Review Sheet 3

1. (a) Conditional on X, Y_1 and Y_0 are orthogonal to D, which means that we can't use Y_1 or Y_0 to predict college attendance for groups homogenous in $X = x$. Independently of X, we see that $\mathbb{E}[Y_1|D] = \mathbb{E}[\mathbb{E}[Y_1|D,X]|D] = \alpha_1 + \gamma_1\mathbb{E}[X|D]$, and similarly $\mathbb{E}[Y_0|D] = \alpha_0 + \gamma_0\mathbb{E}[X|D]$. So, the observed values of Y_1 and Y_2 would indeed be different for different values of D, which means we could use them to predict whether an individual completed college.

   We can't use an individual's observed wage Y to predict whether they went to college without controlling for X, the characteristic measured at the completion of high school, which we can think of as ability. Intuitively, a higher observed wage could mean that an individual attended college, but it also could mean that an individual has a high ability and didn't attend college.

   (b) β_0 is the expected difference in earnings for a college graduate compared to if they did not complete college. Using the law of iterated expectations

   $$\mathbb{E}[Y_1|D=1] = \mathbb{E}[\mathbb{E}[Y_1|D,X]|D=1] = \mathbb{E}[\mathbb{E}[Y_1|X]|D=1] = \mathbb{E}[\alpha_1 + \gamma_1 X|D=1] = \alpha_1 + \gamma_1\mathbb{E}[X|D=1].$$

   Similarly, $\mathbb{E}[Y_0|D=1] = \alpha_0 + \gamma_0\mathbb{E}[X|D=1]$. So $\beta_0 = (\alpha_1 - \alpha_0) + (\gamma_1 - \gamma_0)\mathbb{E}[X|D=1]$.

   (c)

   $$\mathbb{E}[Y|X,D] = (1-D)\mathbb{E}[Y_0|X,D] + D\mathbb{E}[Y_1|X,D] = (1-D)(\alpha_0 + \gamma_0 X) + D(\alpha_1 + \gamma_1 X)$$
   $$\mathbb{E}[Y|X,D] = \alpha_0 + \gamma_0 X + (\alpha_1 - \alpha_0)D + (\gamma_1 - \gamma_0)DX$$

   (d) From (2), $(\alpha_1 - \alpha_0) = 0.1898$ and $(\gamma_1 - \gamma_0) = 0.0023$. From (3), $\mathbb{E}[X|D=1] = 52.27 + 28.35 = 80.62$. So $\hat{\beta}_0 = 0.1898 + 0.0023 \times 80.62 = 0.375$. This coefficient is significantly less than 0.4879, the coefficient on UNDERGRAD from (1). This is because the estimate from (1) does not account for the fact that people with an undergraduate degree would be more likely to earn more even if they did not graduate; there is an omitted variable (ability, measured here by AFQT) that biases the coefficient on UNDERGRAD upwards. The smaller estimate results from computing the direct effect of having an undergraduate degree (controlling for ability) and accounting for the fact that holders of degrees have a higher average AFQT.

2.

3. (a) Since $\mathbb{V}(X|D=1) = \mathbb{V}(X|D=0) = \sigma^2$, $\mathbb{E}[\mathbb{V}(X|D)] = \sigma^2$. Now

   $$\mathbb{V}(\mathbb{E}[X|D]) = Q(\mathbb{E}[X|D=1] - \mathbb{E}[X])^2 + (1-Q)(\mathbb{E}[X|D=0] - \mathbb{E}[X])^2$$

   And since $\mathbb{E}[X] = Q(\mathbb{E}[X|D=1]) + (1-Q)(\mathbb{E}[X|D=0])$, this becomes

   $$\mathbb{V}(\mathbb{E}[X|D]) = Q\{(1-Q)(\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0])\}^2 + (1-Q)\{Q(\mathbb{E}[X|D=0] - \mathbb{E}[X|D=1])\}^2$$
   $$\mathbb{V}(\mathbb{E}[X|D]) = Q(1-Q)\{\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0]\}^2.$$

   So $\mathbb{V}(X) = Q(1-Q)\{\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0]\}^2 + \sigma^2$.

   (b) We know $\lambda = \mathbb{C}(D,X)/\mathbb{V}(X)$, and $\mathbb{C}(D,X) = \mathbb{E}[DX] - \mathbb{E}[D]\mathbb{E}[X]$. So

   $$\mathbb{C}(D,X) = Q(\mathbb{E}[X|D=1]) - Q\{Q(\mathbb{E}[X|D=1]) + (1-Q)(\mathbb{E}[X|D=0])\}$$
   $$\mathbb{C}(D,X) = Q(1-Q)\{\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0]\}$$

   and

$$\lambda = \frac{Q(1-Q)\{\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0]\}}{Q(1-Q)\{\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0]\}^2 + \sigma^2}.$$

(c) If $\beta_0 = 0$, $\gamma_0 = \mathbb{C}(D,Y)/\mathbb{V}(D)$. By the same steps as (b), we can show $\mathbb{C}(D,Y) = Q(1-Q)\{\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]\}$. Since the variance of a binomial random variable is $p(1-p)$, $\gamma_0 = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$.

(d) If $\beta_0 = 0$, $E^*[Y|X] = \alpha_0 + \gamma_0(\kappa + \lambda X)$, so $b = \gamma_0 \lambda$, or

$$b = \frac{Q(1-Q)\{\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]\}\{\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0]\}}{Q(1-Q)\{\mathbb{E}[X|D=1] - \mathbb{E}[X|D=0]\}^2 + \sigma^2}.$$

(e) With these values, $\mathbb{V}(X) = 1.2$ and $b = 0.75$.

(f) A low β_0 but large b indicates that most of an individual's adult income is driven by their race, not by their parental income. Conditional on an individual's race, South Africa is highly mobile, in the sense that among white people, parental income is not predictive of adult income. Unconditional on race, South Africa is not highly mobile, in the sense that black adults are likely to have income similar to their parents and white adults are likely to have income similar to their parents, so in general people are likely to have income similar to their parents.

4. (a) For each of the ten water levels $X = x_l$, sort the 100 observations from smallest to largest, and then take the average of the 50th and 51st observations of lawn growth. This is $\hat{\pi}_l = Q_{Y|X}(1/2|X = x_l)$.

(b) We use the sample variance estimator $\hat{\sigma}_l^2 = N_l(X_{\lceil N_l\tau+l \rceil} - X_{\lfloor N_l\tau-l \rfloor})^2/4(1.96)^2$. Let $l^* = 1.96\sqrt{N_l\tau(1-\tau)} \approx 10$, with $N_l = 100$ and $\tau = 0.5$. and approximate $N_l/4(1.96)^2 = 6.25$. For each of the ten water levels, sort the 100 observations and take the difference of the 60th and 40th observations of lawn growth. Then, square the difference and multiply by 6.25. This is the variance estimate $\hat{\sigma}_l^2$.

(c) Compute the weighted least squares fit of $\hat{\pi}_l$ on x_l and x_l^2 using each of the 10 water levels $X = x_l$. We use $\hat{p}_l^2/\hat{\sigma}_l^2 = 1/(100\hat{\sigma}_l^2)$ as weights (which, since the groups are of equal size, is equivalent to just weighting by the inverse of the variance estimate). The coefficient on x_l will recover $\hat{\beta}_{1/2}$ and the coefficient on x_l^2 will recover $\hat{\gamma}_{1/2}$.

(d) We could instead estimate not the conditional median of lawn growth but, say, the 75th percentile. For the same watering level, the 75th percentile of lawn growth is larger than the 50th percentile, so it takes less water to achieve the same level of lawn growth. If the conditional quantile function is still quadratic in x, we can use the same estimation procedure to optimize the watering level to maximize lawn growth at the more conservative 75th percentile.

5. (a) $\lambda(y; X)$ is the hazard rate given X. That is, $\lambda(y; X) = Pr(Y = y|Y \geq y; X)$, so $\lambda(y; X = 1)$ is the probability that an individual retires at age y conditional on the individual not retiring before age y and conditional on the individual turning seventy during or after 1994. Likewise, $\lambda(y; X = 0)$ is the probability that an individual retires at age y conditional on the individual not retiring before age y and conditional on the individual turning seventy before 1994.

β_0 is the overall difference in the retirement hazard between an individual who turns seventy during or after 1994 and an individual who turned seventy before 1994. γ_0 is the difference in the retirement hazard *after age seventy* between an individual who turns seventy during or after 1994 and an individual who turned seventy before 1994.

(b) We expect $\beta_0 = 0$, that is, turning seventy after 1994 shouldn't affect the probability of retirement unless the individual is older than seventy, and thus the mandatory retirement policy actually affects them. We expect $\gamma_0 < 0$, which means that if an individual turns seventy after the mandatory retirement rules were prohibited, their probability of retiring each year after turning 70 should be lower than for an individual for whom the mandatory retirement rules apply.

From the figure, we see that the two lines track similarly for ages less than seventy, which suggests that indeed $\beta_0 = 0$. After age seventy, the line for which $X = 1$ is significantly smaller than the line for which $X = 0$, which suggests $\gamma_0 < 0$.

2

(c) This is the random censoring assumption, which states than an individual's likelihood of being censored is not affected by their true retirement age, conditional on X. We can imagine this being violated if, for some reason, we lose a group of people to follow-up in a particular year because of a policy that restricts them from disclosing when they retire, or if an idiosyncratic shock in a particular year causes a group of faculty to move abroad and thus be lost to follow-up.

(d) Note that in this notation, $D = 1$ indicates that an individual is *not* censored. Organize the data as follows:

S_{it}	Z_{60}	Z_{61}	Z_{62}	Z_{63}	Z_{64}	Z_{65}	...	Z_{70}	Z_{71}	Z_{72}	X	$1(y \geq 70) \times X$
0	1	0	0	0	0	0	...	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...		...
0	0	0	0	0	0	1	...	0	0	0	0	0
0	1	0	0	0	0	0	...	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...		...
1	0	0	0	0	0	0	...	0	0	1	0	0
0	1	0	0	0	0	0	...	0	0	0	1	0
1	0	1	0	0	0	0	...	0	0	0	1	0
0	1	0	0	0	0	0	...	0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...		...
0	0	0	0	0	0	0	...	1	0	0	1	1

We compute the logistic regression of S_{it} on $Z_{60} \ldots Z_{72}$, X, and $1(y \geq 70) \times X$. These coefficients are the effects of the independent variables on the log-odds of the hazard, so we apply the logistic transformation to recover estimates of $\alpha_{60} \ldots \alpha_{73}$, β_0 and γ_0.

(e) Say we lose some individuals to follow-up at age 65. Then, we will naively estimate the probability of retiring after age 65 to be lower than it actually is, given that some of those individuals would have retired after age 65. So $Pr(Y > y|X) < S(y; X)$.

An estimate $\hat{S}(y; X)$ can be constructed with $\prod_{i=1}^{y} (1 - \hat{\lambda}(y))$. So, for example,

$$\hat{S}(Y = 62; X = 1) = (1 - 0.030) \times (1 - 0.044) \times (1 - 0.089).$$

We could plot the survival functions for $X = 1$ and $X = 0$ separately, and estimate where they cross a horizontal line at 0.5; the difference in the y (age) values of these points is the effect of the end of mandatory retirement on the median retirement age.

Review Sheet 4

1. (a)

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(b) The equation implies that the productivity of a firm is dependent on the firm's inherent productivity α_0, the average productivity of all its suppliers (assuming it has suppliers), and a randomly distributed shock term V_i.

(c) The average productivity of firm 1's direct customers increases by σ/n, where n is the number of suppliers the firm has. So, the increase in productivity of firms 2, 3, and 4 is $\beta_0\sigma$ each, since they have no other direct suppliers, and the increase in productivity of firm 7 is $\beta_0\sigma/2$, since it has one other supplier.

(d) The effect of the shock on the *customers* of firms 2, 3, 4, and 7 – firms 5 and 6 – will be $(\beta_0)^2\sigma$ (since firms 5 and 6 only buy from these firms). Then, the effect on firm 7, which buys from firm 5, will be $(\beta_0)^3\sigma/2$.

(e) So, the aggregate increase productivity from a shock of σ to firm 1 is $\Delta A = \sigma + \frac{7}{2}\beta_0\sigma + 2(\beta_0)^2\sigma + \frac{1}{2}(\beta_0)^3\sigma$, so the social multiplier is $1 + \frac{7}{2}\beta_0 + 2(\beta_0)^2 + \frac{1}{2}(\beta_0)^3$.

(f) A shock to firm 6 only affects firm 6, since it has no customers, so the social multiplier is 1.

Important definitions and formulas

Projections

Projection theorem: $\hat{Y} = \pi(Y|L)$ is unique, and $\langle Y - \hat{Y}, \tilde{Y}\rangle = 0$.

LR: $\mathbb{E}^*[Y|X, W] = \alpha_0 + \beta_0 X + \gamma_0 W$. SR: $\mathbb{E}^*[Y|X] = a_0 + b_0 X$. AR: $\mathbb{E}^*[W|X] = \kappa_0 + \eta_0 X$. Then, $b_0 = \beta_0\eta_0$.

Frisch-Waugh: If $\tilde{W} = W - \mathbb{E}^*[W|X]$, then $E^*[Y|\tilde{W}] = \gamma_0\tilde{W}$.

LLN: $\sqrt{N}(\bar{Y} - \mu) \to N(0, \sigma^2) \Rightarrow T_N = \frac{\bar{Y}-\mu}{se(\bar{y})} \to N(0, 1)$.

Quantile regression

Quantile estimate: $\hat{\pi}_l = \frac{X_j + X_{j+1}}{2}$, with $j < (N_l + 1)\tau < j + 1$.

Variance of the quantile estimate: $\hat{\sigma}_l^2 = \frac{N_l(X_{\lceil N_l\tau + l^*\rceil} - X_{\lfloor N_l\tau - l^*\rfloor})^2}{4(z^{1-\alpha/2})^2}$, with $l^* = z^{1-\alpha/2}\sqrt{N_l\tau(1-\tau)}$.

Conditional quantile function: $Q_{Y|W}(\tau|W = w_l) = \alpha(\tau) + W'\beta(\tau)$. Estimate using WLS with $\hat{p}_l^2/\hat{\sigma}_l^2$ as weights.

Survival analysis

Hazard rate: $\lambda(y) = Pr(Y^* = y|Y^* \geq y)$.

Survival function: $S(y) = Pr(Y^* \geq y) = (1 - \lambda(1)) \times (1 - \lambda(2)) \times \ldots (1 - \lambda(y)) = \prod_{t=1}^{y}(1 - \lambda(y))$.

Variance of survival function: $\mathbb{V}(S(y)) = S(y)^2 \sum_{t=1}^{y}\frac{\lambda(t)}{(1-\lambda(t))N_t} = S(y)^2 \sum_{t=1}^{y}\frac{D_t}{N_t(N_t - D_t)}$.

Logit model: $\log\left(\frac{\lambda(y|X=x)}{1-\lambda(y|X=x)}\right) = \sum_{t=1}^{T}\alpha_t 1(y = t) + X'\beta$.

Networks

Row sum: outdegree $D_{i+} = \sum_{j\neq i} D_{ij}$. Column sum: indegree $D_{+i} = \sum_{j\neq i} D_{ji}$.

Utility depends on average peer action: $v_i(y; D) = (\alpha_0 + U_i)y_i + \beta_0\bar{y}_{n(i)}y_i$, where $\bar{y}_{n(i)} = \sum_{j\neq i} G_{ij}y_j$.

Social multiplier centrality: $c^{SM}(D, \beta_0) = 1_N(I_N - \beta_0 G)^{-1}$, where G is the row-normalized adjacency matrix.

Model of link formation: $D_{ij} = 1(W_{ij}'\beta_0 + A_i + A_j - U_{ij} \geq 0)$. $K + N$ parameters $\Rightarrow$ estimate using joint maximum likelihood or tetrad logit.