Homework 2 Solutions

Hansen 2.5

a) The mean squared error is given by

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

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 to not be too accurate for that value of X.

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

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

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 $h(X) = \mathrm{E}[e^2|X] = \sigma^2(X)$ which implies that MSE is minimized by $\sigma^2(X)$.

Hansen 2.6

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

where the first equality holds by the law of total variance (Theorem 2.8 in the textbook), the second equality holds by substituting for Y, the third equality holds because (i) conditional on X the variance of $m(X)$ equals 0, and (ii) $\mathrm{E}[m(X) + e|X] = m(X)$, the fourth equality holds because $\mathrm{var}(e|X) = \mathrm{E}[e^2|X] - \mathrm{E}[e|X]^2$ (by the definition of conditional variance), the fifth equality holds by the law of iterated expectations, and the last equality holds by the definition of σ^2.

Hansen 2.10

True.

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

Hansen 2.11

False. Here is a counterexample. Suppose that $X = 1$ with probability $1/2$ and that $X = -1$ with probability $1/2$. Importantly, this means that $X^2 = 1$, $X^3 = X$, $X^4 = 1$, and so on; this further implies that $\mathrm{E}[X] = 0$, $\mathrm{E}[X^2] = 1$, $\mathrm{E}[X^3] = 0$ and so on. Also, suppose that $\mathrm{E}[e|X] = X^2$. Then, $\mathrm{E}[Xe] = \mathrm{E}[X\mathrm{E}[e|X]] = \mathrm{E}[X \cdot X^2] = \mathrm{E}[X^3] = 0$. However, $\mathrm{E}[X^2 e] = \mathrm{E}[X^2 \mathrm{E}[e|X]] = \mathrm{E}[X^2 \cdot X^2] = \mathrm{E}[X^4] = 1 \neq 0$

Hansen 2.12

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

Hansen 2.13

False. The same counterexample as in 2.11 works here. In that case, $\mathrm{E}[Xe] = 0$, but $\mathrm{E}[e|X] = X^2$ (in that case $X^2 = 1$, but the main point is that it is not equal to 0 for all values of X).

Hansen 2.14

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

Extra Question

```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
```

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

```
# 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.