ECMA 31360, PSet 3: Solutions

YOUR NAMES, University of Chicago

XX-XX-2026

([ ] out of 10p) PART I: Intent-to-Treat Versus Average Treatment Effect

([ ] out of 10p) Q1: An Exercise to Deep-dive the Difference b/w ITT and ATE

([ ] out of 2p) Q1.a: Describe ATE^o

([ ] out of 1p) Q1.b: Express $(Y_i^o(1), Y_i^o(0))$ as functions of $(Y_i(1), Y_i(0))$

([ ] out of 1p) Q1.c: Analytical Relationship between ATE^o and ATE

([ ] out of 2p) Q1.d: Selection into the NSW program based on flipping an unbalanced coin

---

([ ] out of 2p) Q1.e: Selection into the NSW program based on gains

---

([ ] out of 2p) Q1.f: When ITT and ATE Differ

([ ] out of 10p) PART II: Describe the Pseudo-Observational NSW Data

([ ] out of 3p) Q2: Compute Sample Averages of OPVs and outcome variable for PSID-1 and CPS-1 Samples

Script and Output

([ ] out of 4p) Q3: Compare the PSID-1 and CPS-1 Comparison Groups to the NSW-Treated Sample

---

([ ] out of 4p) **Q4: Why do Dehajia and Wahba mimic observational data by combining the NSW-treated sample with survey data?**

([ ] out of 10p bonus) PART III: Target Estimand: ATE versus ATT of the NSW Offer

([ ] out of 2p) Q5.a

([ ] out of 2p) Q5.b

([ ] out of 4p) Q5.c

([ ] out of 2p) Q5.d

([ ] out of 80p) PART IV: Regression-based Estimation of the Effect of the NSW Offer based on Pseudo-Observational Data

([ ] out of 9p) Q6: Implement the Treated-Control Comparison Estimator

([ ] out of 1p) Q6.a: Get the DM estimate and its SE under Homoskedasticity

Script and output

```r
library(data.table)
library(dplyr)
library(lmtest)
library(sandwich)

# Load pseudo-observational data
df <- fread("nswpsid.csv")

# Per pset guidance: scale re74 and re75 by 1,000
df <- df %>% mutate(
  re74 = re74 / 1000,
  re75 = re75 / 1000
)

# DM regression (Specification 1)
m1 <- lm(re78 ~ treat, data = df)
```

```r
# Point estimate and homoskedastic SE
summary(m1)$coefficients["treat", c("Estimate", "Std. Error")]
```

```
##   Estimate Std. Error
## -15204.776   1154.614
```

```r
# (Optional) full regression output
summary(m1)
```

```
##
## Call:
## lm(formula = re78 ~ treat, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21554  -9732   -866   7705  99620
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21553.9      303.6   70.98   <2e-16 ***
## treat       -15204.8     1154.6  -13.17   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15150 on 2673 degrees of freedom
## Multiple R-squared:  0.06092,    Adjusted R-squared:  0.06057
## F-statistic: 173.4 on 1 and 2673 DF,  p-value: < 2.2e-16
```

([ ] out of 3p) Q6.b: Get the Heteroskedasticity-robust SE of the DM estimator

Script and output

```r
# Heteroskedasticity-robust (HC0) SE
coeftest(m1, vcov. = vcovHC(m1, type = "HC0"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  21553.92     311.67  69.157 < 2.2e-16 ***
## treat       -15204.78     655.67 -23.190 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# (Optional) extract just the robust SE of treat
sqrt(diag(vcovHC(m1, type = "HC0")))[["treat"]]
```

```
## [1] 655.6691
```

([ ] out of 5p) Q6.c: Discuss DM Estimator of ATT of NSW Offer

The difference-in-means (DM) estimator compares average post-treatment earnings in 1978 between the treated group (units with $D_i = 1$) and the control group (units with $D_i = 0$). In population terms, the DM estimand is

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0].$$

Using the potential outcomes framework and the measurement equation $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$, this estimand can be written as

$$\begin{aligned}\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] &= \mathbb{E}[Y_i(1) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0] \\ &= \underbrace{\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]}_{\text{ATT}} + \underbrace{(\mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0])}_{\text{confounding term}}.\end{aligned}$$

Therefore, the DM estimator identifies the average treatment effect on the treated (ATT) if and only if

$$\mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_i(0) \mid D_i = 0],$$

that is, if untreated potential outcomes are mean-independent of treatment assignment (unconditional mean independence).

In the pseudo-observational setting considered here, treatment is not randomly assigned: treated units (from the NSW experimental sample) and control units (from the PSID) differ systematically in baseline characteristics and pre-treatment earnings. As a result, the above condition is unlikely to hold, and the confounding term is generally nonzero. Consequently, the DM estimator does not have a credible causal interpretation as the ATT of the NSW offer in this context.

([ ] out of 8p) Q7: Prove Claim about identification and estimation of ATT

We prove Claim 1.

(i) Under (a)–(b), ATT is identified.

By definition,

$$ATT := \mathbb{E}[Y(1) - Y(0) \mid D = 1] = \mathbb{E}[\,\mathbb{E}(Y(1) - Y(0) \mid D = 1, X) \mid D = 1]\,.$$

For treated units ($D = 1$), the observed outcome equals the treated potential outcome:

$$Y = Y(1) \quad \text{when } D = 1 \;\Rightarrow\; \mathbb{E}[Y(1) \mid D = 1, X] = \mathbb{E}[Y \mid D = 1, X].$$

Assumption (b) (CMIA0) states

$$\mathbb{E}[Y(0) \mid D = 1, X = x] = \mathbb{E}[Y(0) \mid D = 0, X = x] \;\; \forall x \in \mathcal{X}.$$

For control units ($D = 0$), the observed outcome equals the untreated potential outcome:

$$Y = Y(0) \quad \text{when } D = 0 \;\Rightarrow\; \mathbb{E}[Y(0) \mid D = 0, X] = \mathbb{E}[Y \mid D = 0, X].$$

Combining these,

$$\mathbb{E}[Y(1) - Y(0) \mid D = 1, X] = \mathbb{E}[Y \mid D = 1, X] - \mathbb{E}[Y \mid D = 0, X].$$

Hence,

$$\boxed{ATT = \mathbb{E}[\,\mathbb{E}[Y \mid D = 1, X] - \mathbb{E}[Y \mid D = 0, X] \mid D = 1]\,.}$$

Assumption (a) (COC) guarantees that for each x in the relevant support there exist both treated and control units, so the conditional expectations above are well-defined.

(ii) Under (a)–(d), the observed conditional expectation function is linear:

$$\mathbb{E}[Y \mid D = d, X = x] = \alpha + \rho d + \beta' x.$$

When $D = 0$, we observe $Y = Y(0)$, and by (c),

$$\mathbb{E}[Y \mid D = 0, X = x] = \mathbb{E}[Y(0) \mid X = x] = \alpha_0 + \theta_0' x.$$

When $D = 1$, we observe $Y = Y(1)$, and by (d),

$$\mathbb{E}[Y \mid D = 1, X = x] = \mathbb{E}[Y(1) \mid D = 1, X = x] = \alpha_1 + \gamma + \theta_1' x.$$

With the homogeneity restriction $\theta_0 = \theta_1 = \theta$, define

$$\alpha := \alpha_0, \quad \beta := \theta, \quad \rho := (\alpha_1 + \gamma - \alpha_0).$$

Then for $d \in \{0, 1\}$ we can write

$$\mathbb{E}[Y \mid D = d, X = x] = \alpha + \rho d + \beta' x,$$

which is linear in (d, x).

(iii) $\hat{\rho}$ is a consistent estimator of ρ.

Given the correct linear specification in (ii), i.i.d. sampling, and the usual OLS regularity conditions (in particular, no perfect multicollinearity), the OLS estimator $\hat{\rho}$ from the regression of Y on $(1, D, X)$ is consistent for the population coefficient ρ.

(iv) Under (a)–(d), $\rho = ATT$.

Using the conditional means above and $\theta_0 = \theta_1 = \theta$,

$$\mathbb{E}[Y(1) - Y(0) \mid D = 1, X = x] = (\alpha_1 + \gamma + \theta' x) - (\alpha_0 + \theta' x) = \alpha_1 + \gamma - \alpha_0 = \rho.$$

Since this expression does not depend on x,

$$ATT = \mathbb{E}[\mathbb{E}[Y(1) - Y(0) \mid D = 1, X] \mid D = 1] = \mathbb{E}[\rho \mid D = 1] = \rho.$$

Therefore, $\hat{\rho} \xrightarrow{p} \rho = ATT$, i.e., the OLS coefficient on D consistently estimates ATT under the stated assumptions.

([ ] out of 8p) Q8: Implement the Regression-Adjusted DM Estimator

([ ] out of 3p) Q8.a: Get the Adj.DM estimate and SE under Heteroskedasticity

Script and Output

```r
library(dplyr)
library(lmtest)
library(sandwich)

# Assume df is already loaded in Q6 and re74/re75 already scaled by 1,000.
# If not, uncomment the next lines:
# library(data.table)
# df <- data.table::fread("nswpsid.csv") %>% mutate(re74 = re74/1000, re75 = re75/1000)

# Create agesq as instructed
df <- df %>% mutate(agesq = age^2)

# Specification 2 (Adj-DM / Regression Adjustment)
m2 <- lm(re78 ~ treat + age + agesq + edu + nodegree + black + hisp + re74 + re75, data = df)

# Robust SE (White / HC0)
coeftest(m2, vcov. = vcovHC(m2, type = "HC0"))
```

```
##
## t test of coefficients:
##
##                  Estimate Std. Error t value  Pr(>|t|)
```

```
## (Intercept) -2836.7030  2931.8892 -0.9675   0.33336
## treat         217.9438   766.4444  0.2844   0.77616
## age           158.5058   150.7480  1.0515   0.29314
## agesq          -3.2329     2.0994 -1.5399   0.12370
## edu           564.6237   121.4207  4.6501 3.479e-06 ***
## nodegree      502.0912   631.1854  0.7955   0.42641
## black        -699.3353   431.6491 -1.6201   0.10532
## hisp         2226.5351  1216.7992  1.8298   0.06739 .
## re74          279.1682    61.7644  4.5199 6.457e-06 ***
## re75          568.0874    66.2753  8.5716 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# (Optional) extract only treat coef and its robust SE
c(
  rho_hat = coef(m2)[["treat"]],
  se_hc0  = sqrt(diag(vcovHC(m2, type = "HC0")))[["treat"]]
)
```

```
##  rho_hat   se_hc0
## 217.9438 766.4444
```

([ ] out of 5p) Q8.b: May the Adj.DM estimator improve over the DM estimator?

Yes, the Adj-DM estimator *may* improve over the DM estimator.

The DM estimator compares $\mathbb{E}[Y \mid D=1] - \mathbb{E}[Y \mid D=0]$ and identifies ATT only under a strong unconditional comparability condition such as

$$\mathbb{E}[Y(0) \mid D=1] = \mathbb{E}[Y(0) \mid D=0],$$

which is implausible in the pseudo-observational setting because treated (NSW) and controls (PSID) differ systematically in baseline characteristics and pre-treatment earnings.

In contrast, the adjusted difference-in-means (regression adjustment) estimates ρ from

$$Y_i = \alpha + \rho D_i + \beta' X_i + u_i,$$

where $X_i = (age_i, age_i^2, edu_i, nodegree_i, black_i, hisp_i, re74_i, re75_i)$. This approach attempts to account for *observable* differences between treated and controls. Under the conditions in Claim 1—in particular overlap (COC), conditional mean independence of untreated potential outcomes (CMIA0), and correct linear specification with homogeneous slopes—the population coefficient ρ equals ATT, and OLS provides a consistent estimator of ATT.

However, improvement is not guaranteed: if the linear functional form is misspecified or if there are important *unobserved* confounders not captured by X, the Adj-DM estimator may still be biased. —

([ ] out of 6p) Q9: Verify the Partialling-out Interpretation of OLS

([ ] out of 2p) Q9.a: Implement Procedure B

Script and Output

```r
library(lmtest)
library(sandwich)

# We assume you already ran Q8 and created:
#   df (data frame) and m2 (Spec 2 regression)
# If not, run Q8 chunk first.
```

```r
# Procedure B:
# 1) Regress D on X and keep residuals vhat
m_D_on_X <- lm(treat ~ age + agesq + edu + nodegree + black + hisp + re74 + re75, data = df)
vhat <- resid(m_D_on_X)

# 2) Regress Y on vhat; slope should equal rho from the full regression
m_B <- lm(re78 ~ vhat, data = df)

rho_A <- coef(m2)[["treat"]]      # from full regression (Spec 2)
rho_B <- coef(m_B)[["vhat"]]      # from Procedure B
c(rho_A = rho_A, rho_B = rho_B, diff = rho_A - rho_B)
```

```
##         rho_A         rho_B          diff
##  2.179438e+02  2.179438e+02 -9.350742e-12
```

```r
summary(m_B)
```

```
##
## Call:
## lm(formula = re78 ~ vhat, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20705 -11277  -1067   8326 100724
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20502.4      302.3  67.820   <2e-16 ***
## vhat           217.9     1344.3   0.162    0.871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15640 on 2673 degrees of freedom
## Multiple R-squared:  9.833e-06,  Adjusted R-squared:  -0.0003643
## F-statistic: 0.02629 on 1 and 2673 DF,  p-value: 0.8712
```

([ ] out of 2p) Q9.b: Implement Procedure C

Script and Output

```r
# Procedure C:
# 1) Regress D on X and keep residuals vhat (reuse vhat if already computed)

# 2) Regress Y on X and keep residuals ehat
m_Y_on_X <- lm(re78 ~ age + agesq + edu + nodegree + black + hisp + re74 + re75, data = df)
ehat <- resid(m_Y_on_X)

# 3) Regress ehat on vhat (often with no intercept)
m_C <- lm(ehat ~ 0 + vhat, data = df)

rho_C <- coef(m_C)[["vhat"]]
c(rho_A = rho_A, rho_C = rho_C, diff = rho_A - rho_C)
```

```
##         rho_A         rho_C          diff
##  2.179438e+02  2.179438e+02 -1.082867e-11
```

```r
summary(m_C)
```

```
##
## Call:
## lm(formula = ehat ~ 0 + vhat, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -64762  -4325   -523   3823 110481
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## vhat    217.9      864.7   0.252    0.801
##
## Residual standard error: 10060 on 2674 degrees of freedom
## Multiple R-squared:  2.375e-05,  Adjusted R-squared:  -0.0003502
## F-statistic: 0.06352 on 1 and 2674 DF,  p-value: 0.801
```

([ ] out of 2p) Q9.c: Interpretation

The Frisch–Waugh–Lovell (FWL) theorem implies that the coefficient on D_i (here $treat_i$) in the multiple regression

$$Y_i = \alpha + \rho D_i + \beta' X_i + u_i$$

can be obtained by *partialling out* the covariates X_i from both D_i and Y_i.

In particular, let $\hat{v}_i$ be the residual from regressing D_i on X_i:

$$D_i = \pi' X_i + v_i, \qquad \hat{v}_i = D_i - \widehat{\pi' X_i}.$$

Procedure B then regresses Y_i on $\hat{v}_i$ and the resulting slope equals $\hat{\rho}$. Equivalently, let $\hat{\varepsilon}_i$ be the residual from regressing Y_i on X_i:

$$Y_i = \delta' X_i + \varepsilon_i, \qquad \hat{\varepsilon}_i = Y_i - \widehat{\delta' X_i}.$$

Procedure C regresses $\hat{\varepsilon}_i$ on $\hat{v}_i$ (often without an intercept) and the slope again equals $\hat{\rho}$.

Thus, partialling-out means that OLS identifies the effect of D using only the variation in D that is orthogonal (linearly unrelated) to X, and the corresponding orthogonal component of Y. Numerically, Procedures B and C should reproduce exactly the same $\hat{\rho}$ as the full regression (up to numerical precision). —

([ ] out of 18p) Q10: Implement the DML Estimator to estimate the partially-linear specification

([ ] out of 2p) Q10.a: Why may we prefer Specification 3 over Specification 2?

([ ] out of 2p) Q10.b: Relate Procedure D to Procedure C

([ ] out of 1p) Q10.c: Install packages (here or at the top)

Script and Output

([ ] out of 1p) Q10.d: Convert to `data.table` (here or above)

Script and Output

([ ] out of 1p) Q10.e: Collect OPVs names (here or above)

Script and Output

([ ] out of 2p) Q10.f: Specify the "causal model"

Script and Output

Commentary

([ ] out of 1p) Q10.g: Suppress messages

Script and Output

([ ] out of 2p) Q10.h: Specify learners from $m(\cdot)$ and $l(\cdot)$

Script and Output

([ ] out of 2p) Q10.i: Initialize DML object

Script and Output

([ ] out of 2p) Q10.j: Estimate the Partial-Linear Model

Script and Output

([ ] out of 2p) Q10.k: DML estimate of ATT of NSW Offer

Script and Output

([ ] out of 10p) Q11: Compare ATT Estimates based on pseudo-observational data to Experimental estimates

Commentary

([ ] out of 21p) Q12: Imbens and Xu's Defense: The importance of overlap and the use of trimming