# Part 1

## 1.1

Table 1: Mean Characteristics

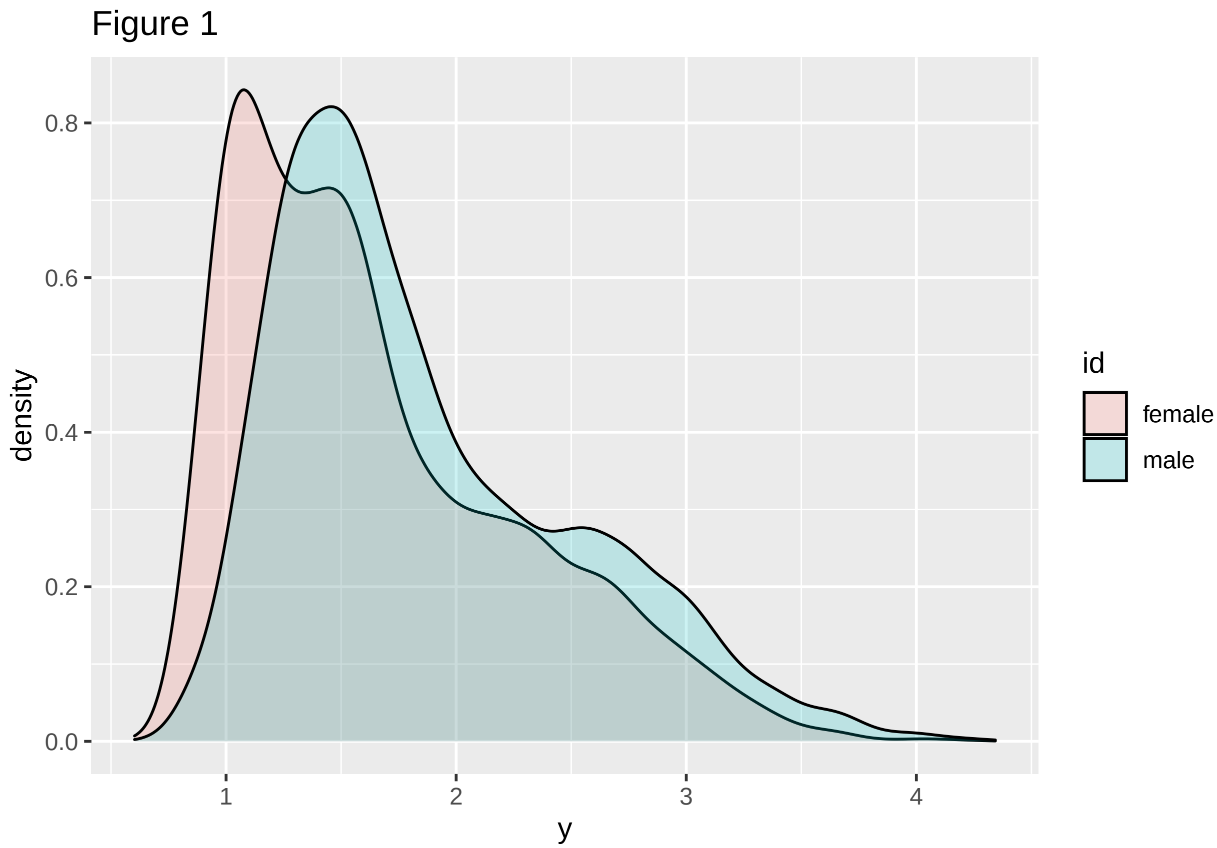
| | All Mean | Male Mean | Female Mean | T-Stat Diff. |
|---|---|---|---|---|
| educ | 10.484 | 10.272 | 10.835 | -9.903 |
| age | 33.559 | 33.574 | 33.534 | 0.436 |
| y | 1.788 | 1.866 | 1.658 | 20.816 |
| owage2 | 1.692 | 1.725 | 1.638 | 11.676 |
| exp | 17.075 | 17.302 | 16.700 | 5.845 |
| female | 0.377 | 0 | 1 | |

Table 2: Education Distribution

| | All Frac. | Male Frac. | Female Frac. |
|---|---|---|---|
| educ 6 | 0.274 | 0.289 | 0.249 |
| educ 9 | 0.227 | 0.247 | 0.195 |
| educ 12 | 0.296 | 0.276 | 0.329 |
| educ 16 | 0.203 | 0.188 | 0.228 |
| Sum | 1 | 1 | 1 |

Table 3: Wage (y) Distribution

| | All Frac. | Male Frac. | Female Frac. |
|---|---|---|---|
| Q1 All | 0.250 | 0.189 | 0.351 |
| Q2 All | 0.250 | 0.260 | 0.233 |
| Q3 All | 0.250 | 0.274 | 0.210 |
| Q4 All | 0.250 | 0.277 | 0.206 |
| Sum | 1.000 | 1.000 | 1 |




In Table 1, men and women in the sample differ significantly on all characteristics aside from age. Men tend to have lower levels of education, which is confirmed in Table 2. Nonetheless, men tend to earn higher wages, and are better represented in higher wage quartiles in Table 3. Figure 1 also supports this result, as the distribution of males wages is shifted rightwards relative to the female distribution. Interestingly male co-workers tend to earn more than female co-workers, so profession may explain part of the wage gap. Males also have higher potential experience by virtue of leaving school earlier on average.

### 1.2

Table 4: Pooled and Individual Regressions

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | y | | | |
| | (1) | (2) | (3) | (4) |
| educ | | $0.149^{***}$ | $0.149^{***}$ | $0.149^{***}$ |
| | | (0.001) | (0.001) | (0.002) |
| exp | | $0.033^{***}$ | $0.034^{**}$ | $0.043^{***}$ |
| | | (0.011) | (0.015) | (0.016) |
| exp2 | | 0.001 | 0.001 | −0.0002 |
| | | (0.001) | (0.001) | (0.001) |
| exp3 | | $-0.00004^{***}$ | $-0.00005^{***}$ | −0.00002 |
| | | (0.00001) | (0.00002) | (0.00002) |
| female | $-0.208^{***}$ | $-0.273^{***}$ | | |
| | (0.010) | (0.007) | | |
| Constant | $1.866^{***}$ | $-0.264^{***}$ | $-0.315^{***}$ | $-0.529^{***}$ |
| | (0.006) | (0.060) | (0.082) | (0.085) |
| Observations | 16,969 | 16,969 | 10,575 | 6,394 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

(1) and (2) use pooled data. (3) and (4) use male and female data, respectively.

## Oaxaca decomposition pooled model:

Explained Difference: 0.0641 (-30.76%)

Unexplained Difference: -0.2725 (130.76%)

## Oaxaca decompotion male and female models:

Explained Difference $(\overline{x}^f - \overline{x}^m)'\hat{\beta}^m$: 0.0619 (-29.68%)

Returns Difference $(\overline{x}^f)'(\hat{\beta}^f - \hat{\beta}^m)$ - $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.0565 (27.09%)

Explained Difference $(\overline{x}^f - \overline{x}^m)'\hat{\beta}^f$: 0.0670 (-32.14%)

Returns Difference $(\overline{x}^m)'(\hat{\beta}^f - \hat{\beta}^m)$ - $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.0616 (29.54%)

Constant Difference $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.2138 (102.60%)

The difference in male and female wages: **-0.2084** as observed in the first regression of Table 4. The decomposition using the pooled regression has both males and females receiving the same returns for observed covariates, while the one using the separate male and female regressions allows for different returns.

The difference in mean wages is not explained by the difference in male and female covariates. Since the explained portion has the opposite sign as the gap, the wage gap is understated. Both females and males have the same returns to education, according to Table 4. Females should actually be earning higher wages based on their higher mean education levels.

Yet, females and males have different returns to potential experience which accounts roughly 27.09% to 29.54% of the wage gap. The remaining portion of the gap is unexplained. Perhaps adding additional controls would allow us to better explain the wage gap.

## 1.3

Table 5: Pooled and Individual Regressions

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | y | | | |
| | (1) | (2) | (3) | (4) |
| educ | | 0.098*** | 0.099*** | 0.094*** |
| | | (0.001) | (0.001) | (0.002) |
| exp | | 0.025*** | 0.022* | 0.039*** |
| | | (0.009) | (0.012) | (0.013) |
| exp2 | | 0.001* | 0.001* | −0.0003 |
| | | (0.001) | (0.001) | (0.001) |
| exp3 | | −0.00003*** | −0.00004*** | −0.00001 |
| | | (0.00001) | (0.00001) | (0.00001) |
| female | −0.121*** | −0.192*** | | |
| | (0.007) | (0.006) | | |
| owage2 | 1.006*** | 0.638*** | 0.662*** | 0.604*** |
| | (0.007) | (0.007) | (0.009) | (0.011) |
| Constant | 0.131*** | −0.716*** | −0.809*** | −0.816*** |
| | (0.013) | (0.049) | (0.068) | (0.070) |
| Observations | 16,969 | 16,969 | 10,575 | 6,394 |

*Note:* *p<0.1; **p<0.05; ***p<0.01
(1) and (2) use pooled data. (3) and (4) use male and female data, respectively.

### Oaxaca decomposition pooled model (2):

Explained Difference: -0.0168 (8.05%)

Unexplained Difference: -0.1916 (91.95%)

### Oaxaca decompotion male and female models:

Explained Difference $(\overline{x}^f - \overline{x}^m)'\hat{\beta}^m$: -0.0120 (9.59%)

Returns Difference $(\overline{x}^f)'(\hat{\beta}^f - \hat{\beta}^m)$ - $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.1820 (87.32%)

Explained Difference $(\overline{x}^f - \overline{x}^m)'\hat{\beta}^f$: -0.0128 (6.13%)

Returns Difference $(\overline{x}^m)'(\hat{\beta}^f - \hat{\beta}^m)$ - $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.1892 (90.77%)

Constant Difference $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.0064 (3.09%)

Differences in covariates now account for 8.05% of the wage gap in the pooled model with equal male and female returns. When allowing for different male and female returns, 6.13% to 9.59% of the wage gap is accounted for.

87.32% to 90.77% of the wage gap can be attributed to higher male returns relative to female returns. Table 5 shows that men get have a slightly higher coefficient for education and a significantly higher coefficient for co-worker wages.

Model 1: High paying occupations such as engineering jobs and business rely heavily on employee referrals. The majority of these occupations are held by men, who tend to have many more male friends than female friends to refer. It's uncertain if men search harder for higher paying jobs than women, or if men are luckier. Though, men tend to be more aggressive in negotiating for wages.

Model 2: Cognitive skills are unlikely to differ substantially among men and women in the sample. Ambition is difficult to quantify, but career ambitions are likely to be similar for men and women. (Though I don't know how the data was collected.)

These models help reconcile the elevated returns to co-worker wage for men. In the decomposition, the portion of the wage gap attributable to return differences is substantial. Hence the more plausible model, which is Model 1, offers a mechanism for understanding the wage gap.

## 1.4

## Figure 2: Mean Wages vs. Event Time

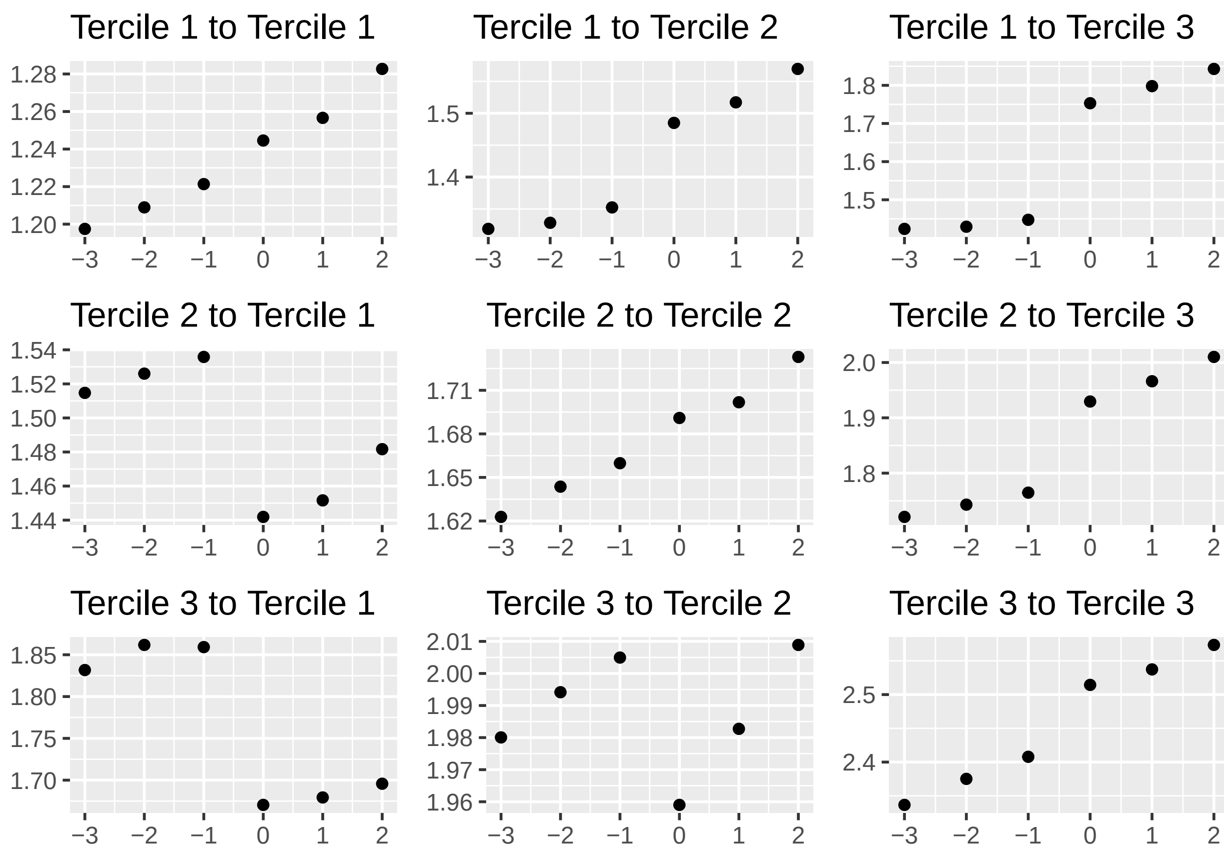


Table 6: Co-worker Tercile Transitions

| | All Frac. | Male Frac. | Female Frac. |
|---|---|---|---|
| T1-T1 | 0.232 | 0.182 | 0.315 |
| T1-T2 | 0.077 | 0.081 | 0.070 |
| T1-T3 | 0.024 | 0.023 | 0.026 |
| T2-T1 | 0.081 | 0.086 | 0.072 |
| T2-T2 | 0.189 | 0.194 | 0.181 |
| T2-T3 | 0.063 | 0.068 | 0.056 |
| T3-T1 | 0.020 | 0.024 | 0.015 |
| T3-T2 | 0.067 | 0.084 | 0.039 |
| T3-T3 | 0.246 | 0.258 | 0.226 |
| Sum | 1 | 1 | 1 |

The sharpest changes in mean wage coincide with changes in co-worker tercile wage. Moving up a tercile results in a sharp increase in mean wage, while moving down a tercile results in a sharp decline in mean wage. There are slightly different trends prior to the event for tercile changes vs. staying in the same tercile. Tercile changes tend to coincide with a flattening of wages before hand, with the exception of the Tercile 3 to Tercile 2 transition. Staying in the same tercile coincides with steadily rising wages. These different patterns before job change are somewhat concerning. Perhaps workers who have consistent wage growth differ systematically from those with flattening wages.

Table 6 shows the fractions of males and females making transitions to different terciles. A substantial fraction of females stay steadily in Tercile 1, while about a quarter of men stay steadily in Tercile 3. Overall, there is a greater fraction of men making transitions between quartiles. The only exception is where a greater fraction of females make the transition is from Tercile 1 to Tercile 3.

Model 2 unlikely to hold. If people self-select into jobs based on skills and ambition there should be few jobs transitions that result in changing terciles. This is because individuals and their peers are unlikely to change substantially in cognitive skills or ambition in a few years, so their wages along with co-worker wages should be steady. While about 2/3 of jobs changes stay within co-worker tercile, 1/3 lead to changing terciles. The presence of upwards mobility contradicts the model.

Model 1 is weakly supported. Suppose through a connection a man gets a job that pays a high wage and has co-workers who are also paid well. Despite the connection, the man performs poorly. He gets fired and starts at a lower paying job. Table 6 shows larger fractions of men transitioning to jobs with lower co-worker wages. We would expect to see steep drops in wages due to these adverse transitions. These are supported in Figure 2.

Table 7: Pooled and Individual Regressions

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Dy | | | |
| | (1) | (2) | (3) | (4) |
| expl1 | | $-0.009^{***}$ | $-0.011^{***}$ | $-0.008^{***}$ |
| | | (0.002) | (0.002) | (0.002) |
| expl1_2 | | $0.0002^{***}$ | $0.0002^{***}$ | 0.0001 |
| | | (0.00005) | (0.0001) | (0.0001) |
| female | $-0.019^{***}$ | $-0.022^{***}$ | | |
| | (0.004) | (0.004) | | |
| Dwage | $0.294^{***}$ | $0.290^{***}$ | $0.332^{***}$ | $0.218^{***}$ |
| | (0.006) | (0.006) | (0.008) | (0.010) |
| Constant | $0.045^{***}$ | $0.153^{***}$ | $0.163^{***}$ | $0.123^{***}$ |
| | (0.003) | (0.013) | (0.018) | (0.018) |
| Observations | 16,969 | 16,969 | 10,575 | 6,394 |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

Dy: y-yl1, (1) and (2) use pooled data. (3) and (4) use male and female data, respectively.

Use the differenced coefficients on Dwage in Table 7 as the true values of the coefficents for owage2 in Table 5. Transform the dependent variable for each regression in Table 5 to: $y - \beta_{Dwage} * owage2$. Use the remaining covariates of the OLS regressions (educ, cubic of exp) to predict this transformed dependent variable. This will lead to improved estimates for these covariates. Then, using the estimated coefficients and the corresponding one for Dwage, perform the decomposition.

Table 8: Pooled and Individual Modified Regressions

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | y_pool_adj1 | y_pool_adj2 | y_male_adj | y_fem_adj |
| | (1) | (2) | (3) | (4) |
| educ | | $0.126^{***}$ | $0.124^{***}$ | $0.129^{***}$ |
| | | (0.001) | (0.001) | (0.001) |
| exp | | $0.029^{***}$ | $0.028^{**}$ | $0.042^{***}$ |
| | | (0.010) | (0.013) | (0.014) |
| exp2 | | 0.001 | 0.001 | $-0.0002$ |
| | | (0.001) | (0.001) | (0.001) |
| exp3 | | $-0.00004^{***}$ | $-0.00004^{***}$ | $-0.00001$ |
| | | (0.00001) | (0.00001) | (0.00002) |
| female | $-0.183^{***}$ | $-0.236^{***}$ | | |
| | (0.009) | (0.006) | | |
| Constant | $1.360^{***}$ | $-0.470^{***}$ | $-0.563^{***}$ | $-0.632^{***}$ |
| | (0.005) | (0.053) | (0.071) | (0.077) |
| Observations | 16,969 | 16,969 | 10,575 | 6,394 |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

(1) and (2) use pooled data. (3) and (4) use male and female data, respectively.

### Oaxaca decomposition pooled model (2):

Explained Difference: 0.0273 (-13.11%)

Unexplained Difference: -0.2357 (113.11%)

### Oaxaca decompotion male and female models:

Explained Difference $(\overline{x}^f - \overline{x}^m)'\hat{\beta}^m$: 0.0208 (-9.99%)

Returns Difference $(\overline{x}^f)'(\hat{\beta}^f - \hat{\beta}^m)$ - $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.1598 (76.67%)

Explained Difference $(\overline{x}^f - \overline{x}^m)'\hat{\beta}^f$: 0.0382 (-18.33%)

Returns Difference $(\overline{x}^m)'(\hat{\beta}^f - \hat{\beta}^m)$ - $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.1772 (85.00%)

Constant Difference $(\hat{\beta_0}^f - \hat{\beta_0}^m)$: -0.0694 (33.32%)

Once again the explained portion has the opposite sign as the wage gap. After accounting for unobserved skills of people who work at high-coworker wage jobs, we can conclude that the male and female wage gap is understated.

# Part 2

## 2.1

### a)

### (i)

From Lecture 15:

$$E[w_i|AT(0)\ or\ C(0)] = \frac{E[w_i|AT(0)] * Pr(AT(0)) + E[w_i|C(0)] * Pr(C(0))}{Pr(AT(0)\ or\ C(0))}$$

Hence:

$$E[w_i|AT(0)\ or\ C(0)] * Pr(AT(0)\ or\ C(0)) = E[w_i|AT(0)] * Pr(AT(0)) + E[w_i|C(0)] * Pr(C(0))$$

Then:

$$E[w_i|C(0)] * Pr(C(0)) = E[w_i|AT(0)\ or\ C(0)] * Pr(AT(0)\ or\ C(0)) - E[w_i|AT(0)] * Pr(AT)$$

Finally:

$$E[w_i|C(0)] = \frac{E[w_i|AT(0)\ or\ C(0)] * Pr(AT(0)\ or\ C(0)) - E[w_i|AT(0)] * Pr(AT(0))}{Pr(C(0))}$$

### (ii)

Using LIE:

$$E[w_iD_i|x_i \to 0, z_i = 1] = E[D_i * E[w_i|D_i, x_i \to 0, z_i = 1]$$

$$E[D_i * E[w_i|D_i, x_i \to 0, z_i = 1] = E[w_i|Di = 1, x_i \to 0, z_i = 1] * Pr(Di = 1|x_i \to 0, z_i = 1)$$

Hence:

$$E[w_iD_i|x_i \to 0, z_i = 1] = E[w_i|AT(0)\ or\ C(0)] * Pr(AT(0)\ or\ C(0))$$

### (iii)

Using LIE:

$$E[w_iD_i|x_i \to 0, z_i = 0] = E[D_i * E[w_i|D_i, x_i \to 0, z_i = 0]$$

$$E[D_i * E[w_i|D_i, x_i \to 0, z_i = 0] = E[w_i|Di = 1, x_i \to 0, z_i = 0] * Pr(Di = 1|x_i \to 0, z_i = 0)$$

Hence:

$$E[w_iD_i|x_i \to 0, z_i = 0] = E[w_i|AT(0)] * Pr(AT(0))$$

**(iv)**

Given:

$$\hat{\beta}_1 = \hat{\delta}_1 / \hat{\pi}_1$$

$$\hat{\delta}_1 = E[w_i D_i | x_i \to 0, z_i = 1] - E[w_i D_i | x_i \to 0, z_i = 0]$$
$$\hat{\delta}_1 = E[w_i | AT(0) \ or \ C(0)] * Pr(AT(0) \ or \ C(0)) - E[w_i | AT(0)] * Pr(AT(0))$$
$$\hat{\pi}_1 = E[D_i | x_i \to 0, z_i = 1] - E[D_i | x_i \to 0, z_i = 0]$$
$$\hat{\pi}_1 = Pr(C(0))$$

Hence:

$$\hat{\beta}_1 = E[w_i | C(0)] = \frac{E[w_i | AT(0) \ or \ C(0)] * Pr(AT(0) \ or \ C(0)) - E[w_i | AT(0)] * Pr(AT(0))}{Pr(C(0))}$$

**2.2**

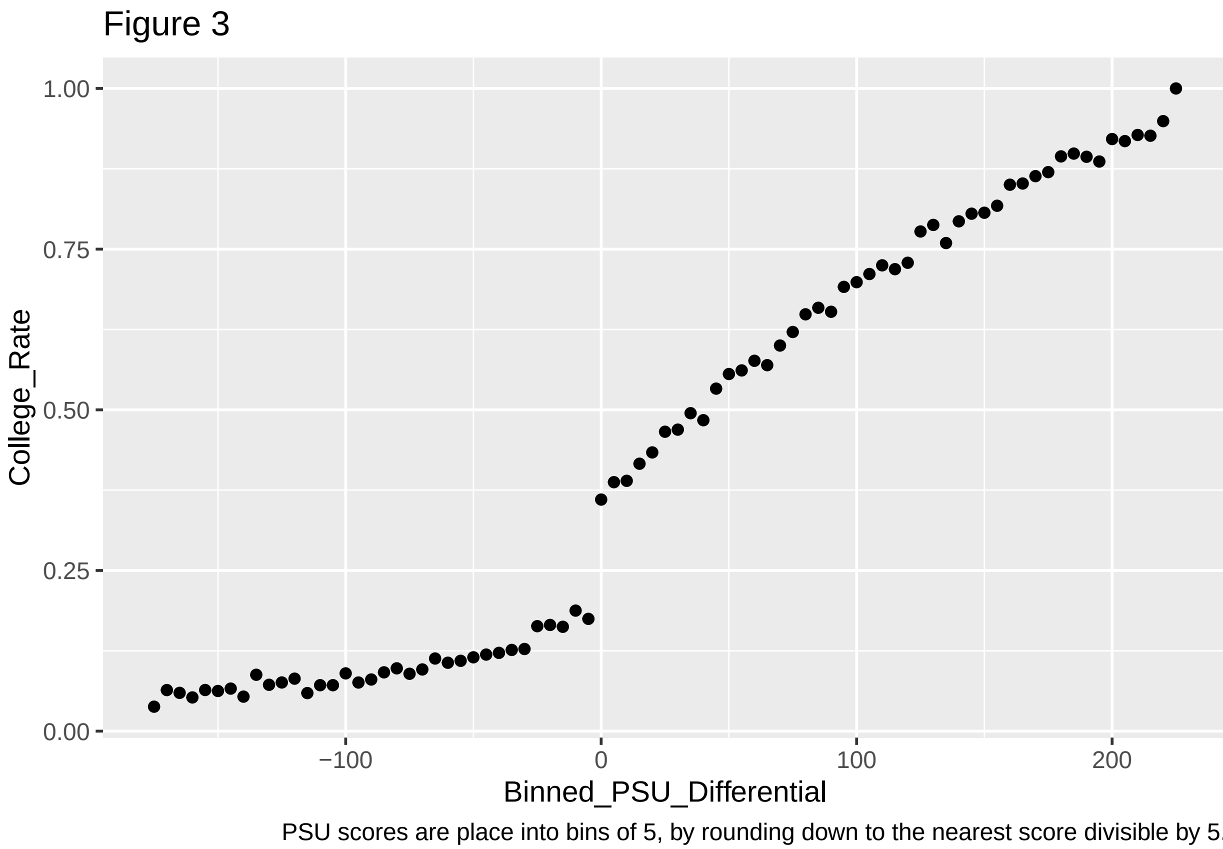
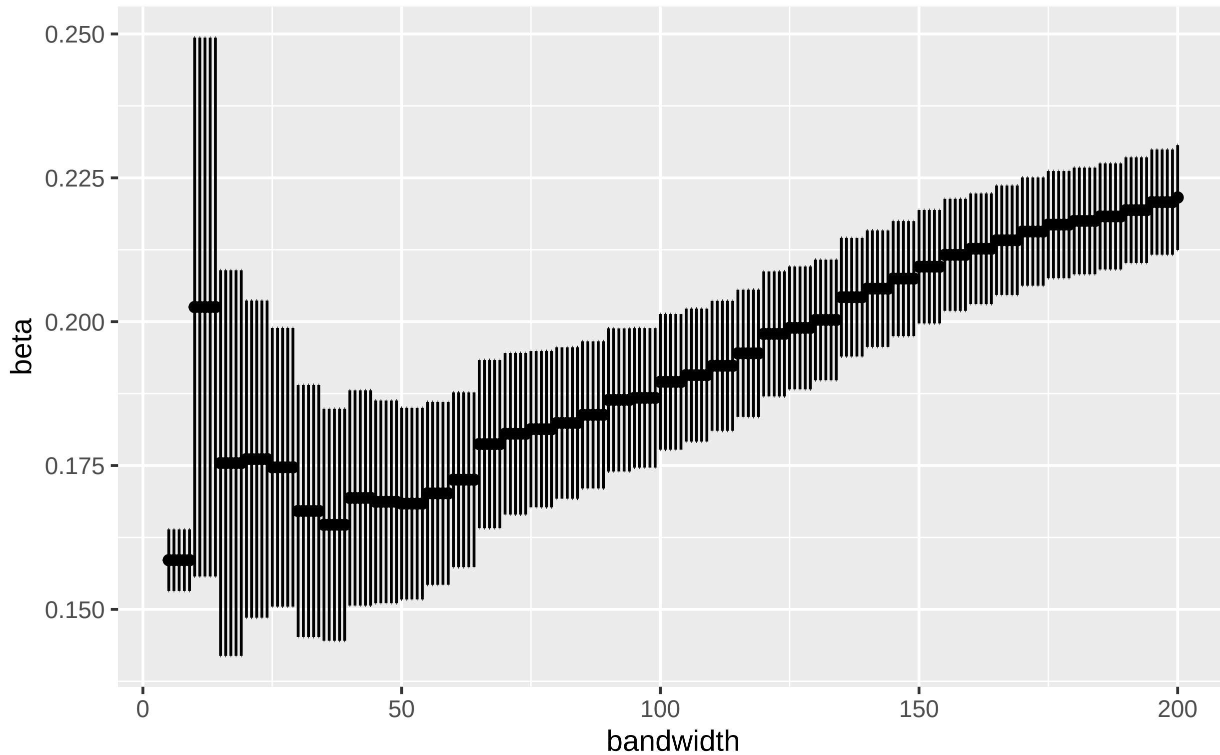



Table 9: Binned Regressions for Bandwidths 25, 50, 75, 100

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | entercollege | | | |
| | (1) | (2) | (3) | (4) |
| over475 | 0.175*** | 0.168*** | 0.181*** | 0.190*** |
| | (0.012) | (0.008) | (0.007) | (0.006) |
| x_binned | 0.001 | 0.002*** | 0.001*** | 0.001*** |
| | (0.001) | (0.0002) | (0.0001) | (0.0001) |
| x_binnedXover475 | 0.003*** | 0.002*** | 0.002*** | 0.002*** |
| | (0.001) | (0.0003) | (0.0002) | (0.0001) |
| Constant | 0.184*** | 0.193*** | 0.185*** | 0.179*** |
| | (0.010) | (0.007) | (0.005) | (0.005) |
| Observations | 24,815 | 46,068 | 65,053 | 81,133 |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

Figure 4




Integer bandwidths from 5 to 200.

Figure 3 shows that a linear fit may be reasonable for a bandwidth of up to 50. Beyond that point the upper part of the plot begins to curve, so a linear model would be more biased at higher bandwidth. Yet, higher bandwidths include more data, so they coincide with estimates that have lower variance. Figure 4 shows that the smaller the bandwidth, the larger than standard errors tend to be. So we don't want to choose a bandwidth that is too small, as it may not lead to a very precise and reproducible estimate. The estimated discontinuity coefficient holds somewhat steady for the bandwidths shown in Table 9. Based on these observation a bandwidth choice between 25 to 50 may strike a reasonable balance between precision and bias. Choose 50 for the following table.

Table 10: Characteristic Shares

| | Full Sample | Bandwith 50 Sample | Compliers | Ratio |
|---|---|---|---|---|
| Share Income Q1 | 0.473 | 0.517 | 0.581 | 1.230 |
| Share Income Q2 | 0.216 | 0.216 | 0.247 | 1.143 |
| Share Income Q3 | 0.159 | 0.146 | 0.123 | 0.771 |
| Share Income Q4 | 0.152 | 0.121 | 0.049 | 0.320 |
| Share Female | 0.569 | 0.588 | 0.642 | 1.129 |
| Share 60<=GPA<=70 | 0.319 | 0.230 | 0.261 | 0.820 |
| Share 50<=GPA<60 | 0.627 | 0.714 | 0.684 | 1.092 |
| Share GPA<50 | 0.055 | 0.056 | 0.054 | 0.998 |
| Share Mom Educ>HS | 0.148 | 0.116 | 0.091 | 0.615 |
| Share Dad Educ>HS | 0.156 | 0.120 | 0.106 | 0.679 |

Table 10 strongly supports the claim that the loan program extends college access to more economically disadvantaged students. Program compliers tend to be from lower family income quintiles, with the greatest fraction of them in the lowest quintile. Compliers are more likely to be female. In addition, fewer compliers are in the highest GPA category, though this is attributable in part to the restricted test score bandwidth. Parents of compliers tend to have less education than those in the larger sample.

Overall, the loan program appears to fulfilling its role. It also promotes greater female education and first generation college students to some extent.