Quant 2, Week 7: Midterm Review

Sylvan Zheng

1 HW2 FWL Application

2. **(5 points)** Now suppose that the following system of linear equations holds for each graph:

(i)	(ii)	(iii)	(iv)
$V = \varepsilon_v$	$V = \varepsilon_v$	$V = \varepsilon_v$	$V = \varepsilon_v$
		$U = \varepsilon_u$	$U = \varepsilon_u$
$X = \varepsilon_x$	$X = V + \varepsilon_x$	$X = U + V + \varepsilon_x$	$X = U + V + \varepsilon_x$
$D = X + \varepsilon_d$	$D = X + \varepsilon_d$	$D = X + \varepsilon_d$	$D = X + U + \varepsilon_d$
$Y = D + X + V + \varepsilon_y$	$Y = D + X + V + \varepsilon_y$	$Y = D + X + V + \varepsilon_y$	$Y = D + X + V + \varepsilon_y$

- We know that (i)-(iii) are unbiased by DAG logic.
- But, we can also prove this formally using FWL and covariance algebra.
- Why does this work? Given the structural equations and the knowledge that the $\epsilon \sim N(0, 1)$, it's not too hard to calculate covariances and variances
- For example, in system (i) the variance of V is 1 and the variance of D is $Var(X + \epsilon_d) = 2$ (the sum because X and ϵ_d are independent).

Exercise (i)

- FWL estimate of OLS coefficient $\beta = Cov(\tilde{D}, \tilde{Y})/Var(\tilde{D})$
- $\tilde{D} = \epsilon_d$
- $\tilde{Y} = \epsilon_d + V + \epsilon_y$
- $\beta = \frac{Cov(\epsilon_d, \epsilon_d + V + \epsilon_y)}{Var(\epsilon_d)}$
- $= \frac{Cov(\epsilon_d, \epsilon_d) + Cov(\epsilon_d, V) + Cov(\epsilon_d, \epsilon_y)}{Var(\epsilon_d)}$
- $= \frac{Var(\epsilon_d) + 0 + 0}{Var(\epsilon_d)} = 1$

Exercise (iv)

For more complex cases, writing $\tilde{}(D)$ is not so easy. In these cases, better to express it formally as $\tilde{}(D) = D - \delta_0 - \delta_1 X$ (where δ are the OLS estimates from regressing D on X). Then, we also know that $\delta_1 = \frac{Cov(X,D)}{Var(X)}$. Similarly, we can write $\tilde{Y} = Y - \gamma_0 - \gamma_1 X$ where $\gamma_1 = \frac{Cov(Y,X)}{Var(X)}$.

- $\tilde{D} = D - \delta_0 - \delta_1 X$
- $\tilde{Y} = Y - \gamma_0 - \gamma_1 X$
- $\beta = \frac{Cov(\tilde{Y}, D - \delta_0 - \delta_1 X)}{Var(D - \delta_0 - \delta_1 X)}$

We can throw out the constants δ_0 and expand the numerator:

- $\beta = \frac{Cov(\tilde{Y},D) - Cov(\tilde{Y},\delta_1 X)}{Var(D-\delta_1 X)}$

We know that $\tilde{Y}$ is uncorrelated with X:

- $\beta = \frac{Cov(\tilde{Y},D)}{Var(D-\delta_1 X)} = \frac{Cov(Y-\gamma_1 X,D)}{Var(D-\delta_1 X)} = \frac{Cov(Y,D)-\gamma_1 Cov(X,D)}{Var(D-\delta_1 X)}$

Then expanding the denominator:

- $\beta = \frac{Cov(Y,D)-\gamma_1 Cov(X,D)}{Var(D)+\delta_1^2 Var(X) - 2\delta_1 Cov(X,D)}$

From here we could do more simplification to reduce future calculation but at this point we can already calculate β. We just need the variances and covariances of the variables.

- $Var(X) = Var(U + V + \epsilon_x) = 3$ (all components of X are independent)
- $Cov(X,U) = Var(U) = 1 = Cov(X,V)$
- $Cov(X,D) = Cov(X, X + U + \epsilon_d) = Var(X) + cov(X,U) = 4$
- $Var(D) = Var(\epsilon_d) + Var(X) + Var(U) + 2Cov(X,U) = 1 + 3 + 1 + 2 = 7$
- $Cov(Y,D) = Cov(D + X + V + \epsilon_y, D) = Var(D) + Cov(X,D) + Cov(V,D) + 0 = 7 + 4 + 1 = 12$

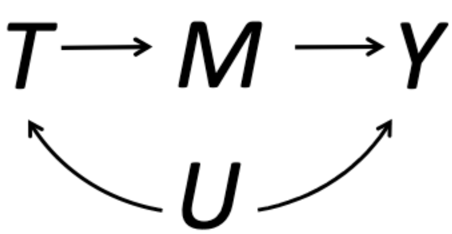
Plugging in:

```
var.d <- 7
var.x <- 3
cov.xd <- 4
cov.yd <- 12
cov.yx <- 8
gamma <- cov.yx/var.x
delta <- cov.xd/var.x
beta <- (cov.yd - gamma * cov.xd) / (var.d + delta * delta * var.x -  2 * delta * cov.xd)
beta
```

```
## [1] 0.8
```

Extension to FDC

We can also apply this to a set of structural equations to illustrate how FDC works.



- $U = \epsilon_u$
- $T = U + \epsilon_t$
- $M = T + \epsilon_m$
- $Y = M + U + \epsilon_y$

We can identify the effect of T on M in an unbiased way. This is just the OLS result $\beta_T = 1$

Then we can identify the effect of M on Y controlling for T.

$$\beta_M = \frac{Cov(\tilde{Y}, \tilde{M})}{Var(\tilde{M})}$$

$$= \frac{Cov(Y - \gamma_1 T, \epsilon_m)}{1} = Cov(Y, \epsilon_m) - Cov(\gamma_1 T, \epsilon_m)$$

T, ϵ_m independent:

$$\beta_M = Cov(T + \epsilon_m + U + \epsilon_y, \epsilon_m) - 0 = 1$$

The total effect is $1 \cdot 1 = 1$.

Non Parametric FDC - ATT derivation

Setup: Binary M and T. Y_i is a function of M_i which is a function of T. So we write $Y_i(M_i(\cdot))$

$$ATT = \mathbb{E}\left[Y_i(M_i(1)) \mid T_i = 1\right] - \mathbb{E}\left[Y_i(M_i(0)) \mid T_i = 1\right]$$

Call the first term A and the second term B.

A is the expected value of Y given $T = 1$ (the observed Y for the treated group). For some people, $M = 0$ even if $T = 1$. For these people, we would write $M_i(1) = 0$. For other units, $M_i(1) = 1$ which means they are actually treated via the mechanism M. We can decompose into these two parts using the law of total probability.

$$\begin{aligned}
A &= \mathbb{E}\left[Y_i(1) \mid M_i(1) = 1, T_i = 1\right] \Pr[M_i(1) = 1 \mid T_i = 1] \\
&\quad + \mathbb{E}\left[Y_i(0) \mid M_i(1) = 0, T_i = 1\right] \Pr[M_i(1) = 0 \mid T_i = 1] \\
&= \mathbb{E}\left[Y_i \mid M_i = 1, T_i = 1\right] \Pr[M_i = 1 \mid T_i = 1] \\
&\quad + \mathbb{E}\left[Y_i \mid M_i = 0, T_i = 1\right] (1 - \Pr[M_i = 1 \mid T_i = 1])
\end{aligned}$$

Ok. Now B is the counterfactual component. This is the expected value of Y if $T = 0$ **for the units where T is actually 1**. However, we can start by doing the same decomposition. For some units, $M_i(0) = 1$ which means that they would have received $M = 1$ even if T had been 0 . For other units, $M_i(0) = 0$ which means that they would have received $M = 0$ even if T had been 0 .

$$\begin{aligned}
B &= \mathbb{E}\left[Y_i(1) \mid M_i(0) = 1, T_i = 1\right] \Pr[M_i(0) = 1 \mid T_i = 1] \\
&\quad + \mathbb{E}\left[Y_i(0) \mid M_i(0) = 0, T_i = 1\right] \Pr[M_i(0) = 0 \mid T_i = 1]
\end{aligned}$$

Now we use our assumptions about FDC setup to get leverage on these counterfactual quantities.

First, the exclusion restriction means that T doesn't affect Y except through M. So the potential outcomes Y for the group $M_i(0) = 1$ are the same for the group where $M_i(1) = 1$. So, $\mathbb{E}[Y_i(1) \mid M_i(0) = 1, T_i = 1] = \mathbb{E}[Y_i(1) \mid M_i(1) = 1, T_i = 1]$

Next, the unconfounded assumption means that $\Pr[M_i(0) = 1 \mid T_i = 1] = \Pr[M_i(0) = 1 \mid T_i = 0]$

$$\begin{aligned} B &= \mathbb{E}[Y_i(1) \mid M_i(1) = 1, T_i = 1] \Pr[M_i(0) = 1 \mid T_i = 0] \\ &\quad + \mathbb{E}[Y_i(0) \mid M_i(1) = 0, T_i = 1] \Pr[M_i(0) = 0_| T_i = 0] \\ &= \mathbb{E}[Y_i \mid M_i = 1, T_i = 1] \Pr[M_i = 1 \mid T_i = 0] \\ &\quad + \mathbb{E}[Y_i \mid M_i = 0, T_i = 1] (1 - \Pr[M_i = 1 \mid T_i = 0]). \end{aligned}$$

Random Sampling - Variance Derivation

Consider a simple randomized experiment.

We draw a random sample of units indexed by $i \in 0, 1, ...N$ from a finite population of size n. We randomly assign $D \in 0, 1$ and observe $Y_{i,D}$ for each unit. $N_1 = \sum_{i=1}^{N} \mathbb{1}(D_i == 1)$ (the number of units assigned to treatment) and $N_0 = N - N_1 = \sum_{i=1}^{N} \mathbb{1}(D_i == 0)$ (the number of units assigned to control).

Then, the difference in means estimator $\hat{\rho} = \frac{1}{N_1} \sum_{i=1;D_i=1}^{N} Y_i - \frac{1}{N_0} \sum_{i=1;D_i=0}^{N} Y_i$

There is uncertainty in $\hat{\rho}$ from two sources. One is sampling uncertainty, because we randomly selected our sample from the population. The other is design-based uncertainty, because we randomly assigned treatment outcomes.

Let's derive the total variance $Var(\hat{\rho})$ with respect to both sources of uncertainty.

Applying the ANOVA theorem :

$$Var[\hat{\rho}] = E_S[Var_D[\hat{\rho}|S]] + Var_S[E_D[\hat{\rho}|S]]$$

Let's focus on the second term. The second term is the random-sample variance of the expected value of $\hat{\rho}$ over different treatment assignments. Given a fixed sample S, this is equivalent to the sample ATE.

$$= E_S[Var_D[\hat{\rho}|S]] + Var_S[\frac{1}{N} \sum_{i=1}^{N} Y_{i1} - Y_{i0}]$$

$$= E_S[Var_D[\hat{\rho}|S]] + Var_S[\frac{1}{N} \sum_{i=1}^{N} \rho_i]$$

$$= E_S[Var_D[\hat{\rho}|S]] + \frac{1}{N^2} Var_S[\sum_{i=1}^{N} \rho_i]$$

Because the ρ_i are independent (SUTVA), the variance of sums is the same as the sum of variances:

$$= E_S[Var_D[\hat{\rho}|S]] + \frac{1}{N^2} \sum_{i=1}^{N} Var_S[\rho_i]$$

Then, define $\sigma_\rho^2 = \frac{1}{N} \sum_{i=1}^{N} Var_S[\rho_i]$. This is the average population variance of ρ.

$$= E_S[Var_D[\hat{\rho}|S]] + \frac{1}{N}\sigma_\rho^2$$

Note that this term is not identified, since we can't actually observe any $\rho_i = Y_{i1} - Y_{i0}$.

Ok. Now for the first term which is the expected value of $Var_D[\hat{\rho}|S]$ over the random sampling process. First, we can just think about $Var_D[\hat{\rho}|S]$ (ie, with a fixed sample S).

$$Var_D[\hat{\rho}|S] = Var_D[\bar{Y_1} - \bar{Y_0}] = Var_D[\bar{Y_1}] + Var_D[\bar{Y_0}] - 2Cov(\bar{Y_1}, \bar{Y_0})$$

$$= \frac{1}{N_1}s_{Y1}^2 * \frac{N - N_1}{N} + \frac{1}{N_0}s_{Y0}^2 * \frac{N - N_0}{N} - 2[-s_{Y1,Y0}/N]$$

Ok what's going on here? s_{Y1}^2 is the sample variance of $Y1$ and $s_{Y1,Y0}$ is the sample covariance.

The population correction factors $\frac{N-N_1}{N}$ arise because of a finite population correction. "Complete random assignment is like SRSWOR for potential outcomes"

$$= \frac{1}{N_1}s_{Y1}^2 - \frac{N_1}{N}s_{Y1}^2 + \frac{1}{N_0}s_{Y0}^2 - \frac{N_0}{N}s_{Y0}^2 - 2[-s_{Y1,Y0}/N]$$

$$= \frac{1}{N_1}s_{Y1}^2 + \frac{1}{N_0}s_{Y0}^2 - \frac{1}{N}\left(s_{Y1}^2 + s_{Y0}^2 - 2[s_{Y1,Y0}]\right)$$

$$= \frac{1}{N_1}s_{Y1}^2 + \frac{1}{N_0}s_{Y0}^2 - \frac{1}{N}s_\rho^2$$

Then, taking expectations over the random sampling:

$$= \frac{1}{N_1}\sigma_{Y1}^2 + \frac{1}{N_0}\sigma_{Y0}^2 - \frac{1}{N}\sigma_\rho^2$$

Then, the last term cancels out with what we derived earlier and the total variance is:

$$= \frac{1}{N_1}\sigma_{Y1}^2 + \frac{1}{N_0}\sigma_{Y0}^2$$