1a.

(a) Estimate model using only data from Minnesota

$$Wage = \beta_0 + \beta_1 after + \beta_2 women + \beta_3 after*women + u$$

$$\hat{\beta}_0 = Wage(after=0, women=0) = 10$$

$$\hat{\beta}_1 = Wage(after=1, women=0) - \hat{\beta}_0$$

$$= 12 - 10 = 2$$

$$\hat{\beta}_2 = Wage(after=0, women=1) - \hat{\beta}_0$$

$$= 8 - 10 = -2$$

$$\hat{\beta}_3 = Wage(after=1, women=1) - \hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_2$$

$$= 9 - 10 - 2 + 2 = -1$$

1b

(b)

Model

$$Wage = \beta_0 + \beta_1\, after + \beta_2\, women + \beta_3\, Minnesota + \beta_4\, after \times women + \beta_5\, after \times Minnesota + \beta_6\, Minnesota \times women + \beta_7\, after \times women \times Minnesota + u.$$

$\rightarrow \hat{\beta}_0 =$ after=0 women=0 Minnesota=0

$= \$12$

$\rightarrow \hat{\beta}_1 = Wage(after=1, women=0\ Minnesota=0) - \hat{\beta}_0$

$= 14 - 12$

$= 2$

$\rightarrow \hat{\beta}_2 = Wage(after=0, women=1, minnesota=0) - \hat{\beta}_0$

$= 9 - 12 = -3$

$\rightarrow \hat{\beta}_3 = Wage(after=0, women=0, minnesota=1) - \hat{\beta}_0$

$= 10 - \hat{\beta}_0 = -2$

$\rightarrow \hat{\beta}_4 = wage(after=1, women=1, Minnesota=0) - \hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_2$

$= 10 - 12 + 3 - 2 = -1$

$\rightarrow \hat{\beta}_5 = Wage(after=1, women=0, Minnesota=1) - \hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_3$

$= 12 - 12 - 2 + 2 = 0$

Continued..

→ $\hat{\beta}_6 = \text{wage}(\text{after}=0, \text{women}=1, \text{Minnesota}=1) - \beta_0 - \beta_2 - \beta_3$

$= 8 - 12 + 3 + 2 = 1$

→ $\hat{\beta}_7 = \text{wage}(\text{after}=1, \text{women}=1, \text{Minnesota}=1) - \hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3 - \hat{\beta}_4 - \hat{\beta}_5 - \hat{\beta}_6$

$= 9 - 12 - 2 + 3 + 2 + 1 - 0 - 1$

$= 0$

1c

(c) The palallel trends assumptions is that the trend of avg wage would be same across Men and Women when the law is not present. The causal impact of the law is found to be a decrease in women's wage by $1, from the $\hat{DiD}$ calculated $\hat{\beta}_3 = -1$

We got a more accurate estimate after including data from a different control group (wisconsin) that leads to $\hat{DiD} = 0$ meaning the new law has no net impact on women's wages. from a and b, the reasonable conclusion is that new law has no causal impact on women's wage.

Question2, next page

2a, 2b

2 (a) Equation

$$Y_{ia} = \beta_0 + \beta_1 D_{a \geq a^*} + f(a) + \varepsilon_{ia}$$

↳ β_2 math-ranking_2001

i = individual

Y_{ia} = outcome variable – math score in national exam 2006.

a = running variable

$D_{a \geq a}$ is indicator for an individual whether in top 15 percentile in math exam, as 85 percentile

$f(a)$, a linear or quadratic function, with slopes above and below a^* (math score percentile) – β_2 Math-ranking_2001

β_0 – Mean score percentile when no cash prize

β_1 – Effect of cashprize in 2006 due to 2001

ε_{ia} – error on individual.

(b) By using the Regression Continuity design, we can find the effect of policy (cash prize) in 2001 on the math exam score in 2006 using coefficients. However we cannot say anything using coefficients about effect of other factors like school books (or) education technology (or) any other factor that may have helped a student to score in the top 15%ile in 2006.

2c,2d next page

2c, 2d

(c) The researchers can use the math scores in 2000 to determine or estimate bunching around the threshold (a^* 85). The presence of bunching can hold the randomization assumption invalid and also the causal effect interpretation found using the Regression Discontinuity design RDD. If there is no presence of bunching around the threshold, then the randomization is valid and consisted across the two periods which says the conclusion, cash prize in 2001 has a causal impact on the score in math exam in 2006.

(d) in-school is +ve correlated with math_score_2001. This correlation would not create bias in the design as this is just a case of exogenous selection and the data or results around the threshold are (a^* 85) considered as random under the assumptions of Regression continuity design. As the high achievers of more than (a^* 85%) are more likely to stay in school there wouldnot be a bias in the estimating the design for scores in 2006.