. However, because we are dealing with panel data it would be better to apply standard errors clustered at the individual level to allow for potential autocorrelated errors.

1. Table 2 column 1, shows the OLS estimate of $employment$ on $jobcourse$ (and covariates), and the parameter on $jobcourse$ is negative and significant. This is consistent with the evidence presented in Figure 1, Panel A, and it means, when taken at face value, that enrolling in the course actually lowers the chances of finding employment. This is likely due to self-selection into the jobcourse where individuals who are ex ante less likely to find a job select into the course. It is noted that employment is positively correlated with the level of education, so that the the group with the highest level of education, $educ4$ is about four times as likely to be employed during the 24 months covered by the data as the group with short education, $educ2$.

2. Table 2, column 2 presents results from regressing employment on $eligibility$ (and

Figure 2. Unemployment, enrolment in the job search course and employment



covariates). The results show that eligibility increases the chance of employment by about 2 percent when measured over the 24 months covered by the data.

3. Table 2, column 3, presents results from regressing *jobcourse* on *eligibility* (and covariates). This regression is also known as the first stage regression. It shows that those individuals who are eligible for the enrolling in the course are also more likely to actually enrol in the course.

4. The results from the two previous questions present the first stage estimate and the reduced form estimate. Using so called "Indirect least squares estimate" (ILS) these can be combined to give the result that would be produced by running $2SLS$ of employment on *jobcourse* using *eligibility* as an instrument. This is done in the following way: $\beta^{ILS} = \frac{\beta^{RF}}{\beta^{1st}} = \frac{0.0221}{0.3841} = 0.0575$. β^{ILS} measures the average effect of the course on employment measured over the 24 months covered by the data. The individuals in the sample, however, face a choice about when to enter the course if they decide to participate. This means that the effect measured is muted compared to the case where all were enrolled in the first month after the unemployment spell began. Relative to that case β^{ILS} measures a lower bound of the effect of taking the course on the chances of finding a job within the first 24 months after the onset of unemployement. The ILS estimate suggest that there is a positive effect of the course. However, based on the ILS estimate we cannot assess whether the effect is statistically significantly different from zero. To assess that we need to estimate the effect by $2SLS$.

Table 2: Regression results

VARIABLES	(1) OLS	(2) Reduced form	(3) 1st stage	(4) 2SLS
jobcourse	-0.0460***			0.0575***
	(0.0042)			(0.0090)
educ2	0.0209**	0.0240***	-0.0525***	0.0270***
	(0.0081)	(0.0082)	(0.0122)	(0.0085)
educ3	0.0567***	0.0605***	-0.0755***	0.0649***
	(0.0045)	(0.0046)	(0.0072)	(0.0048)
educ4	0.0764***	0.0834***	-0.1391***	0.0914***
	(0.0045)	(0.0045)	(0.0073)	(0.0049)
eligible		0.0221***	0.3841***	
		(0.0033)	(0.0056)	
Constant	0.2235***	0.2019***	-0.0219	0.2032***
	(0.0233)	(0.0237)	(0.0400)	(0.0244)
Observations	288,000	288,000	288,000	288,000
R-squared	0.3100	0.3085	0.2771	0.2975

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

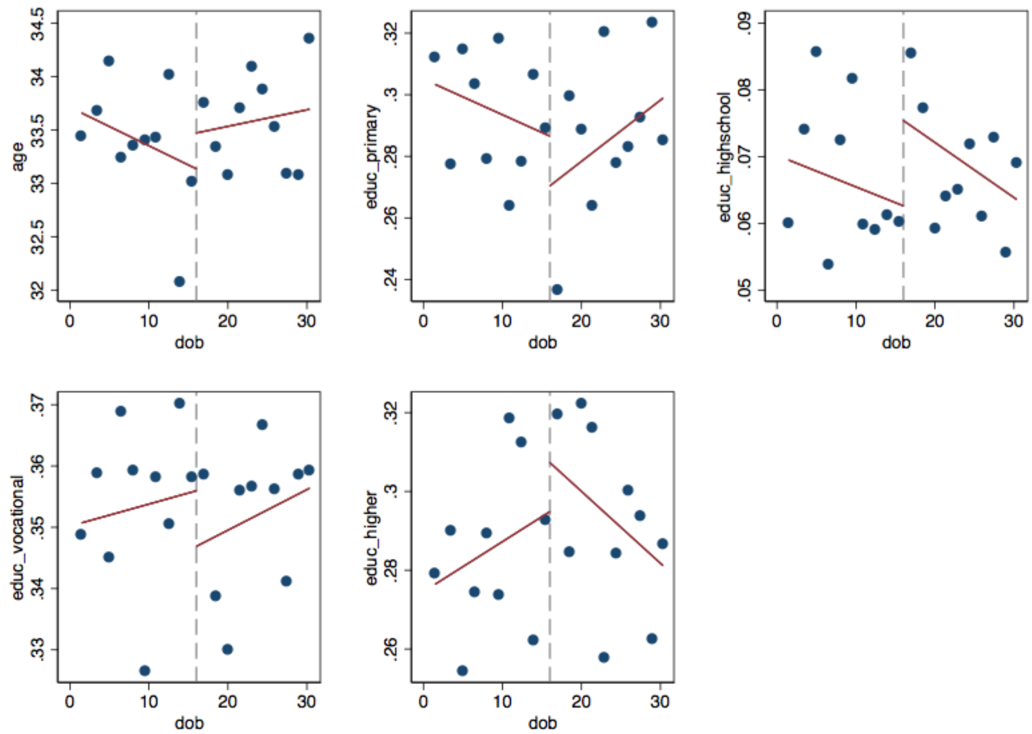
Problem 3 (20%):

1. Eligibility to the course is based on date of birth within the month so that people born in the first half of a month are eligible while people born in the second half are not. If we are willing to assume that birth day within the month is not of direct importance to the employment prospects absent participation in the course, then eligibility is as good as random.

2. Figure 3.1 presents binned scatterplots centred around the 16th day of the month for age and education dummies. None of the pictures show any evidence that education or age varies systematically to one side of the cutoff determining eligibility. This is consistent with the notion that eligibility has not been manipulated.

3. In Table 2, column 4, the $2SLS$ estimate is presented. Also for this regression, only the parameter of interest is presented. The regression does include the same controls used in the regressions presented in columns 1-3, but these are not reported. The results presented in Table 2, column 4, indicate that the there is a positive effect of participating in the course. Given the arguments in the answer to Problem 3.1 this represents a causal estimate of the effect of the course on the chances of obtaining employment. However, it should be noted that the individuals in the sample face a choice about when to enter the course if they decide to participate. This means that the effect measured is muted compared to the case where all were enrolled in the first month after the unemployment spell began.

Figure 3. Binned scatterplot around 16th day of the month



4. (a) $E[employment|eligible = 1] = 0.8622$
$E[employment|eligible = 0] = 0.8406$
$E[jobcourse|eligible = 1] = 0.3853$
$E[jobcourse|eligible = 0] = 0$

(b) The Wald estimator is given by $\frac{E(Y|Z=1)-E(Y|Z=0)}{E(S|Z=1)-E(S|Z=0)} = \frac{0.8622-0.8406}{0.3853-0} = 0.056$. Since no ineligible individuals can participate in the job search course, we have by definition no defiers and no always-takers. Thus, we only have i) persons who do not participate no matter they are ineligible or eligible (i.e. never-takers) and ii) persons who participate when eligible and not when ineligible (i.e. compliers). Therefore, everyone participating in the treatment are compliers, so the LATE estimate is an estimate of ATT

Problem 4 (25%):

1. Our instrument is based on the fact that the timing of birth within the month is quasi random and unrelated to labor market outcomes. Note, however, that the decision to enrol nor the timing of enrolment given enrolment are not randomised. By estimating 2SLS for various months, we might be able to examine treatment effects heterogeneity and whether persons entering treatment early in the unemployment spell has a larger or a smaller treatment effects. If the effect is smaller in the first couple of months after the start of unemployment, this does not necessarily imply that those entering early in the spell have lower treatment effects since the effect of treatment might not materialise immediately. Furthermore, due to the way the estimation is set-up, for a given month we obtain an average effect for all those who

have previously been enrolled into treatment. If, on the other hand, the effect of treatment is higher after only a few months since the beginning of the unemployment spell, this suggest that individuals with the highest gains select into treatment early on. We have estimated the effect separately for 2, 4, 6, 8, 10, and 12 months after the beginning of unemployment. The effect is small and insignificant in month 2, the largely flat over the months 4-8m, after which the effect becomes smaller again. We interpret this as those with the highest gains of treatment selecting into treatment early on. The estimated effects are, however not declining dramatically with elapsed time since the start of unemployment indicating that the selection into the course at different points in time is probably not dramatic. Finally, we note that the the estimated profile can differ slightly across the the ten data sets made available for the exam, but common to all of them is that effects tend to become smaller for months 10 and 12 than for months 4-8.

2. The overall share of compliers is 0.39, but we see that there are more compliers among persons with relatively lower education levels. For example, for the group of unemployed with no education, the share of compliers is 0.53, whereas the share of compliers for those with long education is 0.25. Since persons with relatively more education are more likely to become employed, we have negative selection into treatment. This negative selection is in accordance with the previous results that the estimated effect from OLS is negative, whereas the IV effect is positive.

Problem 5 (25%):

- Treatment/control groups: There is no natural treatment and control groups to identify the effect from. However, we are able to consider people who in 2010, i.e. before the policy was announced, contributed close to but more than 50,000 DKR (treatment group) with people who contributed close to but less than 50,000 DKR to annuity accounts (control group) and to compare the evolution of their contributions around the time of the policy. For example, consider a sample consisting of people who in 2010 contributed 25,000-75,000 DKR.

- Research design and equation: The effect of the policy can be quantified using a difference-in-differences setup where the evolution of pension contributions of the treatment and control groups from before to after the policy are compared. To do this consider Consider the following regression equation.

$$P_{it}^j = \beta_0 + \beta_1^j D_{it}^{TREAT} + \beta_2^j D_t^Y + \beta_{3t}^j D_{it}^{TREAT} \times D_t^Y + \beta_4^j I_{it} + u_{it} \tag{1}$$

Where P_{it}^j are pension contributions made by individual i in year t. $j = (capitalpension, annuity, lifeannuity)$. D_{it}^{TREAT} is a dummy variable taking the value one if individual i contributed 50,000-75,000 DKK to annuity accounts in 2010 and zero otherwise (i.e. if individual i contributed 25,000-49,999 DKK to annuity accounts in 2010. D_t^Y is a vector of year dummies for the years 2005-2009, 2011-2016. I_{it} is income, which acts as a control variable included because savings is known to be positively trending in the level of income irrespective of tax incentives. The effect of the policy is measured on β_{3t}^j, for $t \geq 2012$. For $j = annuity$ we would expect $\beta_{3,t}^j < 0$ (for

$t \geq 2012$) if the policy has any effect on the payments into annuity pension schemes, and if there is substitution (or crowding-out, as substitution is often referred to) we would we expect $\beta^j_{3,2012} > 0$ for $j = (capitalpension, lifeannuity)$ (and $t \geq 2012$).

- Validation of research design: If the two groups are comparable we would expect parallel pre-policy trends. This would amount to not rejecting the null hypothesis $H_0 : \beta^j_{3,2005} = \cdots = \beta^j_{3,2009} = 0$ against the alternative hypothesis $H_1 : \beta^j_{3,2005} \neq 0 \vee \cdots \vee \beta^j_{3,2009} \neq 0$.
  There is no natural treatment and control groups and consequently treatment and control groups are formed based on 2010 contributions. However, it could be that some had unusually high or low contributions in that year, so that they are in fact misclassified. To investigate whether this is a serious concern, the model can be re-estimated omitting observations in the close vicinity around the cut-off defined based on 2010 contributions.