# Homework 4 Solutions

## H 2.3

$$\mathrm{E}[h(X)e] = \mathrm{E}\Big[h(X)\underbrace{\mathrm{E}[e|X]}_{=0}\Big] = 0$$

where the first equality uses the law of iterated expectations.

## H 2.7

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

where the first equality holds from the definition of $\sigma^2(X)$, the second equality uses the definition of $e$, the third equality is just algebra, the fourth equality holds from the conditioning theorem (i.e., functions of $X$ can come out of expectations conditional on $X$), and the last equality combines terms.

## H 2.10

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

## H 2.11

False. Here is a counterexample. Suppose that $X \sim \mathcal{N}(0,1)$ (in this case, $\mathrm{E}[X^3] = 0$ and $\mathrm{E}[X^4] = 3$) and that $e|X \sim \mathcal{N}(0, X^2)$. Then, $\mathrm{E}[Xe] = \mathrm{E}\Big[X\mathrm{E}[e^2|X]\Big] = \mathrm{E}[X^3] = 0$ (as in the problem), but $\mathrm{E}[X^2 e] = \mathrm{E}\Big[X^2\mathrm{E}[e^2|X]\Big] = \mathrm{E}[X^4] = 3 \neq 0$.

## H 7.9

Notice that $\hat{\beta}$ corresponds to the usual least squares estimator of $\beta$ but for the particular case where the model does not include an intercept and where there is a single regressor.

a)

$$
\begin{aligned}
\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i (X_i\beta + e_i) \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i e_i \qquad \text{(A)} \\
&\xrightarrow{p} \beta + \mathrm{E}[X^2]^{-1}\mathrm{E}[Xe]
\end{aligned}
$$

where the second equality plugs in for $Y_i$, the third equality cancels terms, and the fourth equality holds by the weak law of large numbers and the continuous mapping theorem. Moreover, $\mathrm{E}[Xe] = \mathrm{E}[X\mathrm{E}[e|X]] = 0$. Thus, $\hat{\beta} \xrightarrow{p} \beta$, so that $\hat{\beta}$ is consistent for $\beta$. Next,

$$
\begin{aligned}
\tilde{\beta} &= \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{X_i} \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{X_i\beta + e_i}{X_i} \\
&= \beta + \frac{1}{n}\sum_{i=1}^{n} \frac{e_i}{X_i} \qquad \text{(B)} \\
&\xrightarrow{p} \beta + \mathrm{E}\left[\frac{e}{X}\right]
\end{aligned}
$$

where the second equality plugs in for $Y_i$, the second equality cancels terms and rearranges, and the last equality holds by the weak law of large numbers. Moreover, $\mathrm{E}[X^{-1}e] = \mathrm{E}[X^{-1}\mathrm{E}[e|X]] = 0$. Thus, $\tilde{\beta}$ is also consistent for $\beta$.

b)

To talk about efficiency, we need to derive the asymptotic variance of $\hat{\beta}$ and $\tilde{\beta}$. Re-arranging Equation (A), we have that

$$
\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e_i \\
&\xrightarrow{d} \mathcal{N}(0, V_1)
\end{aligned}
$$

where $V_1 = \mathrm{E}[X^2]^{-2}\mathrm{E}[X^2e^2]$. This expression should seem familiar. It is a simplification of the usual asymptotic variance that we derivied in class, but for the special case considered in this problem. Similarly, we can re-arrange Equation (B) to get

$$
\begin{aligned}
\sqrt{n}(\tilde{\beta} - \beta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{e_i}{X_i} \\
&\xrightarrow{d} \mathcal{N}(0, V_2)
\end{aligned}
$$

where $V_2 = \mathrm{E}[X^{-2}e^2]$. It's not immediately clear if $V_1$ or $V_2$ is smaller. One natural starting point for thinking about this is what happens under homoskedasticity (i.e., $\mathrm{E}[e^2|X] = \sigma^2$ / does not depend on $X$). In this case,

$$V_1 = \sigma^2 \mathrm{E}[X^2]^{-1} \quad \text{and} \quad V_2 = \sigma^2 \mathrm{E}[X^{-2}]$$

Next, notice that $X^2 > 0$ (as long as we rule out the case where $X = 0$). The function, $g(z) = z^{-1}$ is convex for $z > 0$; therefore, from Jensen's inequality, we have that $g(\mathrm{E}[Z]) \leq \mathrm{E}[g(Z)] \implies \mathrm{E}[X^2]^{-1} \leq \mathrm{E}[X^{-2}] \implies V_1 \leq V_2$. Thus, this is a case where $\hat{\beta}$ is more efficient than $\tilde{\beta}$.

That said, it is also possible to come up with a case where $\tilde{\beta}$ is more efficient that $\hat{\beta}$. Suppose that $\mathrm{E}[e^2|X] = X^2$ (in this case, there is heteroskedasticity and the variance of $e$ increases for large in magnitude values of $X$). In this case,

$$V_1 = \frac{\mathrm{E}[X^4]}{(\mathrm{E}[X^2])^2} \quad \text{and} \quad V_2 = 1$$

Taking $Z = X^2$, and considering the function $g(z) = z^2$ (which is convex), from Jensen's inequality we have that $g(\mathrm{E}[Z]) \leq \mathrm{E}[g(Z)] \implies (\mathrm{E}[X^2])^2 \leq \mathrm{E}[(X^2)^2] = \mathrm{E}[X^4]$. This implies that $V_1 \geq 1 = V_2$ which implies that $\tilde{\beta}$ is more efficient than $\hat{\beta}$ in this case.

## H 7.27

a)

$$\begin{aligned}
\tilde{\beta} &= \left(\frac{1}{n}\sum_{i=1}^n X_i X_i' \mathbb{1}\{|X_i| \leq c\}\right)^{-1} \frac{1}{n}\sum_{i=1}^n X_i Y_i \mathbb{1}\{|X_i| \leq c\} \\
&= \left(\frac{1}{n}\sum_{i=1}^n X_i X_i' \mathbb{1}\{|X_i| \leq c\}\right)^{-1} \frac{1}{n}\sum_{i=1}^n X_i (X_i'\beta + e_i) \mathbb{1}\{|X_i| \leq c\} \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^n X_i X_i' \mathbb{1}\{|X_i| \leq c\}\right)^{-1} \frac{1}{n}\sum_{i=1}^n X_i e_i \mathbb{1}\{|X_i| \leq c\} \qquad \text{(C)} \\
&\xrightarrow{p} \beta + \mathrm{E}[XX'\mathbb{1}\{|X| \leq c\}]^{-1}\mathrm{E}[Xe\mathbb{1}\{|X| \leq c\}]
\end{aligned}$$

where the second equality plugs in for $Y_i$, the third equality cancels terms, and the last line holds by the weak law of large numbers and the continuous mapping theorem. Moreover, notice that

$$\mathrm{E}[Xe\mathbb{1}\{|X| \leq c\}] = \mathrm{E}\Big[X\mathbb{1}\{|X| \leq c\} \underbrace{\mathrm{E}[e|X]}_{=0}\Big]$$

which uses (as in the problem) that the CEF is linear so that $\mathrm{E}[e|X] = 0$. Thus, $\tilde{\beta} \xrightarrow{p} \beta$.

b) Starting by re-arranging Equation C, we have that

$$\begin{aligned}
\sqrt{n}(\tilde{\beta} - \beta) &= \left(\frac{1}{n}\sum_{i=1}^n X_i X_i' \mathbb{1}\{|X_i| \leq c\}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i e_i \mathbb{1}\{|X_i| \leq c\} \\
&\xrightarrow{d} \mathcal{N}(0, \tilde{V})
\end{aligned}$$

where $\tilde{V} = \mathrm{E}[XX'\mathbb{1}\{|X| \leq c\}]^{-1}\tilde{\Omega}\mathrm{E}[XX'\mathbb{1}\{|X| \leq c\}]^{-1}$ and where $\tilde{\Omega} = \mathrm{E}[XX'e^2\mathbb{1}\{|X| \leq c\}]$ which holds by (i) for $\tilde{\Omega}$, applying the central limit theorem to $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e_i \mathbb{1}\{|X_i| \leq c\}$ (notice that this has mean 0 and that, if you square $\mathbb{1}\{|X| \leq c\}$, it is just equal to itself), and (ii) the extended continuous mapping theorem (to combine everything together).