Exercise 1

$y_i = x_i'\beta + u_i$.

Part (a)

The objective function for the OLS estimator is

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2.$$

We want to find the β that minimizes $Q_n(\theta)$. Take the derivative with respect to β and set it equal to zero,

$$-2\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})x_i = 0 \quad\Longrightarrow\quad \hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n}x_ix_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}y_ix_i.$$

This, of course, requires the inverse **existence assumption**.

Part (b)

First Assumption: Identification. The true model is $y_i = x_i'\beta_0 + u_i$ and β_0 is unique in this respect. Therefore we can write the estimator as

$$\begin{aligned}
\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n}x_ix_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}(x_i'\beta_0 + u_i)x_i \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}x_ix_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}x_ix_i'\beta_0 + \left(\frac{1}{n}\sum_{i=1}^{n}x_ix_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}x_iu_i \\
&= \beta_0 + \left(\frac{1}{n}\sum_{i=1}^{n}x_ix_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}x_iu_i.
\end{aligned}$$

Second Assumption. Let's assume that data is i.i.d. **Third Assumption.** Let's assume that $E[x_ix_i']$ is finite. These two assumptions allow us to invoke Khinchine's LLN to say that

$$\frac{1}{n}\sum_{i=1}^{n}x_ix_i' \xrightarrow{p} E[x_ix_i'].$$

Fourth Assumption. Now let's assume that $E[x_ix_i']$ is positive semidefinite, which implies that $E[x_ix_i']^{-1}$ exists. We can then use the Continuous Mapping Theorem to imply that

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} E[x_i x_i']^{-1} = A_0^{-1}.$$

Fifth Assumption. Assume that $E[x_i u_i] = 0$. This combined with i.i.d. implies

$$\frac{1}{n}\sum_{i=1}^{n} x_i u_i \xrightarrow{p} 0.$$

Therefore by Slutsky's theorem, the product

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i u_i \xrightarrow{p} 0.$$

It follows that $\hat{\beta} \xrightarrow{p} \beta_0$.

Now rewrite

$$\sqrt{n}(\hat{\beta} - \beta_0) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} x_i u_i.$$

Sixth Assumption. Let's assume that $\mathrm{Var}(x_i u_i) < \infty$. This, along with i.i.d., lets us invoke the Lindeberg-Levy CLT for

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} x_i u_i \xrightarrow{d} \mathcal{N}\left(0, E[u_i^2 x_i x']\right).$$

It follows that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}\left(0, E[x_i x_i']^{-1} E[u_i^2 x_i x'] E[x_i x_i']^{-1}\right).$$

Part (c)

For extremum estimator consistency, we need

- **Existence.** $\hat{\beta}$ needs to be a thing. In this case,

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} y_i x_i$$

  needs to exist, which depends entirely upon $n^{-1}\sum_{i=1}^{n} x_i x_i'$ being invertible.

- **Identification.** β_0 is the unique solution to the objective function. We first assume that β_0 is indeed *a* minimizer. Now we need to show that it's *the* minimizer, i.e.

$$E[(y_i - x_i'\beta)^2] - E[(y_i - x_i'\beta_0)^2] > 0$$

for all $\beta \neq \beta_0$. So let's show it.

$$\begin{aligned}
E[(y_i - x_i'\beta)^2] - E[(y_i - x_i'\beta_0)^2] &= E\big[([x_i'\beta_0 + u_i] - x_i'\beta)^2\big] - E[(y_i - x_i'\beta_0)^2] \\
&= E\big[(x_i'[\beta_0 - \beta] + u_i)^2\big] - E[u_i^2] \\
&= E\big[(x_i'[\beta_0 - \beta])^2\big] + 2E[x_i'u_i](\beta_0 - \beta) + E[u_i^2] - E[u_i^2] \\
&= E\big[(x_i'[\beta_0 - \beta])^2\big] + 2E[x_i'u_i](\beta_0 - \beta).
\end{aligned}$$

Now let's use the assumption that $E[x_i'u_i] = 0$. Then we just need to make sure that $\beta_0 \neq \beta$ so that the square in the remaining expectation isn't zero. This gives

$$E\big[(x_i'[\beta_0 - \beta])^2\big] > 0 \quad \Longrightarrow \quad E[(y_i - x_i'\beta)^2] > E[(y_i - x_i'\beta_0)^2].$$

So indeed, β_0 is the unique minimizer.

- **ULLN.** Yeah, we need the uniform law of large numbers, i.e.

$$\sup_{\beta \in B} |Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0.$$

So let's plug and chug.

$$\begin{aligned}
&\sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i'\beta)^2 - E[(y_i - x_i'\beta)^2] \right| \\
&= \sup_{\beta \in B} \left| \frac{1}{n} \sum_{i=1}^{n} y_i^2 - E[y_i^2] - 2\beta \frac{1}{n} \sum_{i=1}^{n} y_i x_i' + 2\beta E[y_i x_i] + \beta^2 \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \beta^2 E[x_i^2] \right| \\
&\leq \left| \frac{1}{n} \sum_{i=1}^{n} y_i^2 - E[y_i^2] \right| + \sup_{\beta \in B} \left| 2\beta \frac{1}{n} \sum_{i=1}^{n} y_i x_i' - 2\beta E[y_i x_i] \right| + \sup_{\beta \in B} \left| \beta^2 \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \beta^2 E[x_i^2] \right| \\
&\leq \left| \frac{1}{n} \sum_{i=1}^{n} y_i^2 - E[y_i^2] \right| + 2 \sup_{\beta \in B} |\beta| \left| \frac{1}{n} \sum_{i=1}^{n} y_i x_i' - \beta E[y_i x_i] \right| + \sup_{\beta \in B} |\beta^2| \left| \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \beta^2 E[x_i^2] \right|.
\end{aligned}$$

Assume that $\sup_{\beta \in B} |\beta| < \infty$, $E[y_i^2] < \infty$, $E[y_i x_i] < \infty$, and $E[x_i^2] < \infty$, where each expectant is also i.i.d. Then we can apply Khinchine's LLN so that each term goes to zero in probability. The result is then established.

Part (d)

Take the derivative of the objective function in part (a) with respect to β, which is

$$\sum_{i=1}^{n} -2\frac{1}{n}x_i(y_i - x_i'\beta).$$

The summand is actually the score, $s(y_i, x_i; \beta)$. Notice that evaluated at $\hat{\beta}$, the sum of the score equals zero as the first order condition.

Take the second derivative of the objective function to get

$$\sum_{i=1}^{n} 2\frac{1}{n}x_i x_i'.$$

The summand is the Hessian, $H(y_i, x_i; \beta)$, which is the derivative of the score.

For C^2 function f, the mean value expansion says that $f(b) = f(a) + f'(c)(b-a)$ for some $c \in (b, a)$. Have the following:

$$f(\theta) = \frac{\partial Q_n(\theta)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} s(y_i, x_i; \theta) = \sum_{i=1}^{n} -2\frac{1}{n}x_i(y_i - x_i'\beta),$$

$$f'(\theta) = \frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'} = \frac{1}{n}\sum_{i=1}^{n} H(y_i, x_i; \theta) = \sum_{i=1}^{n} -2\frac{1}{n}x_i x_i'.$$

Then the mean value expansion gives

$$\frac{1}{n}\sum_{i=1}^{n} s(y_i, x_i; \hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} s(y_i, x_i; \beta_0) + \frac{1}{n}\sum_{i=1}^{n} H(y_i, x_i; \beta^*)(\hat{\beta} - \beta_0)$$

$$\implies \quad -2\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}) = -2\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - x_i'\beta_0) + 2\frac{1}{n}\sum_{i=1}^{n} x_i x_i'(\hat{\beta} - \beta_0).$$

The LHS is zero from first order conditions. Therefore we have

$$\hat{\beta} - \beta_0 = \left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i(y_i - x_i'\beta_0)$$

$$\implies \quad \hat{\beta} - \beta_0 = \left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i u_i.$$

Oh shit it's the same thing. Woo.

Part (e)

Finding the sampling error is exactly the same as before, relying upon the same conditions.

Part (f)

If we do this problem in the M-estimation framework, just have $m(y_i, x_i; \beta) = (y_i - x_i'\beta)^2$ and everything follows as before, same assumptions required.

Part (g)

We're using the model $y_i = x_i'\beta + u_i$. We are told that $E[x_i] = 0$ and $E[y_i] = 0$. We are given the hint that

$$\operatorname*{plim}_{n\to\infty} \hat{\beta} = \beta_0 + \operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^n x_i u_i}{\frac{1}{n}\sum_{i=1}^n x_i^2}.$$

Omitted Variable Bias. Suppose the true model is actually $y_i = \gamma_1 x_i + \gamma_2 z_i + v_i$, where $E[v_i|x_i, z_i] = 0$. Then we will have

$$\operatorname*{plim}_{n\to\infty} \hat{\beta} = \gamma_1 + \operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^n x_i u_i}{\frac{1}{n}\sum_{i=1}^n x_i^2}.$$

Since $E[x_i] = 0$, it follows that $\text{Cov}(x_i, u_i) = E[x_i u_i]$ and $\text{Var}(x_i) = E[x_i^2]$. We want to apply Khinchine's LLN so that these sample averages converge, so let's assume i.i.d. and that $E[x_i u_i] < \infty$ and $E[x_i^2] < \infty$. Then we can apply Slutsky's theorem where

$$\operatorname*{plim}_{n\to\infty} \hat{\beta} = \gamma_1 + \operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^n x_i u_i}{\frac{1}{n}\sum_{i=1}^n x_i^2} = \gamma_1 + \frac{\text{Cov}(x_i, u_i)}{\text{Var}(x_i)}.$$

The true model tells us that $u_i = \gamma_2 z_i + v_i$. Plug this bad boy into the equation above and we have

$$\operatorname*{plim}_{n\to\infty} \hat{\beta} = \gamma_1 + \frac{\text{Cov}(x_i, \gamma_2 z_i + v_i)}{\text{Var}(x_i)} = \gamma_1 + \gamma_2 \frac{\text{Cov}(x_i, z_i)}{\text{Var}(x_i)} + \frac{\text{Cov}(x_i, v_i)}{\text{Var}(x_i)}.$$

The right-most term is zero because $E[x_i] = 0$ and $E[v_i|x_i] = 0$. The middle term is the omitted variable bias. Then $\hat{\beta}$ only converges to the true value if either $\gamma_2 = 0$

or $\text{Cov}(x_i z_i) = 0$. Whether the estimator overshoots or undershoots the true value depends on the signs of $\gamma_2 = 0$ and $\text{Cov}(x_i z_i) = 0$.

Exercise II

Part (a). The function is $f_n(x) = x^n$ where $x \in [0, 1]$. Note that

$$f(x) = \lim_{n\to\infty} f_n(x) = \begin{cases} 0 & \text{if } x \in [0,1) \\ 1 & \text{if } x = 1. \end{cases}$$

We want to show whether or not

$$\lim_{n\to\infty} \left\{ \sup_{x\in[0,1]} |f_n(x) - f(x)| \right\} \to 0.$$

Fix n. The first thing to note is that if $x = 1$, then $f_n(x) = 1$ and $f(x) = 1$, so the absolute value just evaluates to zero.

Now consider $x \in [0, 1)$. The term in the absolute value is $x^n - 0$. So what is the supremum of x^n? Well, we can make x as close to 1 as we want, but without quite hitting 1. The closer to 1 we make x, the closer x^n gets to 1. Arbitrarily close, in fact. So even though it never reaches 1, the least upper bound will be 1. Therefore, for any n, we have

$$\sup_{x\in[0,1)} |1 - 0| = 1.$$

And thus the limit will also be 1. There is no uniform convergence.

Part (b). The function is $f_n(x) = n/(nx+1)$. Divide numerator and denominator by n to see that this limits to $f(x) = 1/x$. Fix n and analyze

$$\sup_{x\in[0,1]} \left| \frac{n}{nx+1} - \frac{1}{x} \right| = \sup_{x\in[0,1]} \left| \frac{1}{x(nx+1)} \right|.$$

Clearly this will be maximized by having $x = 0$, which will blow it up to infinity. So the limit as $n \to$ of infinity sure as hell ain't zero. Not uniformly convergent.

Exercise III

The model is $y_i = g(x_i; \theta_0) + u_i$.

Part (a)

The objective functions are

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\big[y_i - g(x_i;\theta)\big]^2,$$

$$Q(\theta) = E\left[\left(y_i - g(x_i;\theta)\right)^2\right].$$

Part (b)

Suppose that $g(x_i;\theta) = \theta_1 x_i + \theta_2 x_i^2$. We can write $z_i = [x_i, x_i^2]$. The objective function becomes

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\big[y_i - z_i'\theta\big]^2.$$

Therefore the first order condition is just

$$-2\frac{1}{n}\sum_{i=1}^{n} z_i[y_i - z_i'\hat{\theta}] := 0 \quad\implies\quad \frac{1}{n}\sum_{i=1}^{n} z_i[y_i - z_i'\hat{\theta}] := 0.$$

Move some stuff around and you get

$$\hat{\theta} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i y_i$$

Part (c)

Now suppose the function is $g(x_i;\theta) = \theta^{x_i}$. Then the objective function be

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\big[y_i - \theta^{x_i}\big]^2.$$

The first order condition is

$$-2\frac{1}{n}\sum_{i=1}^{n} x_i\hat{\theta}^{x_i-1}[y_i - \hat{\theta}^{x_i}] := 0 \quad\implies\quad \frac{1}{n}\sum_{i=1}^{n} x_i\hat{\theta}^{x_i-1}[y_i - \hat{\theta}^{x_i}] := 0.$$

Yeah, good luck solving that thing for $\hat{\theta}$.

Part (d)

Okay. For consistency, we need:

(a) **Existence.** We don't need to verify this one.

(b) **Identification.** We want to show that $E\big[\big(y_i - g(x_i;\theta)^2\big)\big] > E\big[\big(y_i - g(x_i;\theta_0)^2\big)\big]$ for any $\theta \neq \theta_0$, **assuming** that θ_0 is a parameter that gives the true model.

$$\begin{aligned}
&E\big[\big(y_i - g(x_i;\theta)\big)^2\big] - E\big[\big(y_i - g(x_i;\theta_0)\big)^2\big] \\
&= E\big[\big(g(x_i;\theta_0) + u_i - g(x_i;\theta)\big)^2\big] - E\big[\big(y_i - g(x_i;\theta_0)\big)^2\big] \\
&= E\left[\big(g(x_i;\theta_0) - g(x_i;\theta)\big)^2\right] + 2E\big[u_i\big(g(x_i;\theta_0) - g(x_i;\theta)\big)\big] + E[u_i^2] - E\big[\big(y_i - g(x_i;\theta_0)\big)^2\big] \\
&= E\left[\big(g(x_i;\theta_0) - g(x_i;\theta)\big)^2\right] + 2E\big[u_i\big(g(x_i;\theta_0) - g(x_i;\theta)\big)\big] + E[u_i^2] - E[u_i^2]] \\
&= E\left[\big(g(x_i;\theta_0) - g(x_i;\theta)\big)^2\right] + 2E\big[u_i\big(g(x_i;\theta_0) - g(x_i;\theta)\big)\big].
\end{aligned}$$

We will **assume** that $E[u_i|x_i] = 0$. The LIE then implies that for any function $f(x)$, we have

$$E[u_i f(x_i)] = E\big[E[u_i f(x_i)|x_i]\big] = E\big[f_i(x)E[u_i|x_i] = 0.\big]$$

Since $g(x_i;\theta_0) - g(x_i;\theta)$ is a function of x_i, it follows that

$$E\big[u_i\big(g(x_i;\theta_0) - g(x_i;\theta)\big)\big] = 0.$$

Finally, let's **assume** that $g(x_i;\theta_0) \neq g(x_i;\theta)$ for any $\theta \neq \theta_0$. Then we have shown that

$$E\big[\big(y_i - g(x_i;\theta)\big)^2\big] - E\big[\big(y_i - g(x_i;\theta_0)\big)^2\big]$$

.

(c) **ULLN.** This has its own set of conditions.

(i) **Compactness**. We want Θ to be compact. This need not be verified.

(ii) **Continuity.** We want $g(x_i;\theta)$ to be continuous in θ for all x_i. This will depend on whatever $g(x_i;\theta)$ happens to be.

(iii) **Measurability.** We want $g(x_i;\theta)$ to be measurable in x_i for all θ. This need not be verified.

(iv) **Dominance.** We want to find some function $d(x_i)$ with finite expectation such that

$$\left|\big(y - g(x_i;\theta)\big)^2\right| \leq d(x_i)$$

for all θ. Since it's squared, we can drop the absolute value operator. Now

assume that $g(x_i;\theta) \leq h(x_i)$. Then

$$\begin{aligned}
\big(y_i - g(x_i;\theta)\big)^2 &\leq \big(\,|y_i| + |g(x_i;\theta)|\,\big)^2 \\
&\leq \big(\,|y_i| + |h(x_i)|\,\big)^2 \\
&= y_i^2 - 2\,|y_i h(x_i)| + h(x_i)^2 \\
&:= d(x_i).
\end{aligned}$$

If we add further **assumptions** that $E[y_i^2]$, $E[|y_i h(x_i)|]$ and $E[|h(x_i)|]$ are finite, then $d(x_i)$ is a dominating function.

Part (e).

In part (b), we needed to assume $E[u_i|x_i] = 0$ because it allowed us to show that $E[u_i f(x_i)|x_i] = 0$. Assuming only that $E[u_i x_i] = 0$ was not enough. $E[u_i|x_i] = 0$ implies $E[u_i x_i] = 0$, among other things, so $E[u_i|x_i]$ is the stronger assumption.

Part (f).

We have $m(y_i, x_i;\theta) = \big(y_i - g(x_i;\theta)\big)^2$. It follows that

$$\begin{aligned}
s(y_i, x_i;\theta) &= -2\frac{\partial g(x_i;\theta)}{\partial\theta}\big(y_i - g(x_i;\theta)\big), \\
H(y_i, x_i;\theta) &= -2\frac{\partial^2 g(x_i;\theta)}{\partial\theta\partial\theta'}\big(y_i - g(x_i;\theta)\big) + 2\frac{\partial g(x_i;\theta)}{\partial\theta}\frac{\partial g(x_i;\theta)}{\partial\theta'}.
\end{aligned}$$

So the mean value expansion is going to be

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n s(y_i, x_i;\hat\theta) &= \frac{1}{n}\sum_{i=1}^n s(y_i, x_i;\theta_0) + \frac{1}{n}\sum_{i=1}^n H(y_i, x_i;\theta^*)(\hat\theta - \theta_0) \\
\implies \quad \sqrt{n}(\hat\theta - \theta_0) &= -\left[\frac{1}{n}\sum_{i=1}^n H(y_i, x_i;\theta^*)\right]^{-1}\sqrt{n}\frac{1}{n}\sum_{i=1}^n s(y_i, x_i;\theta_0).
\end{aligned}$$

Now plug in the actual expressions for the score and Hessian.

$$\left[\frac{2}{n}\sum_{i=1}^n \frac{\partial g(x_i;\theta^*)}{\partial\theta}\frac{\partial g(x_i;\theta^*)}{\partial\theta'} - \frac{\partial^2 g(x_i;\theta^*)}{\partial\theta\partial\theta'}\big(y_i - g(x_i;\theta^*)\big)\right]^{-1}\sqrt{n}\frac{2}{n}\sum_{i=1}^n \frac{\partial g(x_i;\theta)}{\partial\theta}u_i.$$

For asymptotic normality, we're going to need the following.

- **Consistency.** Been there done that.
- **Mean Value Expansion Validity.** Which means we need $g(x_i; \theta)$ to be twice continuously differentiable in θ and in the interior of Θ; and θ_0 needs to be in the interior of Θ.
- **Consistency of the Sample Average of the Hessian.** Most importantly, we want the expectation of the Hessian to be positive definite so that it is invertible. And some other stuff I'm going to ignore.
- **Asymptotic Normality of the Sample Average of the Score.** We'll need each element of
$$\frac{\partial g(x_i; \theta)}{\partial \theta} u_i$$
to have finite second moment. Furthermore, we'll need
$$E\left[\frac{\partial g(x_i; \theta)}{\partial \theta} u_i\right] = 0.$$