Homework 2 Solutions

Hansen 2.2

$$\begin{aligned}
\mathbb{E}[YX] &= \mathbb{E}[X\mathbb{E}[Y|X]] \\
&= \mathbb{E}[X(a+bX)] \\
&= \mathbb{E}[aX+bX^2] \\
&= a\mathbb{E}[X]+b\mathbb{E}[X^2]
\end{aligned}$$

where the first equality holds by the law of iterated expectations, the second equality holds by the expression for $\mathbb{E}[Y|X]$ in the problem, and the remaining two equalities are just algebra/basic properties of expectations.

Hansen 2.5

a) The mean squared error is given by

$$MSE = \mathbb{E}[(e^2 - h(X))^2]$$

b) e^2 is closely related to a measure of the magnitude of how far off our predictions of Y given X are. For example, given X, if we predict a "high" value of e^2, it would suggest that we expect our predictions of Y to not be too accurate for that value of X.

c) Recall that $\sigma^2(X) = \mathbb{E}[e^2|X]$ so that

$$\begin{aligned}
MSE &= \mathbb{E}\Big[((e^2 - \mathbb{E}[e^2|X]) - (h(X) - \mathbb{E}[e^2|X]))^2\Big] \\
&= \mathbb{E}\Big[(e^2 - \mathbb{E}[e^2|X])^2\Big] - 2\mathbb{E}\Big[(e^2 - \mathbb{E}[e^2|X])(h(X) - \mathbb{E}[e^2|X])\Big] + \mathbb{E}\Big[(h(X) - \mathbb{E}[e^2|X])^2\Big]
\end{aligned}$$

Let's consider each of these three terms.

- The first term does not depend on $h(X)$ so it is invariant to our choice of h.
- The second term is equal to 0 after applying the law of iterated expectations.
- The third term is minimized by setting $h(X) = \mathbb{E}[e^2|X] = \sigma^2(X)$ which implies that MSE is minimized by $\sigma^2(X)$.

Hansen 2.6

To start with, notice that

$$\mathbb{E}[Y] = \mathbb{E}[m(X) + e] = \mathbb{E}[m(X)]$$

where the last equality holds because $\mathbb{E}[e] = 0$. Thus, we have that

$$\begin{aligned}
\mathrm{var}(Y) &= \mathbb{E}\Big[(Y - \mathbb{E}[Y])^2\Big] \\
&= \mathbb{E}\Big[(m(X) + e - \mathbb{E}[m(X)])^2\Big] \\
&= \mathbb{E}\Big[(m(X) - \mathbb{E}[m(X)])^2\Big] + 2\mathbb{E}\Big[(m(X) - \mathbb{E}[m(X)])e\Big] + \mathbb{E}[e^2] \\
&= \mathbb{E}\Big[(m(X) - \mathbb{E}[m(X)])^2\Big] + \mathbb{E}[e^2] \\
&= \mathrm{var}[m(X)] + \sigma^2
\end{aligned}$$

where first equality holds by the definition of $\mathrm{var}(Y)$, the second equality holds by the expression for $\mathbb{E}[Y]$ in the previous display, the third equality holds by expanding the square, the fourth equality holds by applying the law of iterated expectations to the middle term (and because $\mathbb{E}[e|X] = 0$), and the fifth equality holds by the definition of $\mathrm{var}[m(X)]$ and the fact that $\mathbb{E}[e^2] = \sigma^2$.

Hansen 2.10

True.

$$\mathbb{E}[X^2 e] = \mathbb{E}[X^2 \underbrace{\mathbb{E}[e|X]}_{=0}] = 0$$

Hansen 2.11

False. Here is a counterexample. Suppose that $X \sim \mathcal{N}(0,1)$ and $e|X \sim \mathcal{N}(X^2 - 1, 1)$. Recall that, if $X \sim \mathcal{N}(0,1)$, then $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = 1$, and $\mathbb{E}[X^3] = 0$, and $\mathbb{E}[X^4] = 3$. Notice that $\mathbb{E}[e] = \mathbb{E}[\mathbb{E}[e|X]] = \mathbb{E}[X^2 - 1] = 0$ and $\mathbb{E}[Xe] = \mathbb{E}[X\mathbb{E}[e|X]] = \mathbb{E}[X(X^2 - 1)] = 0$. However, $\mathbb{E}[X^2 e] = \mathbb{E}[X^2 \mathbb{E}[e|X]] = \mathbb{E}[X^2(X^2 - 1)] = 2 \neq 0$.

As a side-comment, perhaps it is helpful to explain how I came up with this counterexample. Remember that $\mathbb{E}[Xe]$ is the correlation between X and e—i.e., it is a scalar, linear measure of the relationship between X and e. On the other hand, $\mathbb{E}[e|X]$ is a function of X. To make X and e uncorrelated, but $\mathbb{E}[X^2 e] = 0$, we need to think of a case where $\mathbb{E}[e|X]$ is a function of X that is not linear in X. Using a normal distribution for X and having $\mathbb{E}[e|X]$ be a quadratic function of X is a natural choice, and then I just chose $\mathbb{E}[e|X] = X^2 - 1$ to make the math work out.

Hansen 2.12

False. Here is a counterexample. Suppose that $\mathbb{E}[e^2|X]$ depends on X, then e and X are not independent. As a concrete counterexample, suppose $e|X \sim \mathcal{N}(0, X^2)$ (that is, conditional on X, e follows a normal distribution with mean 0 and variance X^2). In this case $\mathbb{E}[e|X] = 0$, but e and X are not independent.

Hansen 2.13

False. The same counterexample as in 2.11 works here. In that case, $\mathbb{E}[Xe] = 0$, but $\mathbb{E}[e|X] = X^2 - 1$ 0.

Hansen 2.14

False. In this case, higher order moments can still depend on X. For example, $\mathbb{E}[e^3|X]$ can still depend on X. If it does, then e and X are not independent.

Hansen 2.21

a) Following omitted variable bias types of arguments (also, notice that the notation in the problem implies that X is scalar here), we have that

$$\begin{aligned}\gamma_1 &= \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} \\ &= \frac{\mathbb{E}[X(X\beta_1 + X^2\beta_2 + u)]}{\mathbb{E}[X^2]} \\ &= \beta_1 + \frac{\mathbb{E}[X^3]}{\mathbb{E}[X^2]}\beta_2\end{aligned}$$

Thus, $\gamma_1 = \beta_1$ if either $\beta_2 = 0$ or $\mathbb{E}[X^3] = 0$. $\beta_2 = 0$ if X^2 does not have an effect on the outcome (after accounting for the effect of X); this is similar to the omitted variable logic that we talked about in class. A leading case where $\mathbb{E}[X^3] = 0$ is when X is a mean 0 symmetric random variable; for example, if X is standard normal, then its third moment is equal to 0.

b) Using the same arguments as in part (a), we have that

$$\gamma_1 = \theta_1 + \frac{\mathbb{E}[X^4]}{\mathbb{E}[X^2]}\theta_2$$

Similar to the previous part, γ_1 could equal θ_1 if θ_2 were equal to 0. Unlike the previous part though, here, we cannot have that $\mathbb{E}[X^4] = 0$ except in the degenerate case where $X = 0$ with probability 1 (which would be ruled out here as it would also imply that $\mathbb{E}[X^2] = 0$).

Extra Question 1

a)

$$\begin{aligned}ATE &= \mathbb{E}[Y(1) - Y(0)] \\ &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\ &= \mathbb{E}\Big[\mathbb{E}[Y(1)|X]\Big] - \mathbb{E}\Big[\mathbb{E}[Y(0)|X]\Big] \\ &= \mathbb{E}\Big[\mathbb{E}[Y(1)|X, D=1]\Big] - \mathbb{E}\Big[\mathbb{E}[Y(0)|X, D=0]\Big] \\ &= \mathbb{E}\Big[\mathbb{E}[Y|X, D=1]\Big] - \mathbb{E}\Big[\mathbb{E}[Y|X, D=0]\Big]\end{aligned}$$

where the first equality is the definition of ATE, the second equality pushes the expectation through the difference, the third equality holds by the law iterated expectations, the fourth equality holds by unconfoundedness, and the last equality holds because $Y = Y(1)$ among the treated group and $Y = Y(0)$ among the untreated group. This shows that ATE is identified.

b) In class, we showed that $ATT = \mathbb{E}[Y|D=1] - \mathbb{E}[\mathbb{E}[Y|X, D=0]|D=1]$. These are notably different. The expression for ATE takes the $\mathbb{E}[Y|X, D=1]$ (the mean of Y conditional on X among the treated group) and averages it over the distribution of X for the entire population and then subtracts $\mathbb{E}[Y|X, D=0]$ (the mean of Y conditional on X among the untreated group) averaged over the population distribution of X.

   In contrast, ATT compares the mean of Y among the treated group to $\mathbb{E}[Y|X, D=0]$ where this is averaged over the distribution of X *among the treated group*.

   An intuition for why ATE involves averaging over the distribution of X for the entire population is that ATE is the average treatment effect for the entire population.

c) We have that

$$\begin{aligned} Y_i &= Y_i(0) + D_i(Y_i(1) - Y_i(0)) \\ &= X_i'\beta + e_i + \alpha D_i \\ &= \alpha D_i + X_i'\beta + e_i \end{aligned}$$

   where the first equality holds by writing observed outcomes in terms of potential outcomes, the second equality uses the model for untreated potential outcomes and treatment effect homogeneity, and the last equality rearranges terms.

   Furthermore, unconfoundedness implies that $\mathbb{E}[e|X, D] = 0$ which implies that α can be estimated from the regression of Y on D and X.

d) This is exactly the same regressions as we talked about in class after we had identified the ATT. This should not be surprising because, if we restrict the effect of participating in the treatment to be the same across all units, then $ATT = ATE = \alpha$.

Extra Question 2

```r
# load data
data(Star, package="Ecdat")

# limit data to boys in small or regular class
data <- subset(Star,
               classk %in% c("small.class", "regular") & sex=="boy")

# part (a)
att_a <- mean(subset(data, classk == "small.class")$tmathssk) -
  mean(subset(data, classk=="regular")$tmathssk)
att_a
```

```
[1] 13.67522
```

```r
# part (b)
data <- droplevels(data)               # drop extra factors
X <- model.matrix(~classk, data=data) # get data matrix
```

```r
Y <- as.matrix(data$tmathssk)          # get outcome
bet <- solve(t(X)%*%X)%*%t(X)%*%Y      # estimate beta
att_b <- bet[2]                        # report coefficient on small class
att_b
```

```
[1] 13.67522
```

```r
# part (c)
X <- model.matrix(~classk + totexpk + freelunk, data=data) # X w/ extra vars
bet <- solve(t(X)%*%X)%*%t(X)%*%Y                          # estimate beta
att_c <- bet[2]                                            # report coef. on small
att_c
```

```
[1] 13.42333
```

The results from parts (a) and (b) are exactly identical. The result from part (c) is similar, but not exactly the same — this is exactly what we would expect.