EKN-816 Practice Final: Suggested Solutions

There are 60 points available on this exam. It will be graded out of 50 i.e. there are 10 bonus points available.

You have three hours to complete this exam. You may refer to any written materials you find useful, but no electronic devices are permitted.

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:

	$Y_1 = 0$	$Y_1 = 1$
$Y_0 = 0$	0.35	0.15
$Y_0 = 1$	0.2	0.3

(b) Suppose that 30% of the students would graduate with or without the program; 20% are strictly harmed by it (i.e. they would graduate without it, but not with it); and that 35% 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 $E[Y_1 - Y_0] = E[Y_1] - E[Y_0] = 0.45 - 0.5 = -0.05$. So the treatment actually harms students (on average).

(c) Now suppose that students choose to participate if they are weakly better off with the program (i.e. if $Y_1 \geq Y_0$). What will be the average effect of treatment on the treated?

Answer: We are given that $D = 1$ if and only if $Y_1 \geq Y_0$. Thus, $P(D = 1) = 0.35 + 0.15 + 0.3 = 0.8$.

Next, to compute the ATT, we need to find

$$E[Y_1|D = 1] = \frac{0 \times 0.35 + 1 \times 0.15 + 1 \times 0.3}{0.8} = 0.5625$$

and

$$E[Y_0|D = 1] = \frac{0 \times 0.35 + 0 \times 0.15 + 1 \times 0.3}{0.8} = 0.375$$

so $ATT = 0.5625 - 0.375 = 0.1875$. Notice, by the way, that $E[Y|D = 0] = 0$, so the naive difference in means is 0.5625, much larger than the ATT and the wrong sign compared to the ATE of -0.05.

[$3 \times 5 = 15$ points]

2. Consider a linear regression model with one regressor

$$y_i = \beta_0 + \beta_1 x_i^* + \varepsilon_i \tag{1}$$

with $E[\varepsilon_i|x_i^*] = 0$ and $V[\varepsilon_i|x_i^*] = \sigma_\varepsilon^2$. Unfortunatly, you observe $(y_i, x_i)_{i=1}^n$, where

$$x_i = x_i^* + u_i \tag{2}$$

is a mismeasured version of x^*. Assume

$$E[u_i|x_i^*] = 0 \quad (3)$$
$$V[u_i|x_i^*] = \sigma_u^2 \quad (4)$$

(a) Show that a regression of y on x will be inconsistent for β_1. In particular, show that

$$\text{plim}_{n\to\infty}\widehat{\beta}_1 = \beta_1\left(\frac{V[x^*]}{V[x^*]+\sigma_u^2}\right) \quad (5)$$

(Hint: use the fact that $\widehat{\beta}_1$ is the ratio of $\text{cov}(y,x)$ to $V[x]$ in the sample. Then invoke the LLN and Slutsky's theorem.) This phenomenon is often called *attenuation bias.*

Answer: We have $\text{cov}(y,x) = \beta_1 \times (V[x] - \text{cov}(x,u)) = \beta_1 V[x^*]$. Since $V[x] = V[x^*] + \sigma_u^2 > V[x^*]$, we have the usual attenuation bias result.

(b) Now consider estimating the "reverse" regression

$$x_i = \delta_0 + \delta_1 y_i + \widetilde{\varepsilon}_i \quad (6)$$

Show that $\left(\text{plim}_{n\to\infty}\,\widehat{\delta}_1\right)^{-1}$ is an upper bound for β_1 in absolute value, i.e.

$$|\text{plim}_{n\to\infty}\widehat{\beta}_1| \leq \quad |\beta_1| \quad \leq \left(|\text{plim}_{n\to\infty}\widehat{\delta}_1|\right)^{-1} \quad (7)$$

Answer: Rearrange the expression for y in terms of x^* to get

$$x^* = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1}y - \frac{1}{\beta_1}\varepsilon$$

So, if we add u to both sides and take the covariance with y, we get

$$\text{cov}(y,x) = \frac{1}{\beta_1}V[y] - \frac{1}{\beta_1}\text{cov}(y,\varepsilon) + \text{cov}(y,u)$$

Then, we get

$$\underset{n\to\infty}{\text{plim}}\,\widehat{\delta}_1 = \frac{\text{cov}(x,y)}{V[y]} = \frac{1}{\beta_1}\left(\frac{V[y]-\sigma_\varepsilon^2}{V[y]}\right).$$

Since $V[y] > \sigma_\varepsilon^2 > 0$, we have the required result.

[$2\times 5 = 10$ points]

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

Recall that the model in the paper is as follows: for community j in district d at time t,

$$y_{jdt} = \alpha_0 + \alpha_1 t + \alpha_2 T_{jdt} + \mu_j + \delta_j t + \rho_d + \lambda_d t + \varepsilon_{jdt} \quad (8)$$

where y_{jdt} is the employment-to-population ratio for a given sex. Taking first differences we get

$$\Delta y_{jdt} = \alpha_1 + \alpha_2 \Delta T_{jdt} + \delta_j + \lambda_d + \Delta\varepsilon_{jdt} \quad (9)$$

Dinkelman also includes controls for baseline levels of employment $\mathbf{X}_{jdt}$ (population density, local sex ratio, average education levels, distance to electricity grid, to major roads, to major town…) She then uses average land gradient in the community as an instrument for ΔT_{jdt}: more gently sloped areas are easier to connect to the grid. So, she estimates the IV system

$$\begin{aligned}\Delta y_{jdt} &= \alpha_1 + \alpha_2 \Delta T_{jdt} + \mathbf{X}_{jdt}\beta + \delta_j + \lambda_d + \Delta\varepsilon_{jdt} \qquad (10)\\ \Delta T_{jdt} &= \pi_0 + \pi_1 Z_j + \mathbf{X}_{jdt}\pi_2 + \gamma_d + \tau_{jdt} \qquad (11)\end{aligned}$$

(a) What is the "treatment effect" of interest? Suppose the treatment effect is estimated to be 0.03. Explain in words what this means.

Answer: The coefficient on T_{jdt} is the "treatment effect". If $\alpha_2 = 0.03$, this would mean that (holding all the other variables constant) we would expect the employment-population ratio in a community that got connected to the grid to grow by 3 percentage points more - relative to trend - than one which did not.

(b) Is this model overidentified?

Answer: No, the model is just-identified: there is one endogenous regressor (T_{jdt}) and one instrument (average gradient, Z_i).

(c) Can you say anything about instrument relevance from the evidence provided here?

Answer: No: we only have the IV estimates from the "main equation" (10). To analyze instrument regressors, we'd need the estimates and diagnostic statistics from the first stage (11).

(d) Can you think of a reason why $\Delta\varepsilon_{jdt}$ might be heteroskedastic? Explain the economics of your argument. (Hint: think of what ε_{jdt} represents. What might you expect about the variance of those factors across communities j or across districts d?)

Answer: This is a harder question. The basic idea is to translate what heteroskedasticity means into context and think about whether it sounds plausible. So here "heteroskedasticity" means that $V[\Delta\varepsilon_{jdt}]$ is not constant across communities j. What do the ε_{jdt}'s represent? They are "shocks" to employment.

The model does not take an explicit position on whether the shocks are demand or supply driven, which makes the argument trickier, but we might imagine that there are labor demand shocks in some communities that are larger than in other communities, if for example the composition of employment is different across areas. (Imagine a village where everyone works as casual laborers in construction as compared to one where most people work in manufacturing, which we expect to be more stable.)

(e) Suppose that a community with a poverty rate of 0.65 (i.e. 65%) in 1996 that is unconnected to the electricity grid receives an electrification project during the sample period. Coincidentally the poverty rate in this community also falls to 61% over the same period. What do you predict will happen to the employment-population ratio for women in this community?

Answer: Let $\widehat{\beta}_{PR}$ be the estimated coefficient on poverty rate, which in specification (8) is 0.031. We have the predicted change in employment growth

$$\begin{aligned}E[\Delta y|T=1, PR=0.61] - E[\Delta y|T=0, PR=0.65] &= \widehat{\alpha}_2 + \widehat{\beta}_{PR}(0.61 - 0.65)\\ &= 0.095 + 0.031 \times (0.61 - 0.65)\\ &= 0.09376\end{aligned}$$

(f) Can you quantify the precision of your prediction above? If so, provide the standard error of the prediction you made. If you cannot, what information would you need to calculate it? Be as explicit as you can.

Answer: We saw that the predicted change is $\widehat{\alpha}_2 + -0.04\widehat{\beta}_{PR}$, which therefore has a variance of

$$\begin{aligned}V[\widehat{\alpha}_2 + -0.04\widehat{\beta}_{PR}] &= V[\widehat{\alpha}_2] + -2 \times 0.04\,\mathrm{cov}(\widehat{\alpha}_2, \widehat{\beta}_{PR}) + (0.04)^2 V[\widehat{\beta}_{PR}]\\ &= (0.055)^2 + (0.04)^2(0.013)^2 - 0.08\,\mathrm{cov}(\widehat{\alpha}_2, \widehat{\beta}_{PR})\end{aligned}$$

so we cannot estimate the standard error of the prediction from the information in the given regression table. We need to know the covariance of the two estimators involved, $\mathrm{cov}(\widehat{\alpha}_2, \widehat{\beta}_{PR})$, too.

(g) Comment on the validity of Dinkelman's instrument. Could you test whether the instrument is valid with this data?

Answer: Taking the second question first, we can't test the identifying restriction (that land gradient does not affect employment, except through its effects on the likelihood of electrification and through its correlation with the controls), since we are just-identified. (If we had an overidentified model we could perform a Hansen-Sargan test.)

As for the first question, I was looking for you to say something reasonable. For example you could have said something about agricultural employment being directly affected by gradient (a point discussed in the introduction of the paper).

[$7 \times 5 = 35$ points]

References

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review* 101 (7): 3078–3108.

TABLE 4—EFFECTS OF ELECTRIFICATION ON EMPLOYMENT: CENSUS COMMUNITY DATA

	Δ_t female employment rate							
	OLS regression coefficients				IV regression coefficients			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eskom project	−0.004	−0.001	0.000	−0.001	0.025	0.074	0.090*	0.095*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.045)	(0.060)	(0.055)	(0.055)
A. R. 95 percent C.I.							*[0.05; 0.3]*	*[0.05; 0.3]*
Poverty rate		0.029***	0.033***	0.031***		0.027**	0.032**	0.031**
		(0.011)	(0.010)	(0.010)		(0.012)	(0.013)	(0.013)
Female-headed HHs		0.042**	0.051***	0.047**		0.014	0.036	0.033
		(0.019)	(0.019)	(0.020)		(0.031)	(0.026)	(0.026)
Adult sex ratio		0.019**	0.017**	0.020***		0.033**	0.029**	0.032***
		(0.009)	(0.008)	(0.007)		(0.014)	(0.012)	(0.012)
Baseline controls?	N	Y	Y	Y	N	Y	Y	Y
District fixed effects?	N	N	Y	Y	N	N	Y	Y
Δ_t other services?	N	N	N	Y	N	N	N	Y
N communities	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816

Notes: Robust standard errors clustered at subdistrict level. Eskom project is instrumented with mean community land gradient. See Table 3 for full list of control variables. The last two columns provide confidence intervals (C.I.) from the Anderson-Rubin (A.R.) test for the coefficient on Eskom project. The test is robust to weak instruments and implemented to be robust to heteroskedasticity.

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

TABLE 5—EFFECTS OF ELECTRIFICATION ON EMPLOYMENT: CENSUS COMMUNITY DATA

	Δ_t male employment rate							
	OLS regression coefficients				IV regression coefficients			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eskom project	−0.017**	−0.015***	−0.009	−0.010*	−0.063	0.069	0.033	0.035
	(0.007)	(0.006)	(0.006)	(0.006)	(0.073)	(0.082)	(0.064)	(0.066)
A. R. 95 percent C.I.							*[−0.05; 0.25]*	*[−0.05; 0.25]*
Poverty rate		0.062***	0.064***	0.063***		0.059***	0.064***	0.062***
		(0.020)	(0.018)	(0.018)		(0.022)	(0.019)	(0.019)
Female-headed HHs		0.217***	0.233***	0.227***		0.187***	0.227***	0.220***
		(0.029)	(0.030)	(0.030)		(0.042)	(0.034)	(0.034)
Adult sex ratio		0.018*	0.012	0.017		0.034*	0.018	0.023
		(0.011)	(0.011)	(0.011)		(0.019)	(0.015)	(0.015)
Baseline controls?	N	Y	Y	Y	N	Y	Y	Y
District fixed effects?	N	N	Y	Y	N	N	Y	Y
Δ_t other services?	N	N	N	Y	N	N	N	Y
N communities	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816

Notes: Robust standard errors clustered at subdistrict level. Eskom project is instrumented with mean community land gradient. See Table 3 for full list of control variables. The last two columns provide confidence intervals (C.I.) from the Anderson-Rubin (A.R.) test for the coefficient on Eskom project. The test is robust to weak instruments and implemented to be robust to heteroskedasticity.

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

community-level controls and district fixed effects in columns 2 and 3 increases the coefficient on electrification slightly, with the female employment effect still not significantly different from zero and male employment becoming less negative and less statistically significant. The positive, significant coefficients on poverty rate, sex

Figure 1: Excerpt From @Dinkelman2011