PS3 - Boseong Yun

Boseong Yun

11/03/2020

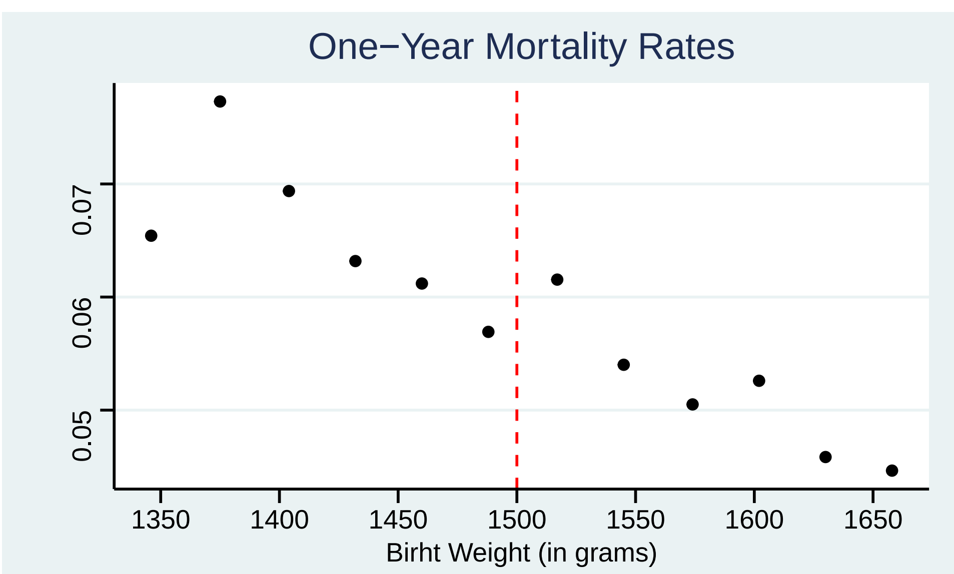
1) Start by getting the descriptive statistics of birth weight in the sample, what is the mean, standard deviation, minimum, and maximum?

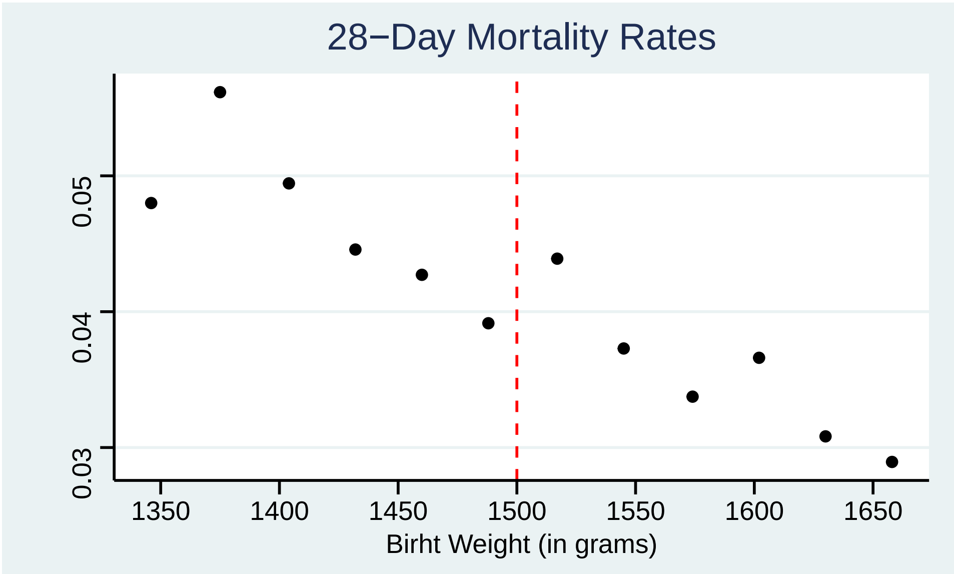
Table 1: The Descriptive Statistics of Birth Weight

mean	sd	min	max
1511.576	89.01614	1350	1650

2) Now plot one year and 28 day mortality rates against our running variable, birth weight. To do so, make bins of one ounce (28.35 grams) around the 1500 grams threshold, and get the mean mortality rate in each bin. Make a separate graph for each outcome. Describe the relationship between birth weight and mortality. Does it appear to be a discontinuity of mortality around the very low birth weight threshold? How does the number of observations in each bin affect your mean estimates?

Answer: Generally, there seems to be negative assocation between birth weight and mortality. Specifically, increase in body weight is negatively associated with mortality rates. Additionally, the graph shows that there appears to be a discontinutiy of mortality around the low birth weight threshold of 1500 grams. There is a sharp decrease in the mortality rates right before the threshold. The number of observation affects the bias and variance of the estimates. Specifically, having a smaller bindwith will have less number of observation in each bin and will thus lessen the bias but increase the variance of the estimates. One should carefully consider the bias-variance tradeoff when setting binwidth.





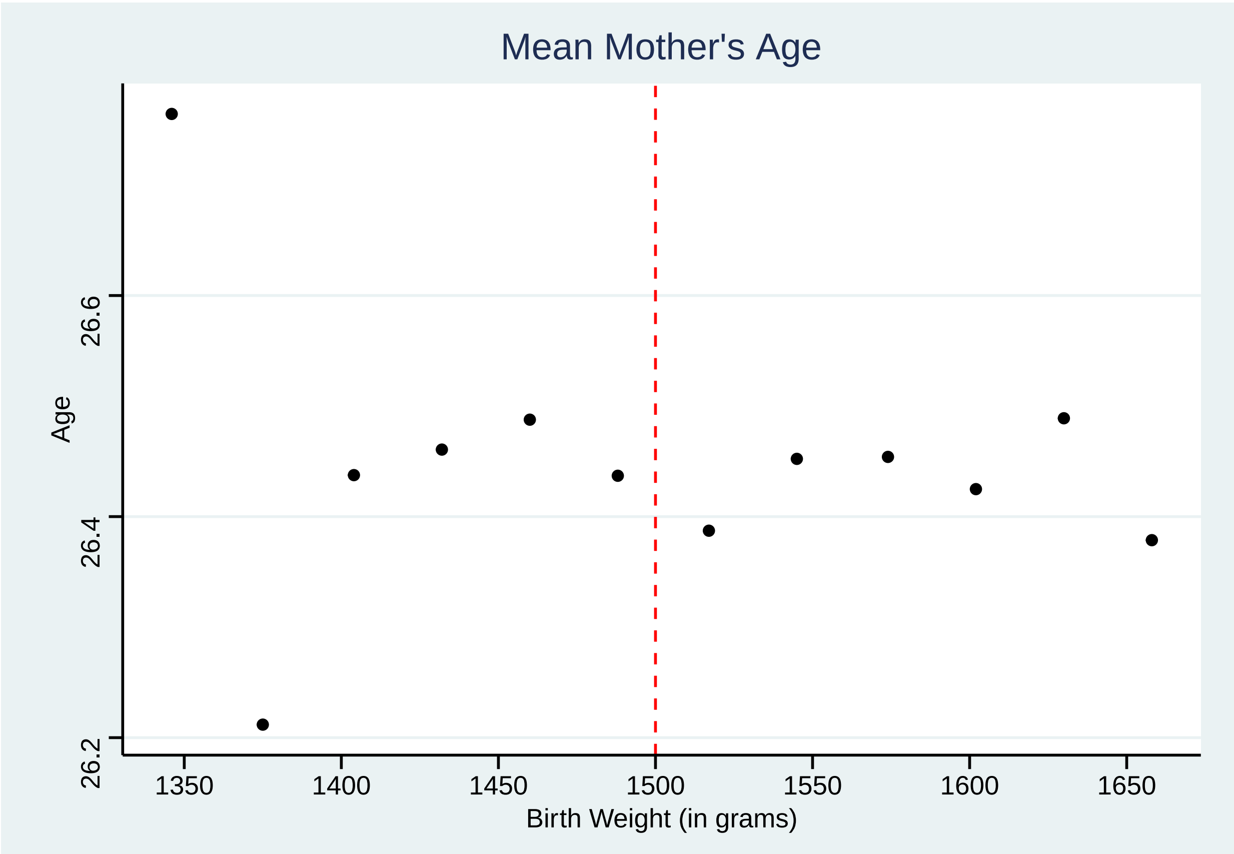
3) A key assumption for an RDD to provide a causal estimate is that individuals are not able to sort according to the running variable, i.e., they should not be able to manipulate its value. Discuss in your own words whether this is a reasonable assumption in this case.

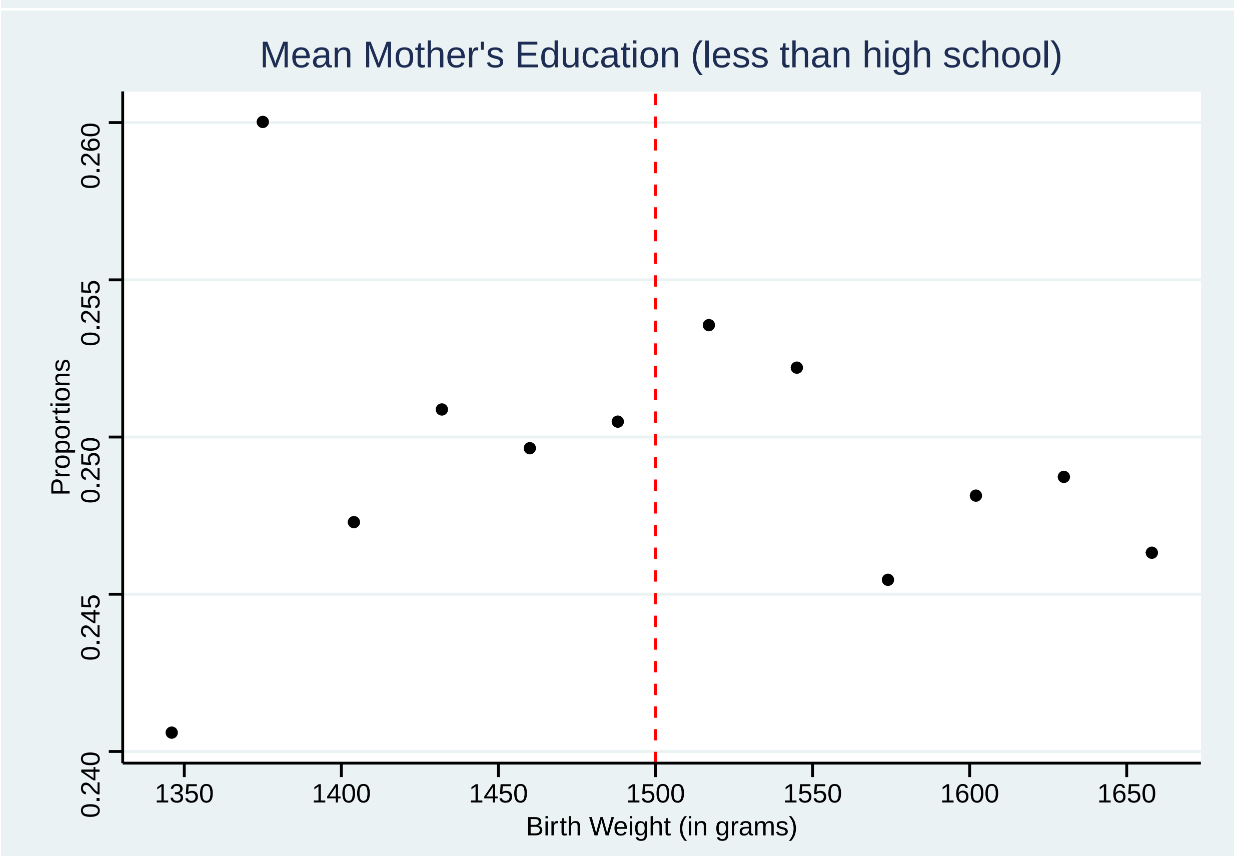
Answer:

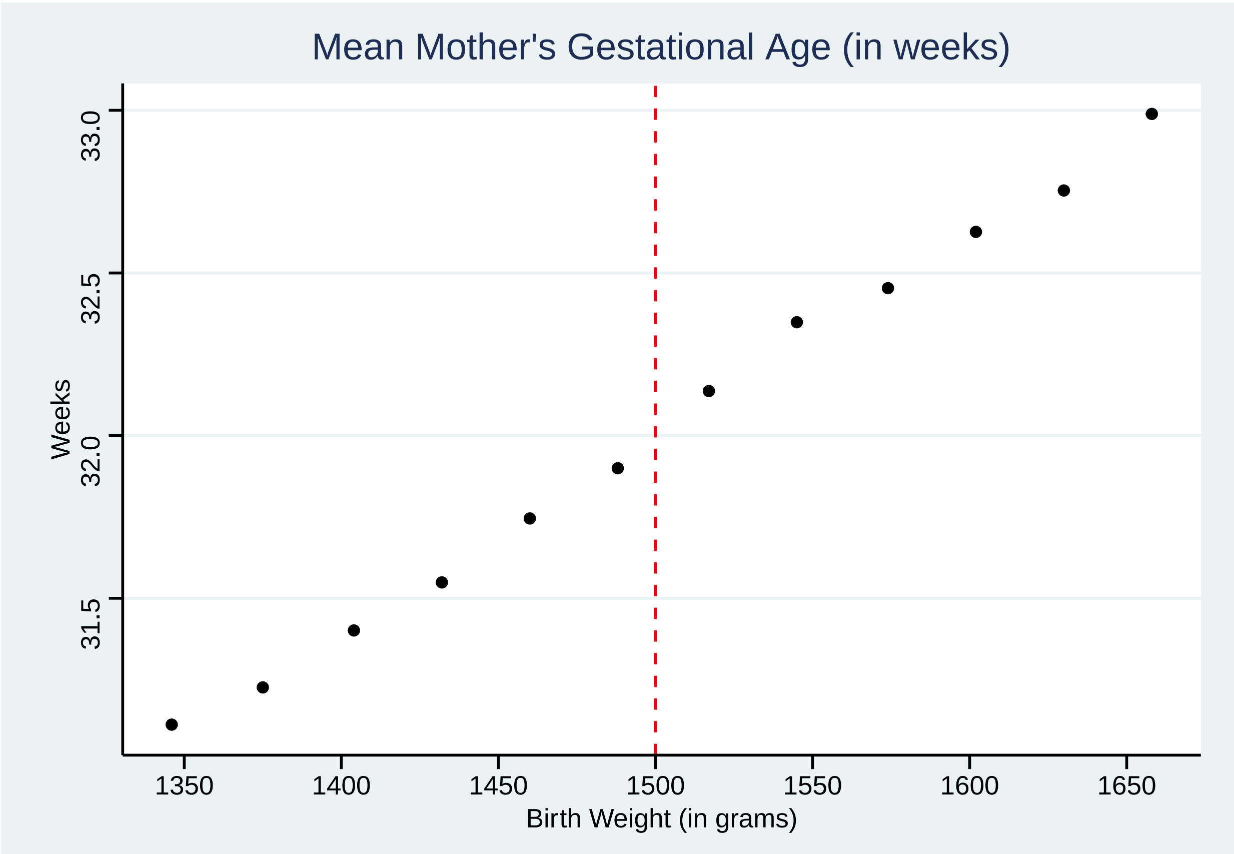
I think it is a reasonable assumption that birth weight, which is the running variable, cannot be manipulated. This is because the weight of the babies are biological in essence and are weighted in the hospitals, reducing the likelihood of manipulating the value. It could be possible, however, that some parents can try to manipulate the weight of the babies when there are government progrmas that provide assistance to the parents of under-weight babies. This may be hard to observe in practice because parents wouldn't intentionally under-feed themselves to be eligible for the programs. This can be dangerous for both the babies and the parents. Therefore, I believe that this is a reasonable assumption.

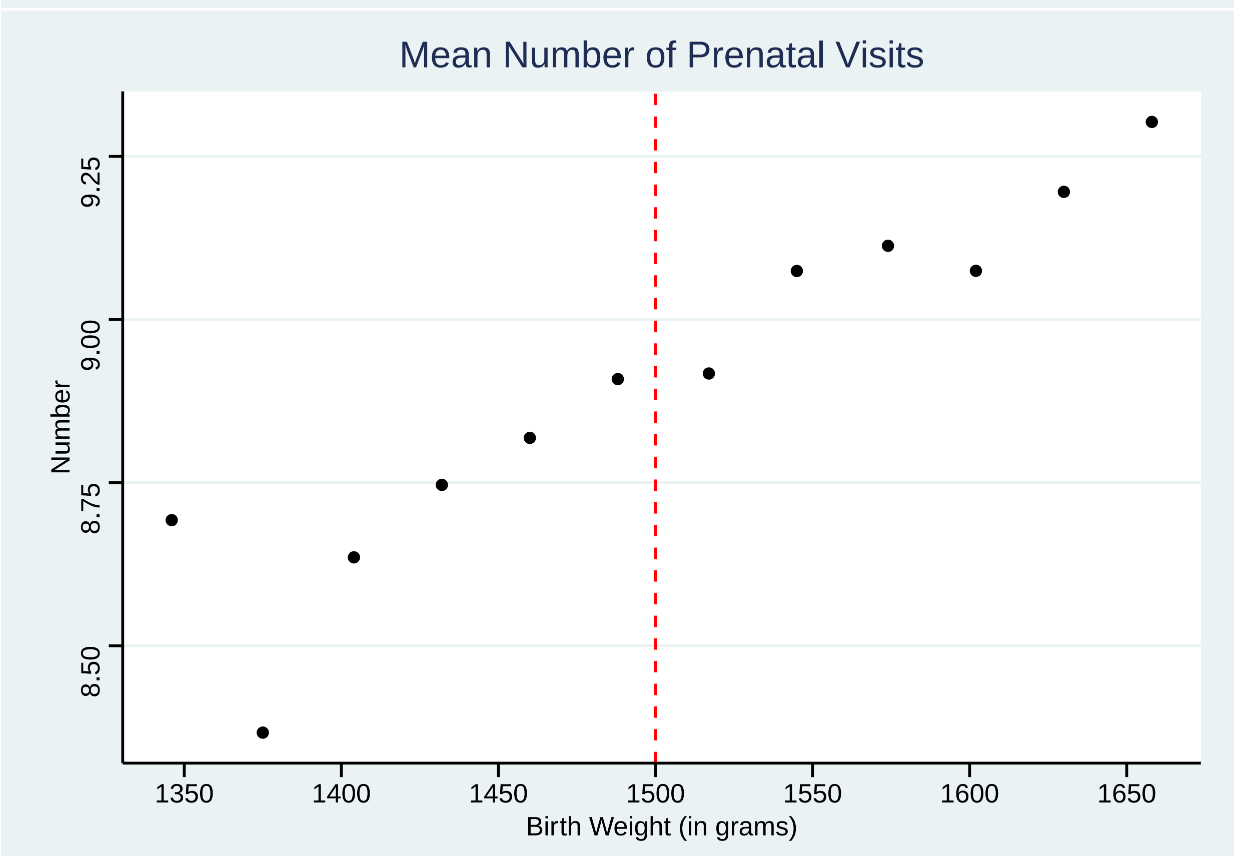
4) Now plot background covariates, including mother's age, mother's education less than high school, gestational age, prenatal care visits, and year of birth, against birth weight as you did in point (2). Do these plots appear to be smooth around the threshold?

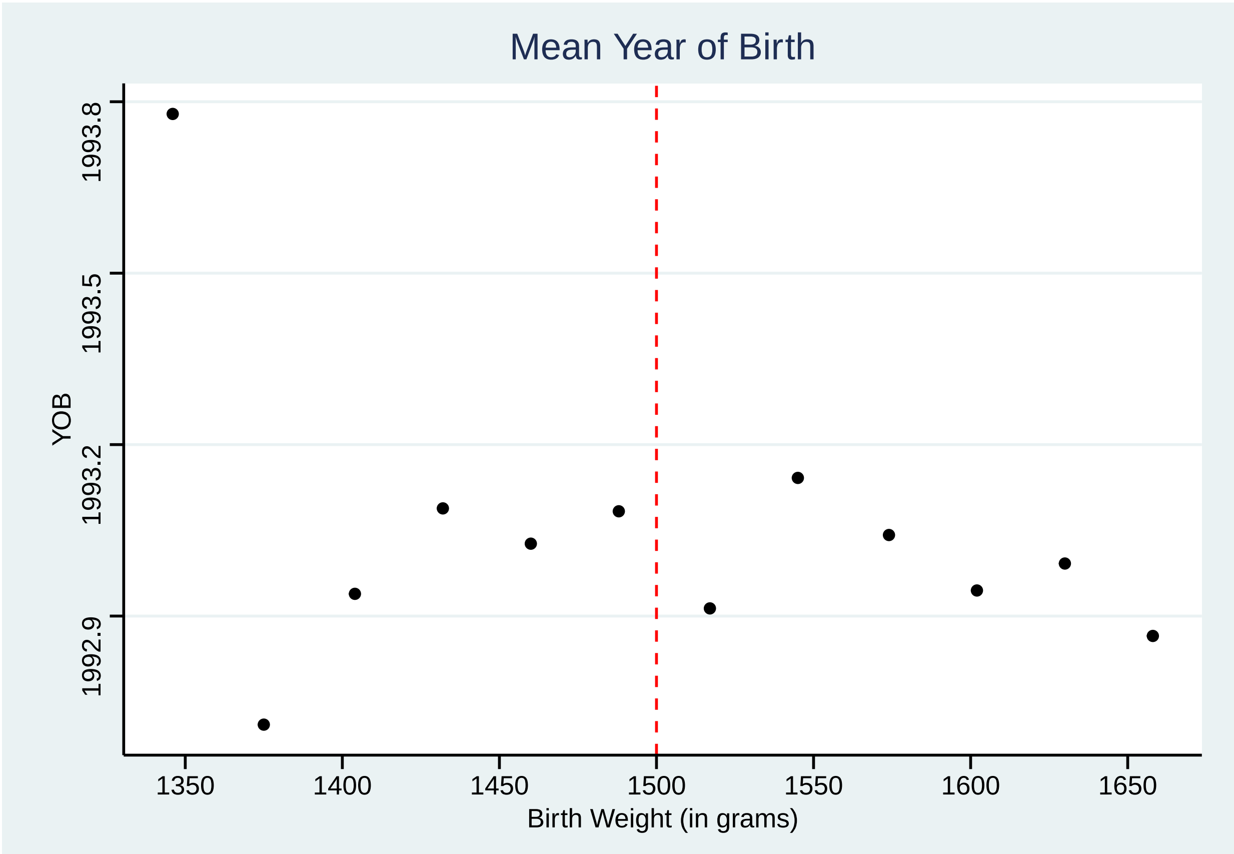
Answer: Although most of the plots appear to be smooth around the threshold, the Year of Birth variable is one that does not appear to be smooth around the threshold among these five background covariates.


Mean Mother's Age
Age
26.2
26.4
26.6
1350
1400
1450
1500
1550
1600
1650
Birth Weight (in grams)


Mean Mother's Education (less than high school)
Proportions
0.240
0.245
0.250
0.255
0.260
1350
1400
1450
1500
1550
1600
1650
Birth Weight (in grams)


Mean Mother's Gestational Age (in weeks)
Weeks
33.0
32.5
32.0
31.5
1350
1400
1450
1500
1550
1600
1650
Birth Weight (in grams)


Mean Number of Prenatal Visits
Number
9.25
9.00
8.75
8.50
1350
1400
1450
1500
1550
1600
1650
Birth Weight (in grams)



5) Now formalize the evidence on smoothness of the covariates by estimating regressions of the form $B_i = \alpha_0 + \alpha_1 VLBW_i + \alpha_2 VLBW_i * (g_i - 1500) + \alpha_3(1 - VLBW_i) * (g_i - 1500) + \varepsilon_i$

Where B_i is a background covariate, $VLBW_i$ indicates that a newborn had very low birth weight (<1500 grams), g_i is birth weight and ε_i a disturbance term. Use a caliper of 85 grams (above and below the threshold). Which coefficient provides a test of smoothness is the vicinity of the very low birth weight threshold? Is there any evidence of discontinuities around the threshold? If they were, how could these affect your RDD estimates?

Answer: The coefficient for VLBM, α_1 in the problem set, provides a test for smoothness in the very low birth weight threshold.

The estimates for VLBM are statistically significant for Mom's Age, Gestatation, and Year of Birth at the 0.05 siginificance level within the calipers of 85 grams. These indiciate evidence of discontinuities around the threshold and break the smoothness assumption that posits there are no differences other than the treatment status in the observations within the calipers. With this assumption broken, my RDD estimates no longer identify the unbiased treatment effects.

Table 2: Coefficients Output for Background Covariates (85g calipers)

vars	term	estimate	std.error	statistic	p.value
mom_age	(Intercept)	26.310251	0.034402	764.794376	0.000000
mom_age	VLBM	0.241361	0.060065	4.018354	0.000059
mom_age	bweight_low	0.003486	0.000810	4.304304	0.000017
mom_age	VLBM:bweight_low	-0.001687	0.001203	-1.402650	0.160723
mom_ed1	(Intercept)	0.254670	0.002291	111.156759	0.000000

vars	term	estimate	std.error	statistic	p.value
mom_ed1	VLBM	-0.002602	0.004000	-0.650471	0.515389
mom_ed1	bweight_low	-0.000123	0.000054	-2.280884	0.022556
mom_ed1	VLBM:bweight_low	0.000158	0.000080	1.967773	0.049095
gest	(Intercept)	32.151642	0.017620	1824.747969	0.000000
gest	VLBM	-0.128477	0.030871	-4.161677	0.000032
gest	bweight_low	0.004670	0.000416	11.214702	0.000000
gest	VLBM:bweight_low	0.001115	0.000618	1.803289	0.071344
nprenatal	(Intercept)	8.903595	0.031298	284.480757	0.000000
nprenatal	VLBM	0.089424	0.054626	1.637014	0.101629
nprenatal	bweight_low	0.003797	0.000739	5.138047	0.000000
nprenatal	VLBM:bweight_low	-0.000455	0.001095	-0.415984	0.677422
yob	(Intercept)	1992.738350	0.032633	61064.300759	0.000000
yob	VLBM	0.589951	0.056977	10.354153	0.000000
yob	bweight_low	0.008469	0.000768	11.024666	0.000000
yob	VLBM:bweight_low	-0.003230	0.001141	-2.831078	0.004640

6) Now get an estimate of the size of the discontinuity in one—year and 28—day mortality, around the 1500 grams threshold. As above, use a caliper of 85 grams. To do so, estimate the following model: $Y_i = \alpha_0 + \alpha_1 VLBW_i + \alpha_2 VLBW_i * (g_i - 1500) + \alpha_3(1 - VLBW_i) * (g_i - 1500) + \varepsilon_i$ where Y_i is the outcome of interest Interpret the coefficients α_1 , α_2, and α_3.

Answer:

I have referred to Professor Grogger's **Piazza Post** for answering this question.

For 1-Year Mortality Rates: The coefficient VLBM, α_1 in the problem set, indicates the magnitude of discontinuity at the threshold with the statistically significant size of -0.009510. The coefficient α_2 indicates the slope of the line to the left of the threshold and the coefficient α_3 indicates the slope of the line to the right of the threshold.

For 28 Days Mortality Rates: The coefficient VLBM, α_1 in the problem set, indicates the magnitude of discontinuity at the threshold with the statistically significant size of-0.008781. The coefficient α_2 in the problem set indicates the slope of the line to the left of the threshold and the coefficient α_3 indicates the slope of the line to the right of the threshold.

Table 3: Mortality Rates for 1 Year and 28 Days (Covariates Uncontrolled)

var	term	estimate	std.error	statistic	p.value
1-Year (Calipers 85g) U/C	(Intercept)	0.063156	0.001233	51.211572	0.000000
1-Year (Calipers 85g) U/C	VLBM	-0.009510	0.002153	-4.416767	0.000010
1-Year (Calipers 85g) U/C	bweight_low	-0.000225	0.000029	-7.744491	0.000000
1-Year (Calipers 85g) U/C	VLBM:bweight_low	0.000090	0.000043	2.081209	0.037416
28-Day (Calipers 85g) U/C	(Intercept)	0.045244	0.001037	43.610193	0.000000
28-Day (Calipers 85g) U/C	VLBM	-0.008781	0.001811	-4.847578	0.000001
28-Day (Calipers 85g) U/C	bweight_low	-0.000200	0.000024	-8.177934	0.000000
28-Day (Calipers 85g) U/C	VLBM:bweight_low	0.000087	0.000036	2.388294	0.016928

7) Now add covariates to the model in (6). Include mother's age, indicators for mother's education and race, indicators for year of birth, indicators for gestational age and prenatal care visits. Use the dummies provided in the data for gestational age and prenatal care visits. Compare your estimates to those obtained in (6) and explain the difference if any.

Answer: For both 1-year mortality and 28-day mortality rates, it can be seen that both the size of the estimates and the standard errors for VLBM decreased compared to the estimates obtain in (6). These difference come from holding the covariates constant some of which were not smooth around the threshold. That is, these estimates partial out the differences that were correlated with the differences in the background covariates.

Table 4: Controlled vs Uncontrolled Estimates for 1-Year Mortality Rates

var	term	estimate	std.error	statistic	p.value
1-Year (Calipers 85g)	(Intercept)	4.9217603	0.1783642	27.5938751	0.0000000
1-Year (Calipers 85g)	VLBM	-0.0077959	0.0021442	-3.6357930	0.0002772
1-Year (Calipers 85g)	bweight_low	-0.0002027	0.0000289	-7.0100642	0.0000000
1-Year (Calipers 85g)	mom_age	-0.0002177	0.0000894	-2.4339513	0.0149359
1-Year (Calipers 85g)	mom_ed1	-0.0101914	0.0022067	-4.6184323	0.0000039
1-Year (Calipers 85g)	mom_ed2	-0.0126558	0.0021211	-5.9666127	0.0000000
1-Year (Calipers 85g)	mom_ed3	-0.0188016	0.0023231	-8.0934676	0.0000000
1-Year (Calipers 85g)	mom_ed4	-0.0196449	0.0024410	-8.0478297	0.0000000
1-Year (Calipers 85g)	white	0.0007673	0.0026445	0.2901433	0.7717069
1-Year (Calipers 85g)	black	-0.0161376	0.0027568	-5.8537666	0.0000000
1-Year (Calipers 85g)	yob	-0.0024204	0.0000897	-26.9736006	0.0000000
1-Year (Calipers 85g)	gest_wks1	-0.0092196	0.0018970	-4.8601494	0.0000012
1-Year (Calipers 85g)	gest_wks2	0.0187102	0.0026393	7.0891614	0.0000000
1-Year (Calipers 85g)	gest_wks3	0.0003349	0.0071588	0.0467780	0.9626902
1-Year (Calipers 85g)	nprenatal_1	-0.0014603	0.0019515	-0.7482906	0.4542858
1-Year (Calipers 85g)	nprenatal_2	-0.0109506	0.0020294	-5.3960523	0.0000001
1-Year (Calipers 85g)	nprenatal_3	-0.0154510	0.0023864	-6.4746513	0.0000000
1-Year (Calipers 85g)	VLBM:bweight_low	0.0000787	0.0000429	1.8333910	0.0667459
1-Year (Calipers 85g) U/C	(Intercept)	0.0631559	0.0012332	51.2115717	0.0000000
1-Year (Calipers 85g) U/C	VLBM	-0.0095102	0.0021532	-4.4167673	0.0000100
1-Year (Calipers 85g) U/C	bweight_low	-0.0002248	0.0000290	-7.7444915	0.0000000
1-Year (Calipers 85g) U/C	VLBM:bweight_low	0.0000897	0.0000431	2.0812093	0.0374160

Table 5: Controlled vs Uncontrolled Estimates for 28-Day Mortality Rates

var	term	estimate	std.error	statistic	p.value
28-Day (Calipers 85g)	(Intercept)	3.6739439	0.1501331	24.4712437	0.0000000
28-Day (Calipers 85g)	VLBM	-0.0075559	0.0018048	-4.1864478	0.0000283
28-Day (Calipers 85g)	bweight_low	-0.0001846	0.0000243	-7.5873513	0.0000000
28-Day (Calipers 85g)	mom_age	-0.0000484	0.0000753	-0.6432018	0.5200939
28-Day (Calipers 85g)	mom_ed1	-0.0126912	0.0018574	-6.8327816	0.0000000
28-Day (Calipers 85g)	mom_ed2	-0.0091505	0.0017854	-5.1252472	0.0000003
28-Day (Calipers 85g)	mom_ed3	-0.0131516	0.0019554	-6.7259155	0.0000000
28-Day (Calipers 85g)	mom_ed4	-0.0122494	0.0020547	-5.9617570	0.0000000
28-Day (Calipers 85g)	white	0.0035871	0.0022259	1.6115177	0.1070685
28-Day (Calipers 85g)	black	-0.0166273	0.0023204	-7.1655650	0.0000000
28-Day (Calipers 85g)	yob	-0.0018089	0.0000755	-23.9496148	0.0000000
28-Day (Calipers 85g)	gest_wks1	-0.0071394	0.0015967	-4.4712494	0.0000078
28-Day (Calipers 85g)	gest_wks2	0.0133661	0.0022215	6.0166463	0.0000000
28-Day (Calipers 85g)	gest_wks3	-0.0012857	0.0060257	-0.2133683	0.8310399
28-Day (Calipers 85g)	nprenatal_1	-0.0011892	0.0016426	-0.7239916	0.4690718
28-Day (Calipers 85g)	nprenatal_2	-0.0081981	0.0017082	-4.7993713	0.0000016

var	term	estimate	std.error	statistic	p.value
28-Day (Calipers 85g)	nprenatal_3	-0.0134264	0.0020087	-6.6841938	0.0000000
28-Day (Calipers 85g)	VLBM:bweight_low	0.0000797	0.0000361	2.2070812	0.0273095
28-Day (Calipers 85g) U/C	(Intercept)	0.0452440	0.0010375	43.6101930	0.0000000
28-Day (Calipers 85g) U/C	VLBM	-0.0087808	0.0018114	-4.8475784	0.0000013
28-Day (Calipers 85g) U/C	bweight_low	-0.0001997	0.0000244	-8.1779338	0.0000000
28-Day (Calipers 85g) U/C	VLBM:bweight_low	0.0000866	0.0000363	2.3882937	0.0169277

8) Use the model in (7) to assess the sensitivity of the estimates to the use of different calipers. Use calipers of 30 and 120 grams (above and below the 1500 threshold). Are the estimates any different to those obtained in (7)? What is the tradeoff that we face when decreasing the caliper?

Answer: As the area of calipers increased, both the size of the estimates and the standard errors for VLBM in both 1-year and 28-days mortality rates have decreased compared to those obtained in (7). These show the bias-variance tradeoff when choosing the levels of caliper. Although lower caliper reduces the bias, it increases the variance due to low number of observations.

Table 6: The Estimates for 1-Year Mortality Rates

term	estimate	std.error	statistic	p.value	var
VLBM	-0.0148483	0.0051705	-2.871726	0.0040835	1-Year (Calipers 30g)
VLBM	-0.0077959	0.0021442	-3.635793	0.0002772	1-Year (Calipers 85g)
VLBM	-0.0064839	0.0017610	-3.681991	0.0002315	1-Year (Calipers 120g)

Table 7: The Estimates for 28-Day Mortality Rates

term	estimate	std.error	statistic	p.value	var
VLBM	-0.0149380	0.0043719	-3.416791	0.0006340	28-Day (Calipers-30g)
VLBM	-0.0075559	0.0018048	-4.186448	0.0000283	28-Day (Calipers 85g)
VLBM	-0.0056937	0.0014831	-3.839091	0.0001235	28-Day (Calipers 120g)

9) Synthetize your findings and discuss what kind of supplementary information would you need to make a cost-benefit analysis of treatment received by newborns close to the very low birth weight threshold.

Answer:

To synthesize my findings, it can be seen that the mortality rates decrease with some magnitude just before the threshold levels. This is plausible because these babies might get additional treatment for their very low weight statuts. It was also found that the mortality rates increased as the birth weight decreased far below the threshold. These findings, however, must be reconciled with the fact that the estimates were changing depending on the level of calipers and that there could be other unobservable characteristics excluded from the regression model that may influence the estimates. In general, however, we can see that children born right underneath the threshold have lower mortality rates.

To evaluate the cost-benefit analysis, I will need supplementary information about the hosptial costs for those who fall short of the low-birth threshold. Additionally, it would be nice to know about the social benefits measured in dollars by saving the baby with the treatment. This way, we can measure whether the cost of the treatment outweighs the benefits and vice versa.