Problem Set #6

Danny Edgel
Econ 709: Economic Statistics and Econometrics I
Fall 2020

November 7, 2020

Collaborated with Sarah Bass, Emily Case, Michael Nattinger, and Alex Von Hafften

Question 1

Find $\mathbb{E}[\mathbb{E}[\mathbb{E}[\mathbf{Y}|\mathbf{X_1},\mathbf{X_2},\mathbf{X_3}]|\mathbf{X_1},\mathbf{X_2}]|\mathbf{X_1}]$

By the Law of Iterated Expectation,

$$\begin{aligned}\mathbb{E}[\mathbb{E}[\mathbb{E}[Y|X_1,X_2,X_3]|X_1,X_2]|X_1] &= \mathbb{E}[\mathbb{E}[Y|X_1,X_2]|X_1]\\ \mathbb{E}[\mathbb{E}[Y|X_1,X_2]|X_1] &= \mathbb{E}[Y|X_1]\end{aligned}$$

Thus, $\mathbb{E}[\mathbb{E}[\mathbb{E}[Y|X_1,X_2,X_3]|X_1,X_2]|X_1] = \mathbb{E}[Y|X_1]$

Question 2

Prove that for any function $h(x)$ such that $\mathbb{E}\,|h(X)e| < \infty$ then $\mathbb{E}[h(X)e] = 0$, where $e = Y - m(X)$ and $m(X) = \mathbb{E}[Y|X]$

According to the conditioning theorem, if $\mathbb{E}\,|Y| < \infty$, then

$$\mathbb{E}[g(X)Y|X] = g(X)\,\mathbb{E}[Y|X]$$

Thus, Since $\mathbb{E}\,|h(X)e| < \infty$ trivially implies $\mathbb{E}\,|Y| < \infty$, we can use the Law of Iterated Expectation to solve:

$$\begin{aligned}\mathbb{E}[h(X)e] &= \mathbb{E}[h(X)Y - h(X)m(X)] = \mathbb{E}[h(X)Y] - \mathbb{E}[h(X)m(X)]\\ &= \mathbb{E}[\mathbb{E}[h(X)Y|X]] - \mathbb{E}[h(X)m(X)] = \mathbb{E}[h(X)\,\mathbb{E}[Y|X]] - \mathbb{E}[h(X)m(X)]\\ &= \mathbb{E}[h(X)m(X)] - \mathbb{E}[h(X)m(X)] = 0\end{aligned}$$

$\therefore$ for any function $h(x)$ such that $\mathbb{E}\,|h(X)e| < \infty$ then $\mathbb{E}[h(X)e] = 0$ ■

Question 3

$$\mathbb{E}[Y|X] = \begin{cases} .4, & X = 0 \\ .3, & X = 1 \end{cases}$$

$$\mathbb{E}[Y^2|X] = \begin{cases} .4, & X = 0 \\ .3, & X = 1 \end{cases}$$

$$Var(Y|X) = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2 = \begin{cases} .24, & X = 0 \\ .21, & X = 1 \end{cases}$$

Question 4

Show that $\sigma^2(X)$ minimizes the mean-squared error and is thus the best predictor.

The variance of $\hat{\beta}_{OLS} = \mathbb{E}(Y|X)$ is

$$\sigma^2(X) = \mathbb{E}\left[(Y - h(X))^2\right] = \sigma^2(X'X)^{-1}$$

Where $\sigma^2 = \mathbb{E}(\varepsilon|X)$ and $h(X)$ is the predictor of Y, which, in this case, is $E(Y|X)$. It is clear that minimizing the variance of $\hat{\beta}$ will also minimize mean-squared error. Thus, we can show that this minimizes mean-squared error among all linear unbiased estimators by comparing this variance to the variance of an arbitrary linear estimator, $\tilde{\beta} = a + Ay$. In order for $\tilde{\beta}$ to be unbiased, it must be the case that $\mathbb{E}(\tilde{\beta}) = \mathbb{E}(\tilde{\beta}|X) = \beta$. Thus,

$$\beta = \mathbb{E}(\tilde{\beta}|X) = \mathbb{E}(a + Ay|X) = a + A\,\mathbb{E}(y|x) = a + A\beta$$

This only holds if $a = 0$, so $\tilde{\beta} = Ay$. Then, the variance of $\tilde{\beta}$ is:

$$\begin{aligned}
\mathbb{E}(\tilde{\beta}|X) &= V(Ay|X) = AV(y|X)A' = \sigma^2 AA' \\
&= \sigma^2 \left[A - (X'X)^{-1} + (X'X)^{-1}X'\right]\left[A' - (X'X)^{-1} + (X'X)^{-1}X'\right] \\
&= \text{... skipping intermediate steps for brevity} \\
&= \sigma^2[A - (X'X)^{-1}X'][A - (X'X)^{-1}X']' + \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X) + \sigma^2[A - (X'X)^{-1}X'][A - (X'X)^{-1}X']' \\
&\geq \sigma^2(X)
\end{aligned}$$

Thus, $\sigma^2(X)$ minimizes the mean-squared error over any other linear estimator.

Question 5

Compute $\mathbb{E}[Y|X]$ and $Var(Y|X)$ for a Poisson-distributed Y given X. Does this justify a linear regression model?

The conditional mean and variance are:

$$\mathbb{E}[Y|X] = x'\beta \text{ , } Var(Y|X) = x'\beta$$

A linear regression model is satisfied if the three Gauss-Markov assumptions are violated. Provided rank$X = k$, that is the case in this situation, because $Var(Y|X) = x'$ implies that the model has constant variance, and:

$$\mathbb{E}[Y|X] = x'\beta \iff \mathbb{E}(\varepsilon|X) = 0$$

Thus, a linear regression model is justified.

Question 6

2.10 True. By the law of iterated expectations,

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

2.11 False. Consider $X = \{-1, 1\}$ with constant error $e = \overline{e}$ and $\Pr(X = -1) = \Pr(X = 1) = \frac{1}{2}$. In this case, $\mathbb{E}[Xe] = 0$ and $\mathbb{E}[X^2 e] = \overline{e}$

2.12 True. $\mathbb{E}[e] = 0$ by definition, so e is independent of X if and only if $\mathbb{E}[Xe] = \mathbb{E}[e]\,\mathbb{E}[X] = 0$. By the law of iterated expectations,

$$\mathbb{E}[Xe] = \mathbb{E}[\mathbb{E}[Xe|X]] = \mathbb{E}[X\,\mathbb{E}[e|X]] = 0$$

2.13 False. Consider the same example from 2.11. $E[eX] = 0$, but $\mathbb{E}[e|X] = \overline{e}$.

2.14 True. Using the law of iterated examples explanation from 2.12, $\mathbb{E}[e|X] = 0 \Rightarrow \mathbb{E}[Xe] = 0$.

Question 7

Let X and Y have the joint density $f(x,y) = \frac{3}{2}(x^2+y^2)$ on $0 \le x \le 1,\ 0 \le y \le 1$. Compute the coefficients of the best linear predictor $Y = \alpha + \beta X + e$. Compute the conditional expectation $m(x) = \mathbb{E}[Y|X=x]$. Are the best linear predictor and conditional expectation different?

The conditional expectation, $m(x) = \mathbb{E}[Y|X=x]$, can be calculated as:

$$
\begin{aligned}
f(y|x) &= \frac{f(x,y)}{\int_0^1 f(x,y)dy} = \frac{\frac{3}{2}(x^2+y^2)}{\int_0^1 \frac{3}{2}(x^2+y^2)dy} = \frac{x^2+y^2}{x^2+\frac{1}{3}} \\
\mathbb{E}[Y|X=x] &= \int_0^1 f(y|x)ydy = \int_0^1 \frac{x^2+y^2}{x^2+\frac{1}{3}}ydy = \frac{1}{x^2+\frac{1}{3}}\left(x^2\int_0^1 ydy + \int_0^1 y^3dy\right) \\
&= \frac{1}{x^2+\frac{1}{3}}\left(\frac{1}{2}x^2+\frac{1}{4}\right) = \frac{2x^2+1}{4x^2+4/3} \\
&= \frac{6x^2+3}{12x^2+4}
\end{aligned}
$$

And the coefficients α and β can be calculated as:

$$
\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}(\mathbb{E}(Y)-\mathbb{E}(X)\beta \\ \frac{\mathbb{E}(XY)-\mathbb{E}(X)\,\mathbb{E}(Y)}{\mathbb{E}(X^2)-(\mathbb{E}(X))^2}\end{pmatrix} = \frac{1}{\mathbb{E}(X^2)-(\mathbb{E}(X))^2}\begin{pmatrix}\mathbb{E}(Y)\,\mathbb{E}(X^2)-\mathbb{E}(X)\,\mathbb{E}(XY) \\ \mathbb{E}(XY)-\mathbb{E}(X)\,\mathbb{E}(Y)\end{pmatrix}
$$

Where:

$$
\begin{aligned}
f(x) &= \int_0^1 f(x,y)dy = \int_0^1 \frac{3}{2}(x^2+y^2)dy = \frac{3}{2}x^2 + \int_0^1 \frac{3}{2}y^2dy = \frac{3}{2}x^2+\frac{1}{2} \\
\mathbb{E}(X) &= \int_0^1 f(x)xdx = \int_0^1 \frac{3}{2}x^3dx + \frac{1}{2}\int_0^1 xdx = \frac{5}{8} \\
\mathbb{E}(X^2) &= \int_0^1 f(x)x^2dx = \int_0^1 \frac{3}{2}x^4dx + \frac{1}{2}\int_0^1 x^2dx = \frac{7}{15} \\
\mathbb{E}(XY) &= \int_0^1\int_0^1 f(x,y)xydxdy = \int_0^1\int_0^1 \frac{3}{2}(x^2+y^2)xydxdy \\
&= \text{... (intermediate steps omitted for brevity)} \\
\mathbb{E}(XY) &= \frac{3}{8}
\end{aligned}
$$

Since X and Y have symmetric marginal distributions, $\mathbb{E}(X) = \mathbb{E}(Y)$. Thus,

$$
\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \frac{1}{\frac{7}{15}-\frac{25}{64}}\begin{pmatrix}\left(\frac{5}{8}\right)\left(\frac{7}{15}\right)-\left(\frac{5}{8}\right)\left(\frac{3}{8}\right) \\ \left(\frac{3}{8}\right)-\left(\frac{5}{8}\right)\left(\frac{5}{8}\right)\end{pmatrix} = \frac{1}{73}\begin{pmatrix}55\\15\end{pmatrix}
$$

Question 8

4.1 **For $k \in \mathbb{Z}$, set $\mu_k = \mathbb{E}[Y^k]$.**

(a) **Construct an estimator, $\hat{\mu}_k$, for μ_k**

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} Y_i^k$$

(b) **Show that $\hat{\mu}_k$ is unbiased**
As long as $\{Y_i\}_{i=1}^n$ are i.i.d., then,

$$\mathbb{E}[\hat{\mu}_k] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} Y_i^k] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i^k] = \frac{1}{n}\sum_{i=1}^{n}\mu_k = \mu_k$$

(c) **Calculate the variance of $\hat{\mu}_k$. What assumption is needed for it to be finite?**
By the weak law of large numbers, if $Var\left(\hat{\mu}_k\right) < \infty$,

$$\sqrt{n}\left(\hat{\mu}_k - \mu_k\right) \to_d \mathcal{N}\left(0, V\right)$$

Where $V = Var(\mu_k)$. Since $\sqrt{n}\left(\hat{\mu} - \mu\right) \to_d \mathcal{N}\left(0, \sigma^2\right)$ where σ^2 is the variance of Y, by the delta method, with $g(\theta) = \theta^k$,

$$V = g'(\mu)^2\sigma^2 = \mu^{2k-2}k^2\sigma^2$$

(d) **Propose an estimator for the variance of $\hat{\mu}_k$.**

$$\widehat{Var(\hat{\mu}_k} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\mu}_k\right)^3$$

4.2 **Calculate $\mathbb{E}\left[(\bar{y}-\mu)^3\right]$. Under what condition is it zero?**

$$
\begin{aligned}
E\left[(\overline{Y}_n-\mu)^3\right] &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu)\right)^3\right] \\
&= \frac{1}{n^3}E\left[\left(\left(\sum_{i=1}^{n}(Y_i-\mu)\right)^2 + 2\sum_{i\neq j}^{n}(Y_i-\mu)(Y_j-\mu)\right)\left(\sum_{i=1}^{n}(Y_i-\mu)\right)\right] \\
&= \frac{1}{n^3}E\left[\left(\sum_{i=1}^{n}(Y_i-\mu)\right)^3\right] + E\left[\sum_{i\neq j}^{n}(Y_i-\mu)(Y_j-\mu)\right] \\
&+ E\left[2\sum_{i\neq j}^{n}(Y_i-\mu)(Y_j-\mu) + 3\sum_{i\neq j\neq k}^{n}(Y_i-\mu)(Y_j-\mu)(Y_k-\mu)\right] \\
&= \frac{1}{n^3}E\left[\left(\sum_{i=1}^{n}(Y_i-\mu)\right)^3\right] \\
&= \frac{1}{n^2}E\left[(Y_i-\mu)^3\right]
\end{aligned}
$$

This value is zero if $\mathbb{E}(Y_i)=\mathbb{E}(Y)=\mu=0$.

4.3 **Explain the difference between $\bar{y}$ and μ, and the difference between $\frac{1}{n}\sum_{i=1}^{n}X_iX_i'$ and $\mathbb{E}[X_iX_i']$.**

The difference between $\overline{Y}$ and μ is that $\overline{Y}$ is a random variable that equals μ *in expectation*, whereas μ is a constant. Furthermore, $\overline{Y}$ is both the mean of a random sample and an esimator of μ, whereas μ is the mean of the full distribution of Y.

Similarly, $\frac{1}{n}\sum_{i=1}^{n}X_iX_i'$ is a random variable and estimator of $\mathbb{E}[X_iX_i']$, which is a constant.

4.4 **True or False: if $\hat{e}_i$ is the OLS residual from the linear regression of Y_i on X_i, then $\sum_{i=1}^{n}X_i^2\hat{e}_i=0$**

True. Since $\mathbb{E}[e_i|X_i]=0$, $\mathbb{E}[\hat{e}_i]=e_i$. Then,

$$
E\left[\sum_{i=1}^{n}X_i^2\hat{e}_i\right] = E\left[E\left[\sum_{i=1}^{n}X_i^2\hat{e}_i|X_i\right]\right] = \left[X_i^2E\left[\sum_{i=1}^{n}\hat{e}_i|X_i\right]\right] = 0
$$

4.5 **Prove (4.15) and (4.16)**

$$\begin{aligned}
\mathbb{E}[\widehat{\beta}|X] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\sum_{i=1}^{n} X_i Y_i\right) \mid X\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\mathbb{E}\left[\left(\sum_{i=1}^{n} X_i Y_i\right) \mid X\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\sum_{i=1}^{n}\mathbb{E}\left[X_i Y_i \mid X\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\sum_{i=1}^{n} X_i\,\mathbb{E}\left[Y_i \mid X\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\sum_{i=1}^{n} X_i X_i'\beta \\
\mathbb{E}[\widehat{\beta}|X] &= \beta \qquad (4.15) \\
Var[\widehat{\beta}|X] &= Var\left(\left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\sum_{i=1}^{n} X_i Y_i\right) \mid X\right) \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} Var\left(\sum_{i=1}^{n} X_i Y_i \mid X\right)\left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} X_i' Var\left(\sum_{i=1}^{n} Y_i \mid X\right) X_i \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}(X_i'\sigma^2 I X_i)\left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \\
Var[\widehat{\beta}|X] &= (X'X)^{-1}(X'\Omega X)(X'X)^{-1} \qquad (4.16)
\end{aligned}$$

4.6 **Prove Theorem 4.5**

Suppose that $Var(\varepsilon|X) = \Omega$ for some positive semi-definite Ω. Then for $Y = X'\beta + \varepsilon$, the linear unbiased estimator of β is:

$$\tilde{\beta} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$$

Then, the variance of $\tilde{\beta}$ is:

$$\begin{aligned}
Var\left(\tilde{\beta} \mid X\right) &= Var\left(\left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \mid X\right) \\
&= \left(X'\Omega^{-1}X\right)^{-1} Var\left(X'\Omega^{-1}Y \mid X\right) \left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1} Var\left(Y \mid X\right) \Omega^{-1}X \left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\Omega\Omega^{-1}X \left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}X \left(X'\Omega^{-1}X\right)^{-1} \\
Var\left(\tilde{\beta} \mid X\right) &= \left(X'\Omega^{-1}X\right)^{-1}
\end{aligned}$$