EKT-816 Final Exam: Suggested Solutions

There are 55 points available on this exam. It will be graded out of 45 i.e. there are 10 bonus points available.

1. Consider a population of high school students who have the option to take an extracurricular program which teaches life skills, allows for some socializing, and provides academic enrichment activities and mentoring. However, the program (and travelling to the venue where it is held) requires a significant time commitment. Students also differ in their abilities, so while participating may help some students' performance in school, it could possibly hurt others.

For a given student, let Y_0 be an indicator for whether he or she would graduate without participating in the program, and let Y_1 be an indicator for whether he or she would graduate when participating.

(a) List the four possible configurations of the potential outcomes for a given student.

Answer: Whether you graduate or not is binary. The joint distribution of potential outcomes is as follows:

	$Y_1 = 0$	$Y_1 = 1$
$Y_0 = 0$	0.25	0.5
$Y_0 = 1$	0.15	0.1

(b) Suppose that 10% of the students would graduate with or without the program; 15% are strictly harmed by it (i.e. they would graduate without it, but not with it); and that 25% would not graduate whether or not they participate. What is the average treatment effect of the program?

Answer: Filling in the joint distribution as above, we have $ATE = E[Y_1 - Y_0] = E[Y_1] - E[Y_0] = 0.6 - 0.25 = 0.35$. So the treatment helps 35% of the students (on average).

(c) Suppose that students choose to participate if they are strictly better off with the program. What fraction of the population will participate? What will be the average effect of treatment on the treated?

Answer: We are given that $D = 1$ if and only if $Y_1 > Y_0$. Thus, $P(D = 1) = P(Y_1 = 1, Y_0 = 0) = 0.5$. For this subset of the population, $Y_1 - Y_0 = 1$ (for everyone who participates), so $ATT = E[Y_1 - Y_0|D = 1] = 1$.

(d) A journalist, who has never studied the potential outcomes model, decides to measure the effect of the program by comparing the graduation rate for participants with the graduation rate of non-participants. What would his estimate of the program's effect be? Is it an overestimate or an underestimate of the average treatment effect you computed in (b)? How would you explain the apparent difference in the estimates?

[$4 \times 5 = 20$ points]

Answer: The naive difference in means is $E[Y|D = 1] - E[Y|D = 0] = E[Y_1|D = 1] - E[Y_0|D = 0]$. Using the given numbers, we have

$$E[Y_0|D = 0] = \frac{0.25 \times 0 + 0.15 \times 1 + 0.1 \times 1}{0.5} = \frac{0.25}{0.5} = 0.5$$

and $E[Y_1|D = 1] = 1$. So the journalist would think the program raises graduation rates by $1 - 0.5 = 0.5$, i.e. by 50 percentage points. We know the ATE is only $0.35 < 0.5$, a much smaller effect. Interestingly, although the journalist overestimates the ATE, he *underestimates* the ATT, which is 1! Of course, the discrepancy arises from selection bias: the set of students who choose $D = 0$ includes many who would never graduate or would have graduated anyway, with or without the program.

2. "Puerperal fever" refers to any type of bacterial infection of a woman's reproductive organs. Many such infections occur during or immediately following childbirth, which is why they are also called

"postpartum infections". If untreated, these infections can be fatal. In the 19th century, however, antibiotics had not yet been invented and indeed the germ theory of disease was not well established.

Figures 1-2 below display data on death rates from the two maternity clinics of Vienna's General Hospital during the 1840s. In the First Clinic, deliveries of newborn children were conducted by physicians, while deliveries in the Second Clinic, they were handled by midwives and nurses.

(a) What do you notice about the death rates in each clinic from Figures 1-2? Can you conclude *from the evidence in Figure 2* that - at least during this time - the midwives are more skilled than physicians at delivering children? Why or why not? What might be some other factors that affect death rates in one clinic vs the other?

Answer: Death rates are always higher in the first clinic, where the physicians handle births. That difference *could* be due to differences in skill, but many other things might differ between the two clinics - they might serve different types of women (e.g. maybe sicker women go to one vs the other), or maybe there are environmental factors that make one clinic more dangerous. (Indeed this was the case, as we discuss below.)

(b) In 1841, the hospital introduced a new training program for physicians, and this training program involved the handling of cadavers. Midwives did not participate in this training program. And, in May of 1847, the hospital introduced a new policy that required all physicians to wash their hands with chlorine after handling cadavers. If you only had access to the data on death rates from 1846 and 1848, could you estimate the effect of the handwashing policy on death rates?

Answer: The diff-in-diff estimate of the effect of the handwashing policy is

$$(1.27 - 11.4) - (1.33 - 0.9) = -10.13 - 0.43 = -10.56$$

i.e. the policy seems to have reduced deaths by 10.56 percentage points per year. Although we don't have standard errors on this estimate, it is an incredibly large effect, reducing deaths (at least from postpartum infections) at the physician's clinic to nearly zero. (Given what we know about the mechanisms of infection, this isn't so surprising in retrospect!)

(c) What assumption do you need to make about the trends in death rates in the physician's clinic relative to the midwives' clinic in order to interpret your estimate in (b) above as causal? Do you see any evidence in favor of that assumption, or any evidence against it, in Figures 1-2?

[$3 \times 5 = 15$ points]

Answer: As discussed in (a), there may be other unobserved factors that account for the differences in the *level* of death rates. But, if the *trends* in both clinics would have been the same in the absence of the policy, we can interpret the DD estimate as causal. Figure 2 would seem to provide us with a lot of confidence in that assumption: for the period 1841-1846, it seems like death rates in both clinics do indeed move in parallel.

	Physicians' Division			Midwives' Division		
		Deaths			Deaths	
Year	Births	No.	%	Births	No.	%
1841	3036	237	7.7	2442	86	3.5
1842	3287	518	15.8	2659	202	7.5
1843	3060	274	8.9	2739	169	6.2
1844	3157	260	8.2	2956	68	2.3
1845	3492	241	6.8	3241	66	2.03
1846	4010	459	11.4	3754	105	2.7
1847				3306	32	0.9
January–May	2134	120	5.6			
	Intervention introduced in May					
June–December	1841	56	3.04			
1848	3556	45	1.27	3219	43	1.33

Figure 1: Death Rates for Obstetrics Patients in Vienna General Hospital, 1841-1848

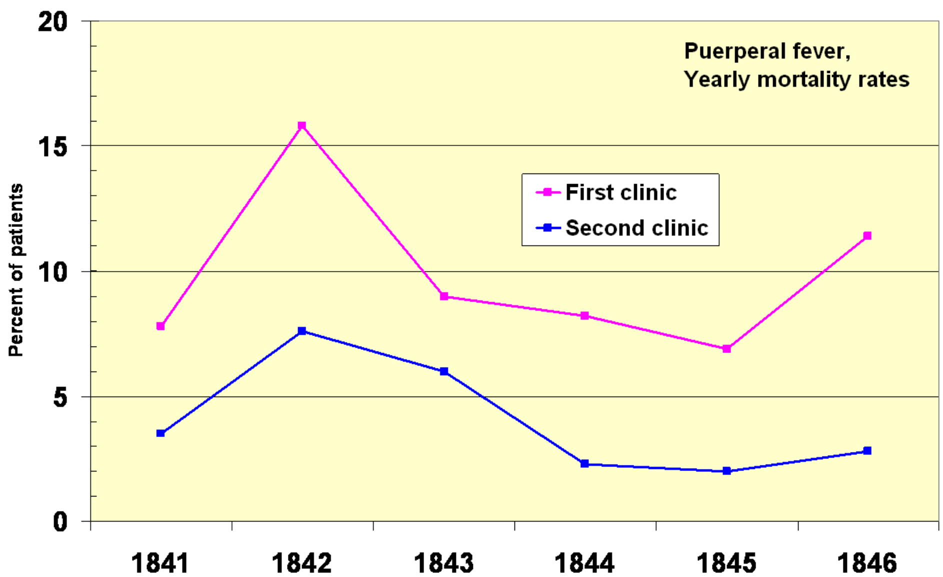

Figure 2: Death Rates for Obstetrics Patients in Vienna General Hospital, 1841-1846

3. Consider the excerpt from Dinkelman (2011), reproduced as Figure 3 below.

(a) The dependent variable in each of the regressions in Figure 3 is a dummy for whether a given community area received an Eskom project within the sample period. What sense can you make of a coefficient of -0.083 on the variable "gradient $\times$ 10" in this case? How should you interpret such a coefficient?

Answer: Because the dependent variable is binary, the coefficients have the interpretation as the effects on the probability of recieving an Eskom project. In particular, a community with average land gradient that is 10 degrees steeper than another (otherwise equal) community, would have been 8.3 percentage points less likely to be connected to the electricity grid over the sample period.

(b) Can you think of a problem that might arise in using OLS with a binary dependent variable? Suggest an alternative way of estimating the "effect" of these covariates on the binary dependent variable of project receipt that would avoid this problem.

Answer: A generic problem here is that predicted values can lie outside of the unit interval, which doesn't make sense for probabilities. You could think of using a logit or probit specification instead, although that seems less desirable in context because we actually want to use these predicted values in the "second stage". (See the discussion of "forbidden regressions" in section 4.6.1 of "Mostly Harmless Econometrics".)

(c) At the bottom of the table is a reported F-statistic for each of the specifications. Why does Dinkelman report this F-statistic in particular? What concerns might be raised by a low value for this test statistic, and what implications should they have for the reliability of her estimation strategy?

Answer: The F-statistic in question is for a test of the null hypothesis that land gradient can be excluded from the first-stage regression. Low values of this test statistic can indicate that the instrument (in this case, land gradient) is weak. Weak instruments are a problem because they tend to increase the bias of IV estimates.

(d) From your knowledge of the paper, explain what exclusion restriction is necessary for Dinkelman's estimates to be valid. Can you suggest a way to test whether these restrictions hold?

[$4 \times 5 = 20$ points]

Answer: The exclusion restriction we need is that land gradient does not affect employment, except through its effects on the likelihood of electrification. If we had multiple instruments, we could perform an overidentification test (although that is really just a test of whether the instruments agree with each other, not whether they are "valid" - they could both be invalid and the test might still fail to reject). But, the econometric model in this paper is just-identified, so even that is not possible.

References

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review* 101 (7). American Economic Association: 3078–3108. https://doi.org/10.1257/aer.101.7.3078.

TABLE 3—ASSIGNMENT TO ESKOM PROJECT: FIRST STAGE OLS ESTIMATES

Dependent variable: Eskom project = [1 or 0]	(1)	(2)	(3)	(4)
Gradient × 10	−0.083**	−0.075**	−0.078***	−0.077***
	(0.040)	(0.034)	(0.027)	(0.027)
Kilometers to grid × 10		−0.040*	−0.012	−0.011
		(0.021)	(0.023)	(0.023)
Household density × 10		0.017***	0.012**	0.013**
		(0.004)	(0.006)	(0.006)
Poverty rate		0.023	0.019	0.017
		(0.069)	(0.070)	(0.069)
Female-headed HHs		0.393***	0.165	0.155
		(0.120)	(0.107)	(0.107)
Adult sex ratio		−0.173***	−0.130***	−0.121***
		(0.052)	(0.042)	(0.042)
Indian, white adults × 10		−1.236***	−1.116**	−1.105**
		(0.401)	(0.459)	(0.452)
Kilometers to road × 10		0.003	−0.010	−0.010
		(0.009)	(0.010)	(0.010)
Kilometers to town × 10		0.016	0.008	0.008
		(0.015)	(0.015)	(0.016)
Men with high school		−0.269	−0.185	−0.152
		(0.500)	(0.411)	(0.417)
Women with high school		1.046**	0.965**	0.984**
		(0.475)	(0.413)	(0.409)
Δ_t water access				0.012
				(0.048)
Δ_t toilet access				0.155
				(0.104)
District fixed effects	N	N	Y	Y
Mean of outcome variable	0.20	0.20	0.20	0.20
N communities	1,816	1,816	1,816	1,816
R^2	0.01	0.07	0.18	0.18
F-statistic on gradient	4.20	4.87	8.34	8.26
Pr > F	0.04	0.03	0.00	0.00

Notes: Robust standard errors clustered at subdistrict level. Ten district fixed effects included in columns 3 and 4. Change in fraction of households with access to water and flush toilet measured between 1996 and 2001.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

The inclusion of district fixed effects in this first stage is important, as a large amount of the variation in gradient comes from cross-district variation (see Figure 3). This means that without controlling for district, the first stage compares project assignment across very different places in terms of gradient and in terms of local labor market conditions. By controlling for district as in columns 3 and 4, I compare places that are in the same local labor market, but which are slightly flatter or steeper.

The two other cost variables have coefficients of expected sign in the first stage results of Table 3: a three-quarter–standard deviation increase in distance from the grid (about 10 kilometers) reduces the probability of electrification by 1 percentage point, although this is not significant when all other controls are added. A one-third–standard deviation increase in household density (10 households) per square kilometer increases the probability of electrification by about 1.3 percentage

Figure 3: Excerpt From @Dinkelman2011