Introduction to Causal Inference

Solutions to Problem Set 4

Professor: Teppei Yamamoto

Due Tuesday, July 19 (at beginning of class)

Only the required problems are due on the above date. The optional problems will not directly count toward your grade, though you are encouraged to complete them as your time permits. If you choose not to work on the optional problems, use them for your self-study during the summer vacation.

Required Problems

Problem 1

In this problem we will assess one of the most famous social science articles using instrumental variables, Acemoglu, Johnson, & Robinson's 2001 paper "The Colonial Origins of Comparative Development: An Empirical Investigation" (henceforth AJR).[1]

First, we will begin with a stylized characterization of the study. Assume that AJR use the following variables for any country i that was previously colonized:

Instrument $Z_i \in \{0,1\}$: Mortality in the 17th, 18th, and early 19th centuries, 0 if low mortality, 1 if high.

Treatment $D_i \in \{0,1\}$: Modern property rights institutions, 0 if weak, 1 if strong.

Outcome Y_i: Modern log GDP per capita.

In our stylized characterization, assume that AJR use instrumental variables to estimate the effect of D_i on Y_i by instrumenting for D_i with Z_i. They find that having strong modern property rights institutions causes higher GDP per capita. (**Note:** As we will see in a minute, AJR include various specifications in which they also control for some pre-treatment covariates, but for now we will focus on the "simplest" empirical strategy.)

(a) Assuming that their empirical strategy is valid, draw a simple DAG to represent the instrumental variables approach used by AJR. Include a hypothetical unobserved confounder (U_i) that creates a back-door path between treatment and outcome. Why is it important to include this hypothetical unobserved confounder?

$$\begin{tikzcd} & & U \arrow[dl, dashed] \arrow[dr, dashed] & \\ Z \arrow[r] & X \arrow[rr] & & Y \end{tikzcd}$$

Figure 1: AJR's Empirical Strategy

The hypothetical confounder is important because without it we would not need to use IV as an empirical strategy at all. That is, if there was no unobserved backdoor path connecting D_i to Y_i, then we would have a simply randomized experiment in which $\{Y_i(d)\} \perp\!\!\!\perp D_i$.

(b) Name the five assumptions underpinning instrumental variables as a strategy for identifying the effect of an endogenous treatment on the outcome for compliers. Write out each assumption formally in terms of Z_i, D_i, and Y_i. In your own words, interpret each assumption with regard to the specific setup of AJR's study. Finally, discuss the plausibility of each assumption. (**Hint:** It may be useful to refer to your DAG from (a) in interpreting and assessing some assumptions.)

First, define potential outcomes for both the outcome variable as $Y_i(z, d)$ and for the treatment indicator as $D_i(d)$.

(a) Ignorability: $\{Y_i(z = 1), Y_i(z = 0), D_i(z = 0), D_i(z = 1)\} \perp\!\!\!\perp Z_i$.
Ignorability in this case requires that settler mortality (Z_i) is independent of potential outcomes for Y_i *and* potential outcomes for treatment status D_i under z. That is, knowing the value of Z_i does not tell us anything about the *potential outcomes* for log GDP per capita Y_i, nor the potential treatment status for property rights strength, D_i.
Is this plausible? First, note that we typically cannot assess this directly with data. Theroetically, this assumptions seems kind of unlikely in the AJR case. Settler mortality is not assigned randomly, nor in a quasi-random way – it is a function,of, among many othr things, the disease environment, the latitude/longitude of an area, natural geography, agricultural feasibility, water access, etc. Many of these factors could simultaneously correlate with (create a backdoor path to) both property rights institutions (treatment D_i) or log GDP per capita (outcome Y_i).

(b) Exclusion restriction: $Y_i(1, d) = Y_i(0, d)$ for $d = (1, 0)$.
The exclusion restriction in this case requires that the potential outcomes for log GDP per capita, when property rights institutions are weak (strong), are equal irrespective of the level of settler mortality. That is, there is no effect of settler mortality on log GDP per capita *other than through* the effect of Z_i on D_i. In our DAG, assuming that there is

[1]This paper has over 8,000 citations, and is heavily debated in a range of disciplines including economics, political science, and history. This problem set question is highly stylized, and if you are really interested in the substantive and methodological details of the paper we encourage you to read the paper and surrounding debates carefully. Also remember, it is always easier to criticise something than to build it yourself!

no backdoor path connecting Z_i and Y_i, the exclusion restriction is satisfied when there is no directed edge departing Z_i and entering Y_i.

Is this plausible for AJR? Again, we cannot directly assess this with data. Theoretically, we can probably imagine a number of potential ways in which Z_i could affect Y_i other than through levels of D_i. For instance, higher levels of settler mortality might have induced more conflict over resources, lowering log GDP per capita in the future.

(c) First-stage relevance: $0 < P(Z_i = 1) < 1$ and $P(D_i(1) = 1) \neq P(D_i(0) = 1)$

First-stage relevance in this case means that higher levels of settler mortality correlates with (causes?) weaker property rights institutions. In the DAG, this means that the hypothesized directed edge connecting Z_i to D_i exists.

The plausibility of this assumption can be tested in data, and seems probable in this case (especially if you read the paper, and see their empirical tests of this).

(d) Monotonicity: $D_i(1) \geq D_i(0)$ for all i.

This assumption rules out defiers: colonies whose potential treatment status with high settler mortality is strong property rights institutions, and conversely whose potential treatment status with low settler mortality is weak property rights institutions, $D_i(0) = 1$ and $D_i(1) = 0$, or $D_i(1) < D_i(0)$.

This assumption does seem reasonably plausible in the case of AJR. It's hard to think of a (plausible) reason why some colonies would be defiers.

(e) SUTVA:

$Y_i(Z_1, Z_2, ..., Z_N, D_1, D_2, ..., D_N) = Y_i(Z_1', Z_2', ..., Z_N', D_1', D_2', ..., D_N') \quad \text{if} \quad D_i = D_i', Z_i = Z_i', \ \forall\, i,$

and $D_i(Z_1, Z_2, ..., Z_N) = D_i(Z_1', Z_2', ..., Z_N') \quad \text{if} \quad Z_i = Z_i', \ \forall\, i.$

The first part of SUTVA outlined above is what we are used to – for any unit i, potential outcomes for log GDP per capita would be equal irrespective of the strength of all other units' property rights instititions or their levels of settler mortality, as long as unit i's property rights institutions and level of settler mortality remain the same. The second line says the same thing, but about unit i's potential outcome of property rights institutions' strength, depending on the settler mortality level of all other units.

There are some reasons to question the plausibility of this set of assumptions, especially when dealing with national-level questions. For instance, if we assume that international investment comes from a fixed pot and is divided up on the basis of property rights regimes, then the first equality may not be true. Similarly, if colonial investment decisions occurred in the same way, we might believe that each country's property rights strength is contingent on settler mortality for other units. Of course, such concerns plague any study conducted at the national level.

(c) We will replicate the main specifications from AJR using their publicly available replication data. Download the data `ajr_data.dta` from the course website and read it into `R`. The data has the following variables:

`shortnam`: Three letter country code for each unit.
`logpgp95`: Log purchasing power parity GDP per capita, 1995.
`avexpr`: Average protection against expropriation risk.
`logem4`: Log settler mortality.
`lat_abst`: Absolute value of latitude of capital city.
`africa`: Dummy=1 if African.
`asia`: Dummy=1 if Asian.
`other`: Dummy=1 if Other.
`america`: Dummy=1 if American.

Estimate the effect of `avexpr` on `loggp95` in two ways using naïve regression approaches. First, run a linear regression of `loggp95` on `avexpr` as the lone regressor (do not include any other covariates). Second, do the same but include, linearly and additively, `lat_abst`, `africa`, `asia`, and `other`. Present the results in a table, including HC2 robust standard errors. Interpret the direction and statistical significance of the estimates. Why should we be concerned about whether these are good estimates of the causal quantity of interest?

```
library(foreign)
library(xtable)
library(stargazer)
library(lmtest)
library(zoo)
library(sandwich)
d = read.dta("ajr_data.dta")

ols.mod1 = lm(logpgp95 ~ avexpr, d)
ols.mod1HC2 = coeftest(ols.mod1, vcov = vcovHC(ols.mod1, type = "HC2"))[,2]
ols.mod2 = lm(logpgp95 ~ avexpr + lat_abst + africa + asia + other, d)
ols.mod2HC2 = coeftest(ols.mod2, vcov = vcovHC(ols.mod2, type = "HC2"))[,2]

stargazer(ols.mod1, ols.mod2,
          se = list(ols.mod1HC2, ols.mod2HC2),
          no.space=T, omit.stat = c("f","ser"), title="Naive OLS estimates")
```

Using OLS, we find a positive association between average expropriation protections and log GDP per capita, significant at convential levels. Of course, we should be cautious about interpreting these findings as causal. It seems very likely that there would be ommitted variables that simultaneously affect both levels of expropriation risk protection and levels of log GDP per capita. Importantly, note that these concerns are not with the estimation strategy, but with the identification strategy – we don't believe that, even with infinite data, the observed data would necessarily map to a unique parameter (causal quantity of interest). That is, we have potential problems of observational equivalence.

Table 1: Naive OLS estimates

	Dependent variable:	
	logpgp95	
	(1)	(2)
avexpr	0.522***	0.401***
	(0.050)	(0.066)
lat_abst		0.875
		(0.628)
africa		−0.881***
		(0.154)
asia		−0.577*
		(0.307)
other		0.107
		(0.251)
Constant	4.660***	5.740***
	(0.322)	(0.396)
Observations	64	64
R^2	0.540	0.714
Adjusted R^2	0.533	0.689
Note:	*p<0.1; **p<0.05; ***p<0.01	

(d) Now, use the same two approaches as in part (c) and estimate the effect of `logem4` on `loggp95`. Interpret the direction and statistical significance of the estimate of the causal effect. What does this "reduced form" estimator purport to estimate? Under what conditions can we interpret this result as causal?

```
ols.mod3 = lm(logpgp95 ~ logem4, d)
ols.mod3HC2 = coeftest(ols.mod3, vcov = vcovHC(ols.mod3, type = "HC2"))[,2]
ols.mod4 = lm(logpgp95 ~ logem4 + lat_abst + africa + asia + other, d)
ols.mod4HC2 = coeftest(ols.mod4, vcov = vcovHC(ols.mod4, type = "HC2"))[,2]

stargazer(ols.mod3, ols.mod4,
          se = list(ols.mod3HC2, ols.mod4HC2),
          no.space=T, omit.stat = c("f","ser"),
          title="Reduced form or intent-to-treat OLS estimates")
```

Our results from this exercise show that higher levels of settler mortality are associated with lower GDP per capita – that is, there is a negative association betwee the two, significant at convential levels. This exercise can be seen as an attempt to estimate the "intent-to-treat" effect of settler mortality on log GDP per capita. That is, we are estimating the reduced form – the total effect of Z_i on Y_i. Of course, we can only interpret this as causal if (1) we

Table 2: Reduced form or intent-to-treat OLS estimates

	Dependent variable:	
	logpgp95	
	(1)	(2)
logem4	−0.573***	−0.377***
	(0.074)	(0.145)
lat_abst		1.050
		(0.886)
africa		−0.723***
		(0.262)
asia		−0.525
		(0.382)
other		0.185
		(0.257)
Constant	10.700***	10.000***
	(0.385)	(0.767)
Observations	64	64
R^2	0.477	0.584
Adjusted R^2	0.469	0.548
Note:	*p<0.1; **p<0.05; ***p<0.01	

have SUTVA, and (2) settler mortality is ignorable with respect to potential outcomes of log GDP per capita. Finally, do note that this is not always a particularly useful or interesting quantity – do we really want to estimate the effect of log settler mortality on log GDP per capita?

(e) Finally, use the `ivreg` function in the `AER` package to estimate the complier local average treatment effect (LATE) of `avexpr` on `loggp95`, with `logem4` used as the instrument for `avexpr`. As in (c) and (d), first include no covariates, and second include linearly and additively `lat_abst`, `africa`, `asia`, and `other`. In your result, be sure to report and interpret the F-statistic from a test for weak instrumentation, which you can retrieve from an output of `ivreg`. What do you find?

```r
library(AER)
#lm1 = lm(avexpr ~ logem4, d)
#lm2 = lm(avexpr ~ logem4+lat_abst+africa+asia+other, d)
#stargazer(lm1,lm2, no.space=T, title = "First stage estimates using lm")
ivmod1 = ivreg(logpgp95~avexpr | logem4, data=d)
ivmod2 = ivreg(logpgp95~avexpr+lat_abst+africa+asia+other
               | logem4+lat_abst+africa+asia+other, data=d)
```

```r
sum.iv1 = summary(ivmod1, diagnostics=T)
sum.iv2 = summary(ivmod2, diagnostics=T)

stargazer(ivmod1, ivmod2,
          add.lines = list(c("F-statistic",
                             round(sum.iv1$diagnostics[1, "statistic"], 2),
                             round(sum.iv2$diagnostics[1, "statistic"], 2))),
          no.space=T, omit.stat = c("f","ser"),
          title = "IV estimates using ivreg function")
```

Table 3: IV estimates using ivreg function

	Dependent variable:	
	logpgp95	
	(1)	(2)
avexpr	0.944***	1.110**
	(0.157)	(0.464)
lat_abst		−1.180
		(1.750)
africa		−0.437
		(0.424)
asia		−1.050*
		(0.525)
other		−0.990
		(0.998)
Constant	1.910*	1.440
	(1.030)	(2.840)
F-statistic	22.95	3.46
Observations	64	64
R^2	0.187	0.011
Adjusted R^2	0.174	−0.074
Note:	*p<0.1; **p<0.05; ***p<0.01	

Fortunately, our point estimates are exactly the same, which shows that our hand-rolled function is producing the correct IV estimate. Interestingly, in this case we can see that the F-test for weak instrumentation (i.e., does our instrument have a first stage effect on our endogenous regressor) is very strong for the model with no covariates (22.95), but considerably lower for the model with covariates (3.46). The typical "rule of thumb" is that an F-statistic of under 5 suggests that we should be careful about the possibility of a weak instrument. So in this case, we should be wary of the model with covariate adjustment.

Problem 2

De Kadt and Rosenzweig, graduate students at MIT, investigate partisan ties and their effects on national resource allocation in Ghana. The specific research question of interest is whether constituencies in Ghana that elect MPs who are from the same party as the President (the "ruling party") receive more electrification over the next four years. Using electoral data from the 1996 parliamentary elections and nightlights data the authors use a sharp regression discontintuity (RD) design to investigate the effect of the treatment (an MP from the ruling party winning the 1996 election) on the outcome (change in nightlights over the next four years). The forcing variable the authors use is voteshare of the ruling party MP candidate. The unit of analysis is the constituency. In this problem you will similarly conduct a sharp RD analysis using the dataset `Ghana_RD.csv`.

The dataset contains 152 observations each corresponding to a constituency in Ghana, with the following variables:

- `constit`: Name of consistuency
- `voteshare`: Voteshare (% votes obtained) for the ruling party MP candidate
- `treatment`: The treatment status indicator (1 if the ruling party MP won the 1996 election in that constituency, 0 if the ruling party candidate lost)
- `changeNL_1996_2000`: Change in nightlights between 1996 and 2000
- `mean_1995`: Mean level of nightlights in 1995

(a) First, let's look at the data to see if sharp RD makes sense for our dataset. Plot treament (y-axis) as a function of the forcing variable (x-axis), where the forcing variable is the margin of victory/loss for the ruling party MP candidate. In other words, adjust the variable `voteshare` such that the cutpoint lies at 50%. Does it seem appropriate to use a sharp regression discontinuity design in this case?

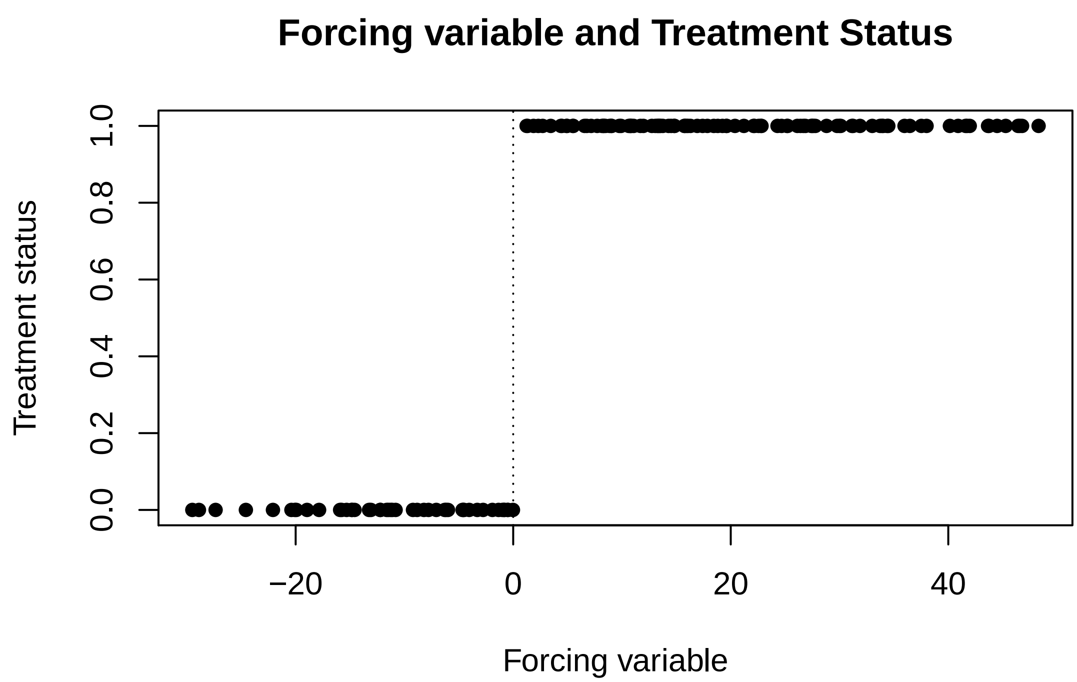
The plot reveals that our forcing variable does fully determine treatment status. Ruling party MP candidates who receive greater than 50% of the voteshare do indeed win the election and those who do not lose. It seems like a sharp regression discontinuity design is appropriate.

```
library(foreign)
library(rdd)
library(xtable)
library(stargazer)

#load data
d<-read.csv("Ghana_Rd.csv")

# code forcing variable so threshold at 50%
# in other words = margin of victory
d$forcing<-d$voteshare-50
```

```r
# make plot of treatment status and forcing variable
plot(d$forcing,d$treatment, main="Forcing variable and Treatment Status",
     xlab="Forcing variable",ylab="Treatment status",pch=16)
abline(v=0,lty=3)
```



(b) Estimate the local average treatment effect (LATE) at the threshold using a linear model with common slopes for treated and control units. What are the additional assumptions required for this estimation strategy? Provide a plot of the change in nightlights from 1996 to 2000 (y-axis) and forcing variable (x-axis for the range of -30 to 50) in which you show the fitted curves and the underlying scatterplot of the data. Interpret your resulting estimate.

The assumptions required for this strategy are $E[Y_i(0)|X_i = x]$ is linear in x and treatment effect τ does not depend on X_i.

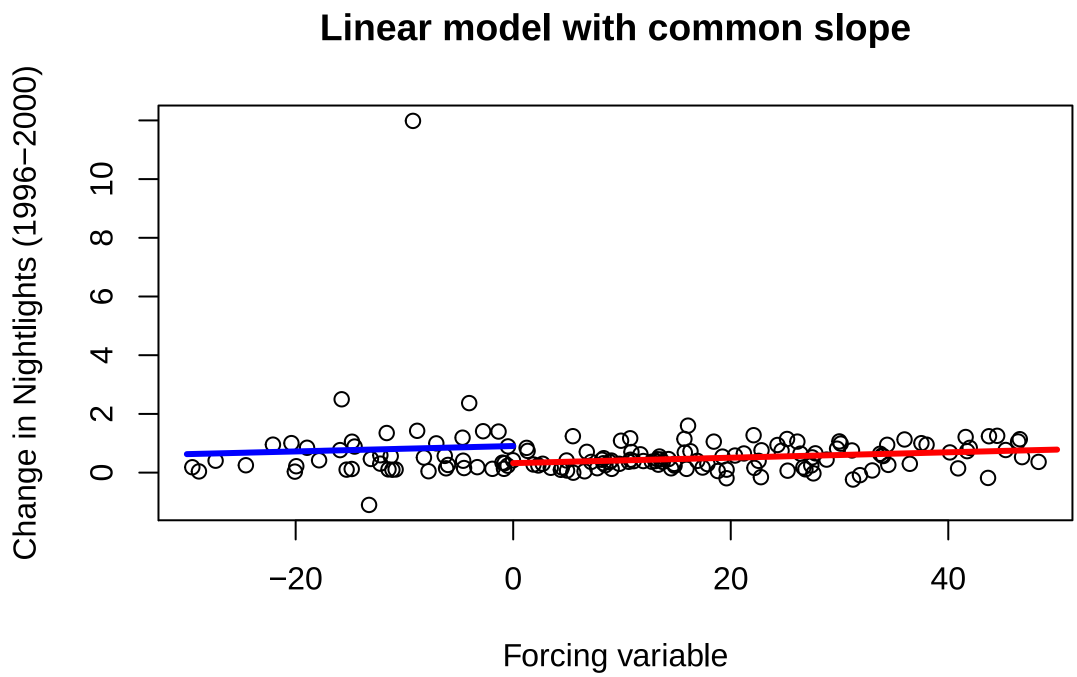
```r
library(lmtest)
#recode outcome variable name
d$Y<-d$changeNL_1996_2000

mod1<-lm(Y~treatment+forcing,d)
mod1.se<-coeftest(mod1,vcov=vcovHC(mod1,"HC2"))[,2]
stargazer(mod1,se=list(mod1.se),title="Linear common slope")
```

Table 4: Linear common slope

	Dependent variable:
	Y
treatment	−0.582*
	(0.306)
forcing	0.009***
	(0.003)
Constant	0.906***
	(0.272)
Observations	152
R^2	0.027
Adjusted R^2	0.014
Residual Std. Error	1.030 (df = 149)
F Statistic	2.040 (df = 2; 149)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
plot(d$forcing, d$Y, main="Linear model with common slope",
     xlab="Forcing variable", ylab="Change in Nightlights (1996-2000)")
curve(mod1$coefficient[1]+mod1$coefficient[2]+mod1$coefficient[3]*x,
      0, 50, add=T, lwd=3, col="red")
curve(mod1$coefficient[1]+mod1$coefficient[3]*x,
      -30, 0, add=T, lwd=3, col="blue")
```



The point estimate is -.58, which we can interpret as the Local Average Treatment Effect (LATE) of D on Y. Interpreting this estimate we see that moving from an opposition-won constituency to a ruling-party-won constituency in the 1996 election results in a .58 reduction in nightlights. The estimate is statistically significant at the 10% level. Remember the LATE in Sharp RDD is only really identified when $X = c$, so in general it's a good idea to decrease the size of the bandwidth around the threshold.

(c) Conduct the same analysis as in part (b) except that you should now use a linear model with *different* slopes for treated and control units.

The assumptions required for this strategy are $E[Y_i(0)|X_i = x]$ and $E[Y_i(1)|X_i = x]$ are both linear in x and now the treatment effect τ can vary with X_i.

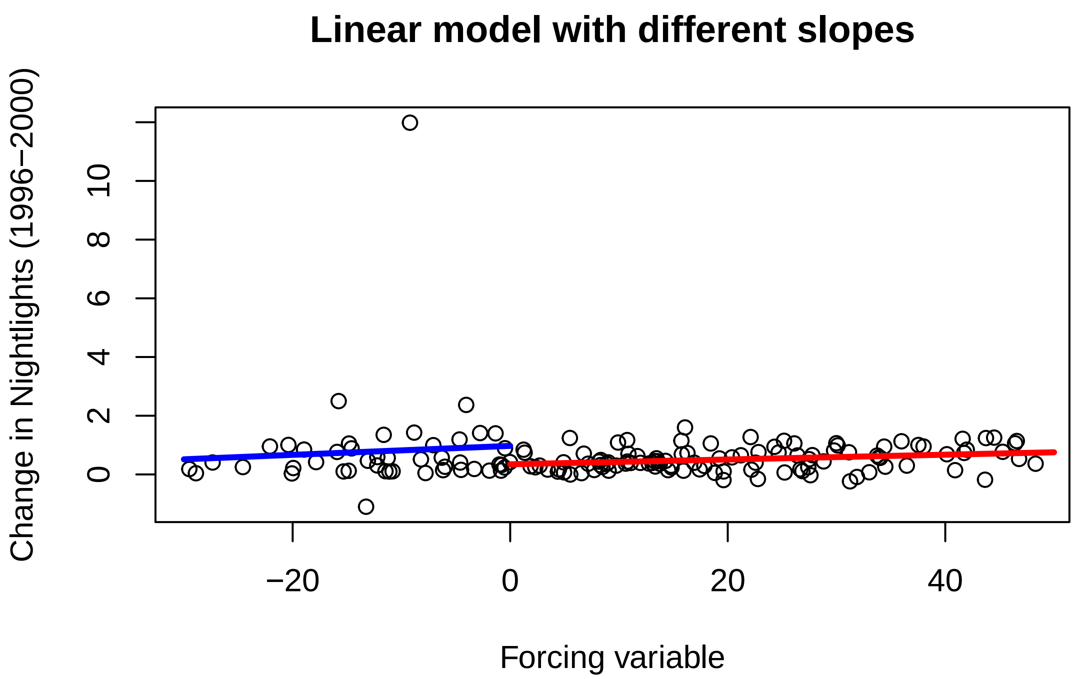
```
mod2<-lm(Y~treatment*forcing,d)
mod2.se<-coeftest(mod2,vcov=vcovHC(mod2,"HC2"))[,2]
stargazer(mod2,se=list(mod2.se),title="Linear different slopes")
```

```
plot(d$forcing, d$Y, main="Linear model with different slopes",
     xlab="Forcing variable", ylab="Change in Nightlights (1996-2000)")
curve(mod2$coefficient[1]+mod2$coefficient[2]
      +mod2$coefficient[3]*x+mod2$coefficient[4]*x,
      0, 50, add=T, lwd=3, col="red")
```

Table 5: Linear different slopes

	Dependent variable:
	Y
treatment	-0.627^{*}
	(0.361)
forcing	0.015
	(0.012)
treatment:forcing	-0.007
	(0.013)
Constant	0.972^{***}
	(0.355)
Observations	152
R^2	0.027
Adjusted R^2	0.008
Residual Std. Error	1.040 (df = 148)
F Statistic	1.390 (df = 3; 148)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
curve(mod2$coefficient[1]+mod2$coefficient[3]*x,
      -30, 0, add=T, lwd=3, col="blue")
```



The point estimate is essentially unchanged from the previous model. It is -.63 and is statistically significantly at the 10% level. In this case switching from an opposition MP to a ruling party MP candidate winning the 1996 election results in a .63 decrease in nightlights over the next four years.

(d) Conduct the same analysis as in part (b) except that you should now use a quadratic model with different model coefficients for the treated and control groups.

In this case $E[Y_i(0)|X_i = x]$ and $E[Y_i(1)|X_i = x]$ are allowed to be non-linear in x, but must be correctly specified to provide unbiased estimates of the LATE. The treatment effect τ can vary with X_i.

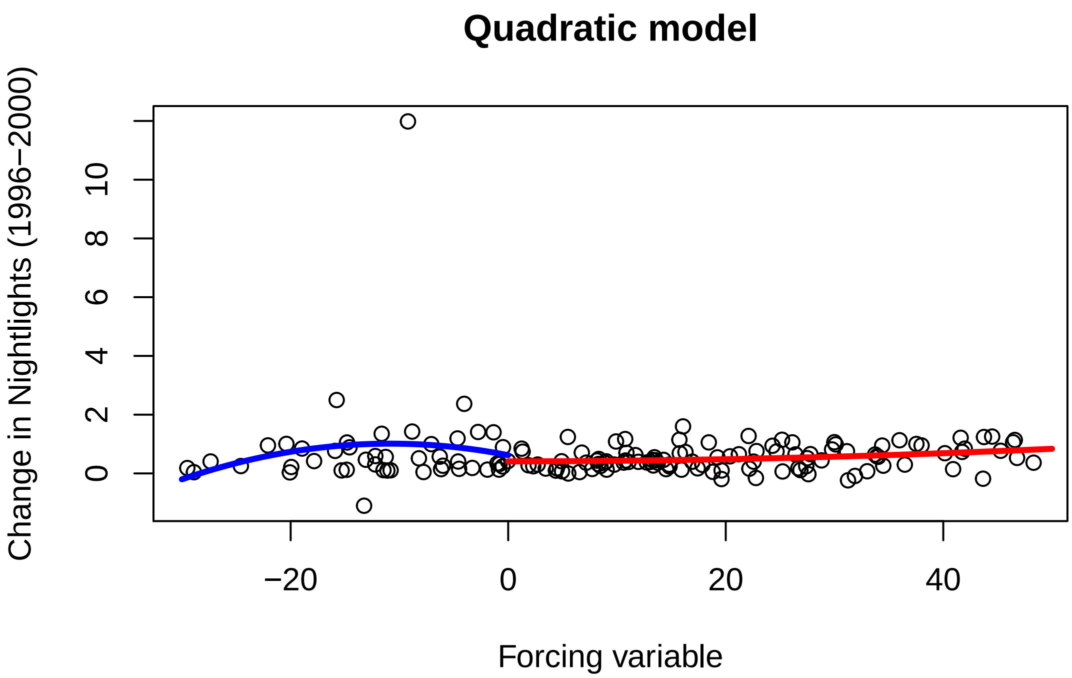
```
mod3<-lm(Y~treatment+forcing+I(treatment*forcing)
         +I(forcing^2)+I(treatment*(forcing^2)), d)
mod3.se<-coeftest(mod3,vcovHC(mod3,"HC2"))[,2]
stargazer(mod3,se=list(mod3.se),title="Quadratic model")
```

```
plot(d$forcing, d$Y, main="Quadratic model",
     xlab="Forcing variable", ylab="Change in Nightlights (1996-2000)")
```

Table 6: Quadratic model

	Dependent variable:
	Y
treatment	−0.212
	(0.221)
forcing	−0.072
	(0.069)
I(treatment ∗forcing)	0.073
	(0.070)
I(forcingˆ2)	−0.003
	(0.003)
I(treatment ∗(forcingˆ2))	0.003
	(0.003)
Constant	0.620***
	(0.195)
Observations	152
R^2	0.043
Adjusted R^2	0.011
Residual Std. Error	1.040 (df = 146)
F Statistic	1.330 (df = 5; 146)
Note:	*p<0.1; **p<0.05; ***p<0.01

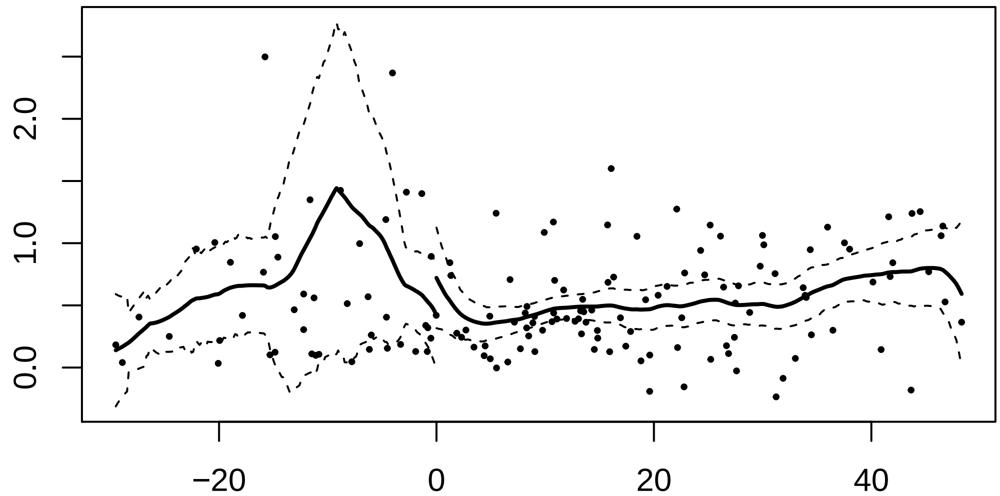
```r
curve(mod3$coefficient[1]+mod3$coefficient[2]
      +mod3$coefficient[3]*x+mod3$coefficient[4]*x
      +mod3$coefficient[5]*(x^2)+mod3$coefficient[6]*(x^2),
      0, 50, add=T, lwd=3, col="red")
curve(mod3$coefficient[1]+mod3$coefficient[3]*x
      +mod3$coefficient[5]*(x^2),
      -30, 0, add=T, lwd=3, col="blue")
```



The point estimate is now smaller (-.21) with a very large p-value. In this case switching from an opposition MP to a ruling party MP candidate winning the 1996 election results in a .21 decrease in nightlights over the next four years.

(e) Use the `rdd` package in `R` to estimate the LATE at the threshold using a local linear regression with a triangular kernel. Note that the function `RDestimate` automatically uses the Imbens-Kalyanamaran optimal bandwidth calculation. Report your estimate for the LATE and an estimate of uncertainty.

```r
library(rdd)
llrmod<-RDestimate(Y~forcing,d,se.type="HC2")
plot(llrmod)
```



The LATE estimate is now completely different. The estimate is now positive, .30, and is not statistically significant, with a standard error of .27. This local linear regression estimate suggests that switching from an opposition MP to a ruling party MP winning the 1996 election results in a .30 *increase* in nightlights over the next four years.

(f) How do the estimates of the LATE at the threshold differ based on your results from parts (b) to (e)? In other words, how robust are the results to different specifications of the regression? What other types of robustness checks might be appropriate?

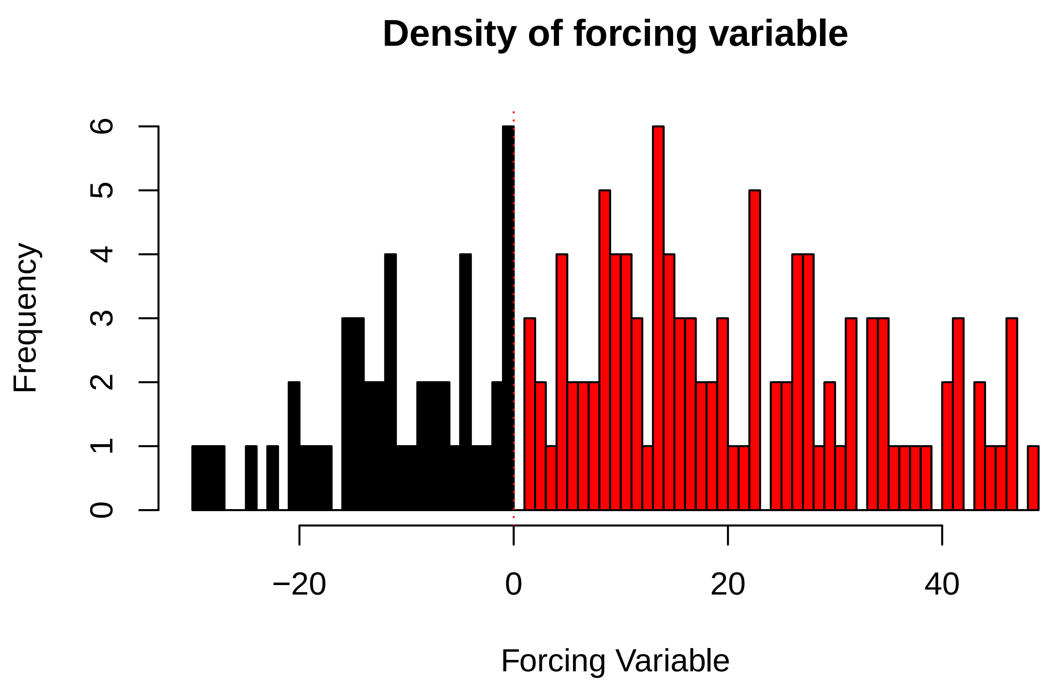
The estimates of the LATE seem to vary based on the functional form used. The way we estimate the conditional expectation function of Y_1 and Y_0 is going to significantly impact the results we get. Recall that the LATE is only identified when $X = c$; this means that we really want to pay special attention to getting the conditional expectation function right in the local area around c. So we don't want to draw too heavily on the rest of the joint distribution. Local linear regression with a triangular kernel (the last estimator we use) is the current "best practice" for getting the conditional expectation right—this is the estimate we should trust since we did not limit the bandwidth around the threshold in our other estimation procedures.

Other robustness checks we might want to conduct are examining continuity in covariates and the forcing variable around the threshold c. We could also conduct placebo regressions. We could also estimate the LATE for different sized bandwidths around the threshold.

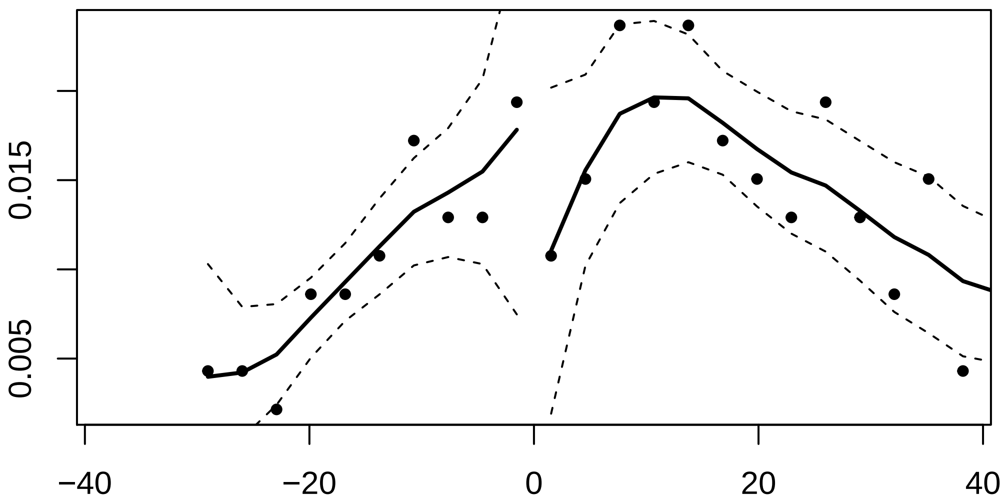
(g) Conduct one such test by examining the density of the forcing variable around the cutoff.

First, create a histogram plot of the forcing variable (binned by 1 percentage points). Second, conduct a formal test of the difference in density of the forcing variable around the cutoff by using McCrary's test (`DCdensity` function in the `rdd` package will be helpful) and report the p-value from the test. Why is this analysis a good diagnosis of the identification assumption for the sharp RD design? What can you say about the plausibility of no manipulation in voteshare?

```
h<-hist(d$forcing,breaks=100,plot=F)
cuts<-cut(h$breaks,c(-Inf,0,Inf))
plot(h,col=cuts,main="Density of forcing variable",
     xlab="Forcing Variable")
abline(v=0,lty=3,col="red")
```




```
# conduct McCrary formal test
DCdensity(d$forcing, 0, plot = TRUE)
```



[1] 0.195

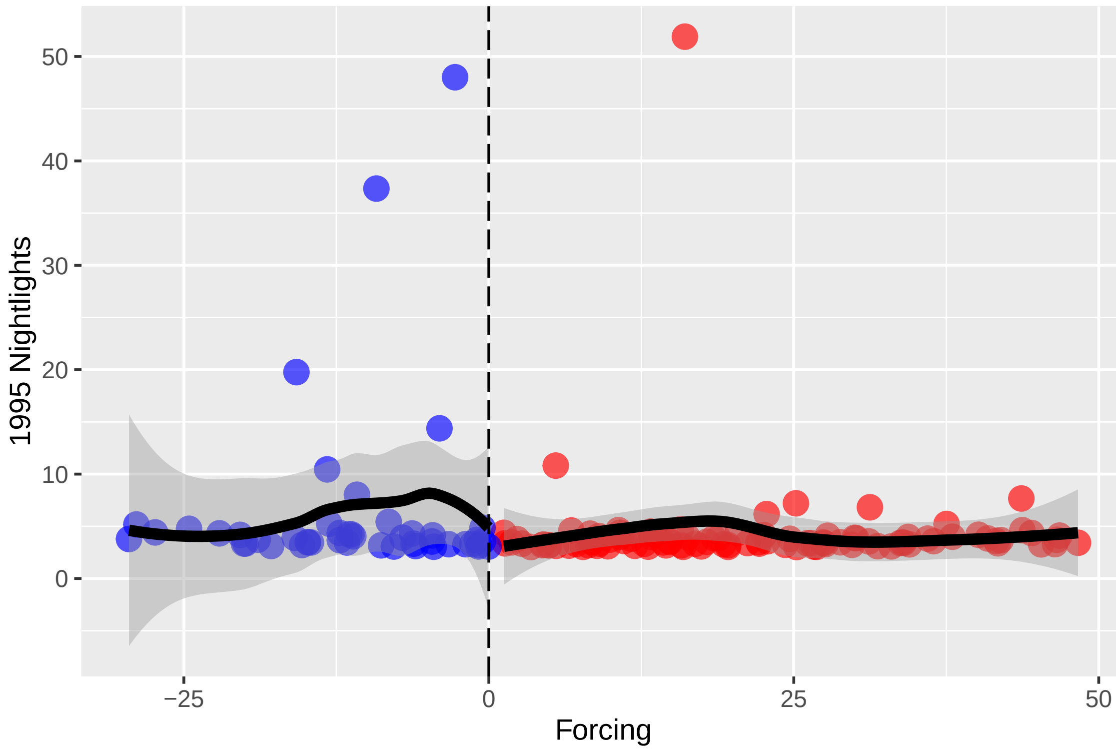
This test serves as a good diagnosis because continuity assumption would be less plausible if units (MPs in each constituency) have the ability to precisely manipulate their voteshare. One signal of this type of manipulation would be if the number of treated observations just above the cutoff (MPs that narrowly win) is very different from the number of control observations below it (MPs that narrowly lose). First, from the histogram we can see that there are no observations where a MP candidate won by between 50% and 51% of the vote (e.g. where the forcing variable or margin of victory is between 0 and 1). This would generally be a cause for concern but due to our small sample size might just be a weird feature of the sample. Second, there seem to be several observations just below the cutoff threshold. This is weird because these are the margins of victory/loss for ruling party MP candidates and it wouldn't make much sense for these candidates, or their party, to manipulate the election such that they narrowly *lose*. Finally, we gain a bit more confidence in the lack of manipulation from our formal McCrary test, which produces an insignificant p-value of .195.

(h) Finally conduct a placebo test using nightlights measured in 1995 as the outcome in a sharp RD analysis for the 1996 election. Use local linear regression as you did in part (e). Also graph the points for the forcing variable (x-axis) and 1995 nightlights (y-axis). What does this placebo test say about the relationship between the 1996 election of ruling party MPs and nightlights measured in 1995?

From this placebo regression we can see that, as expected, there is little (statistically significant) relationship between the 1996 election close winners and losers from the ruling party and 1995 nightlights. However, the LATE estimate from the local linear regression is large(!) -1.9 with a standard error of 2.55. We can see visually from the graph that there is no discontinuity in the data and the confidence intervals overlap so our large estimate may be due to outlying data points around the threshold.

```
# run regression
pla1<-RDestimate(mean_1995~forcing,d,se.type="HC2")

library(ggplot2)
#This line gives your data, then defines x and y.
ggplot(d, aes(x = forcing, y = mean_1995)) +
#this adds the points for treated, with a particular size and color
  geom_point(data=d[which(d$forcing>.5 ),],aes(forcing, mean_1995),
             size=4, alpha=.65, col="red") +
#this adds the points for control
  geom_point(data=d[which(d$forcing<.5 ),], aes(forcing,mean_1995),
             size=4, alpha=.65, col="blue") +
#this adds a smoothing function for treated side of the disc.
#You can choose any method/form.
  stat_smooth(data = d[which(d$forcing>.5 ),], aes(forcing,mean_1995),
              method = "loess", se=T,colour = "black", level=.95, size = 2) +
#this adds a smoothing function for control side of the disc.
#You can choose any method/form.
  stat_smooth(data = d[which(d$forcing<.5 ),], aes(forcing,mean_1995),
              method = "loess", se=T,colour = "black", level=.95, size = 2) +
#note you can add CIs at whatever level you want.
# add x-axis label
  scale_x_continuous("Forcing") +
# add y-axis label
  scale_y_continuous("1995 Nightlights")+
# add vertical dashed line at 0
  geom_vline(xintercept = 0, linetype="longdash")
```



Optional Problems

Problem A

(a) Suppose that we are interested in the effect of a potentially endogenous causal variable X_i on an outcome variable of interest Y_i. Assume that we have another variable Z_i, which is binary and is an instrumental variable for X_i. Show that the IV estimator for the effect of X_i on Y_i

$$\hat{\beta}_{IV} = \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)}$$

Can be written as

$$\frac{(\bar{Y}_1 - \bar{Y}_0)}{(\bar{X}_1 - \bar{X}_0)},$$

where $\text{cov}(\cdot)$ is the sample covariance; $\bar{Y}_0$ and $\bar{X}_0$ are the sample averages of Y_i and X_i over the part of the sample with $Z_i = 0$; and $\bar{Y}_1$ and $\bar{X}_1$ are the sample averages of Y_i and X_i over the part of the sample with $Z_i = 1$.

We are given the (univariate) IV estimator $\hat{\beta}_{IV} = \frac{\text{cov}(Z_i,Y_i)}{\text{cov}(Z_i,X_i)}$. We are asked to show how this is equal to $\frac{(\bar{Y}_1 - \bar{Y}_0)}{(\bar{X}_1 - \bar{X}_0)}$. Let's start by multiplying the IV estimator by 1 to give us quantities we can do something with:

$$\begin{aligned}\hat{\beta}_{IV} &= \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)} \\ &= \frac{\text{cov}(Z_i, Y_i)/\text{cov}(Z_i, Z_i)}{\text{cov}(Z_i, X_i)/\text{cov}(Z_i, Z_i)} \\ &= \frac{\hat{\gamma}_1}{\hat{\pi}_1}\end{aligned}$$

Where $\hat{\pi}_1$ comes from $X_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i + u_{1i}$, and $\hat{\gamma}_1$ comes from $Y = \hat{\gamma}_0 + \hat{\gamma}_1 Z_i + u_{2i}$. Now we can replace $\hat{\gamma}_1$ and $\hat{\pi}_1$ with differences in conditional sample means:

$$\begin{aligned}\gamma_1 &= \frac{1}{N_1}\sum_i^{N_1} Y_i - \frac{1}{N_0}\sum_i^{N_0} Y_i \\ &= \bar{Y}_1 - \bar{Y}_0\end{aligned}$$

And similarly,

$$\begin{aligned}\pi_1 &= \frac{1}{N_1}\sum_i^{N_1} X_i - \frac{1}{N_0}\sum_i^{N_0} X_i \\ &= \bar{X}_1 - \bar{X}_0\end{aligned}$$

Plugging these in, we arrive at

$$\frac{(\bar{Y}_1 - \bar{Y}_0)}{(\bar{X}_1 - \bar{X}_0)}.$$

(b) Let $\mathbf{X} = [1, X_1, X_2, \cdots, X_k, D]$ and $\mathbf{Z} = [1, X_1, X_2, \cdots, X_k, Z]$. The matrix $\mathbf{X}$ contains the covariates (including a vector of 1s) and your treatment vector D, and $\mathbf{Z}$ is a matrix of the same covariates and the instrument for the treatment variable in place of the actual treatment. Y is a vector of observed outcomes. We can construct the following system of linear equations, with error terms u_2 and u_1 respectively:

$$\begin{aligned}Y &= \mathbf{X}\beta + u_2 \\ D &= \mathbf{Z}\pi + u_1\end{aligned}$$

with coefficient vectors $\beta = [\beta_0, \beta_1, \cdots, \beta_k, \beta_D]$ and $\pi = [\pi_0, \pi_1, \cdots, \pi_k, \pi_Z]$.

The IV estimator can be obtained by:

$$\hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'Y$$

i) What are the conditions under which the treatment effect estimate $\hat{\beta}_{D,IV}$ is consistent?

$$\begin{aligned}
\text{plim}(\hat{\beta}_{IV}) &= \text{plim}((\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'Y)) \\
&= \text{plim}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{X}\beta + \text{plim}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'u_2 \\
&= \beta + \text{plim}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'u_2
\end{aligned}$$

This goes to zero only if $\text{plim}(\mathbf{Z}'u_2) = 0$, which occurs when $cov(\mathbf{Z}, u_2) = 0$. That is, the instruments – including the covariates other than D that are used as instruments for themselves, must be "exogenous".

ii) You can obtain the Two-stage Least Squares estimator by the following steps.

(a) Run the first stage regression: $D = \mathbf{Z}\pi + u_1 \Rightarrow \hat{\pi} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'D$

(b) Get fitted values: $\hat{D} = \mathbf{Z}\hat{\pi}$

(c) Regress Y on $\hat{\mathbf{X}} = [1, X_1, X_2, \cdots, X_k, \hat{D}]$: $Y = \hat{\mathbf{X}}\beta_{2SLS} + u_3$

Show formally that $\hat{\beta}_{2SLS} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'Y = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'Y = \hat{\beta}_{IV}$. Comment on the steps along the way to reach your conclusion. If you need any additional assumptions, please state them.

The trick here is to realize the $\hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$. The action of $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})\mathbf{Z}'$ on $\mathbf{X}$ is to fit each column of $\mathbf{X}$ using the columns of $\mathbf{Z}$ (this is why $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})\mathbf{Z}'$ is sometimes called the projection matrix $P_{\mathbf{Z}}$). For the X's that are included as themselves in $\mathbf{Z}$, the best combination of Z's that predict those X's are just the original X's, so they are replicated exactly in $\hat{\mathbf{X}}$. For the treatment, D, it is replicated by a linear combination of the instruments, so you get $\hat{D} = \mathbf{Z}\pi$.

Plugging $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ in for $\hat{\mathbf{X}}$ in $(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}(\hat{\mathbf{X}}'Y)$ you get the desired results, $(\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'Y)$:

$$\begin{aligned}
(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}(\hat{\mathbf{X}}'Y) &= (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Y) \\
&= (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Y) \\
&= (\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'Y).
\end{aligned}$$

Problem B

An alternative way of thinking about regression discontinuity designs is to think of them as "locally randomized" experiments. Consider the following setup:

$$Y = \tau D + \delta_1 W + U$$

$$
\begin{aligned}
D =& 1[X \geq c] \\
X =& \delta_2 W + V
\end{aligned}
$$

where Y is the dependent variable, $D \in \{0, 1\}$ is the treatment indicator, X is the forcing variable, and c is the cutpoint. V and U are unobserved variables, and W is a set of predetermined and observable characteristics. Making no further assumptions about the correlations between V, U, and W, the identification assumption for the RDD under this approach can be stated as:

$$\Pr(W = w, U = u | X = x) \text{ is continuous in } x.$$

(**Hint:** Refer to sections 2 and 3 of Lee & Lemieux (2010) for more detail about this setup and the identification assumption.)

(a) Interpret the identification assumption given above, in light of the setup. Why does it represent "local randomization"? Intuitively, why does this assumption allow us to identify the causal estimand, the LATE at the discontinuity threshold?

The assumption above states that the probability of observing any pair of values of W and U, w and u, conditional on values of X, is smooth over X. That is, it rules out any discontinuous jump in the values of W (observed covariates) and U (unobserved covariates) in the support of X. This represents "local randomization" because, in the local space around the cutoff, we assume that any pair of values of W and U does not perfectly predict treatment status. Thus we break the dependence between the covariates/confounders and potential outcomes, as in randomization. Intuitively, this means that we can treat the local area around the cutpoint as a randomized experiment – we have balance in both observed and unobserved covariates across treatment and control.

(b) Prove that the LATE at the threshold can be identified under the above assumption as:

$$\tau_{RDD} = \lim_{\eta \downarrow 0} \mathbb{E}[Y \mid X = c + \eta] - \lim_{\eta \uparrow 0} \mathbb{E}[Y \mid X = c + \eta]$$

where η is a real number in the range of X.

$$
\begin{aligned}
\tau_{RDD} =& \lim_{\eta \downarrow 0} \mathbb{E}[Y \mid X = c + \eta] - \lim_{\eta \uparrow 0} \mathbb{E}[Y \mid X = c + \eta] \\
=& \tau + \lim_{\eta \downarrow 0} \sum_{w,u} (\delta_1 w + u) \cdot \Pr[W = w, U = u \mid X = c + \eta] \\
& - \lim_{\eta \uparrow 0} \sum_{w,u} (\delta_1 w + u) \cdot \Pr[W = w, U = u \mid X = c + \eta] \\
=& \tau + \Pr[W = w, U = u \mid X = c + 0] \Big[\lim_{\eta \downarrow 0} \sum_{w,u} (\delta_1 w + u) \\
& - \lim_{\eta \uparrow 0} \sum_{w,u} (\delta_1 w + u) \Big] \because \Pr(W = w, U = u | X = x) \text{ is continuous in } x. \\
=& \tau + \Pr[W = w, U = u \mid X = c] \Big[0 \Big] \;=\; \tau
\end{aligned}
$$