ECON 709B - Problem Set 1

Alex von Hafften*

11/11/2020

1. 2.1 - 2.2[1]

2.1 Find $E[E[E[Y|X_1, X_2, X_3]|X_1, X_2]|X_1]$.

By the law of iterated expectations,

$$E[E[E[Y|X_1, X_2, X_3]|X_1, X_2]|X_1] = E[E[Y|X_1, X_2]|X_1] = E[Y|X_1]$$

2.2 If $E[Y|X] = a + bX$, find $E[YX]$ as a function of moments of X.

By the conditioning theorem,

$$E[YX] = E[E[YX|X]] = E[XE[Y|X]] = E[X(a + bX)] = E[aX + bX^2] = aE[X] + bE[X^2]$$

2. 2.3 Prove conclusion (4) of Theorem 2.4.

If $E|Y| < \infty$ then for any function $h(x)$ such that $E|h(X)e| < \infty$ then $E[h(X)e] = 0$.

Proof: Let h be a function such that $E|h(X)e| < \infty$. By the conditioning theorem and conclusion (1) of Theorem 2.4 (i.e., $E[e|X] = 0$),

$$E[h(X)e] = E[E[h(X)e|X]] = E[h(X)E[e|X]] = E[h(X)(0)] = E[0] = 0$$

□

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Emily Case, Sarah Bass, and Danny Edgel.

[1]These problems come from *Econometrics* by Bruce Hansen, revised on October 23, 2020.

3. 2.4 Suppose that the random variables Y and X only take the values 0 and 1, and have the following joint probability distribution

	$X = 0$	$X = 1$
$Y = 0$	.1	.2
$Y = 1$	.4	.3

Find $E[Y|X]$, $E[Y^2|X]$ and $\text{var}[Y|X]$ for $X = 0$, $X = 1$.

$$\begin{aligned}
E[Y|X = 0] &= (1)P[Y = 1|X = 0] + (0)P[Y = 0|X = 0] = (1)(.4)/(.5) = .8 \\
E[Y|X = 1] &= (1)P[Y = 1|X = 1] + (0)P[Y = 0|X = 1] = (1)(.3)/(.5) = .6 \\
E[Y^2|X = 0] &= (1)^2P[Y = 1|X = 0] + (0)^2P[Y = 0|X = 0] = (1)^2(.4)/(.5) = .8 \\
E[Y^2|X = 1] &= (1)^2P[Y = 1|X = 1] + (0)^2P[Y = 0|X = 1] = (1)^2(.3)/(.5) = .6 \\
\text{var}[Y|X = 0] &= E[Y^2|X = 0] - (E[Y|X = 0])^2 = (.8) - (.8)^2 = 0.16 \\
\text{var}[Y|X = 1] &= E[Y^2|X = 1] - (E[Y|X = 1])^2 = (.6) - (.6)^2 = 0.24
\end{aligned}$$

4. 2.5 (c) Show that $\sigma^2(X)$ is the best predictor of e^2 given X. Show that $\sigma^2(X)$ minimizes the mean-squared error and is thus the best predictor.

For $S(X)$ some predictor of e^2 given X:

$$\begin{aligned}
E[(e^2 - S(X))^2] &= E[(e^2 - \sigma^2(X) + \sigma^2(X) - S(X))^2] \\
&= E[(e^2 - \sigma^2(X))^2] + 2E[(e^2 - \sigma^2(X))(\sigma^2(X) - S(X))] + E[(\sigma^2(X) - S(X))^2]
\end{aligned}$$

The middle term is zero:

$$\begin{aligned}
E[(e^2 - \sigma^2(X))(\sigma^2(X) - S(X))] &= E[E[(e^2 - \sigma^2(X))(\sigma^2(X) - S(X))|X]] \\
&= E[E[(e^2 - \sigma^2(X))|X](\sigma^2(X) - S(X))] \\
&= E[(E[(e^2|X] - E[\sigma^2(X)|X])(\sigma^2(X) - S(X))] \\
&= E[(\sigma^2(X) - \sigma^2(X))(\sigma^2(X) - S(X))] \\
&= 0
\end{aligned}$$

Thus,

$$E[(e^2 - S(X))^2] = E[(e^2 - \sigma^2(X))^2] + E[(\sigma^2(X) - S(X))^2]$$

The choice of S does not change the first term and the second term is minimized when $S(X) = \sigma^2(X)$. Thus, $\sigma^2(X)$ is the best predictor of e^2.

5. 2.8 Suppose that Y is discrete-valued, taking values only on the non-negative integers, and the conditional distribution of Y given $X = x$ is Poisson: $P[Y = j|X = x] = \frac{\exp(-x'\beta)(x'\beta)^j}{j!}, j = 0, 1, 2, ...$ Compute $E[Y|X]$ and $\text{var}[Y|X]$. Does this justify a linear regression model of the form $Y = X'\beta + e$?[2]

Using the hint, we know that $E[Y|X] = x'\beta$ and $\text{var}[Y|X] = x'\beta$.

Yes, this justifies a linear regression model because $E[e|X] = E[Y - X'\beta|X] = E[Y|X] - E[X'\beta|X] = x'\beta - x'\beta = 0$.

6. 2.10 - 2.14 Explain your answers.

2.10 If $Y = X\beta + e, X \in \mathbb{R}$, and $E[e|X] = 0$, then $E[X^2 e] = 0$.

True, by the conditioning theorem:

$$E[X^2 e] = E[E[X^2 e|X]] = E[X^2 E[e|X]] = E[X^2(0)] = E[0] = 0$$

2.11 If $Y = X\beta + e, X \in \mathbb{R}$, and $E[Xe] = 0$, then $E[X^2 e] = 0$.

False, for a counter example, assume $X \sim N(0, 1)$ and e is a degenerate random variable equal to 1. Notice that $E[Xe] = E[X] = 0$ and $E[X^2 e] = E[X^2] = 1$.

2.12 If $Y = X'\beta + e$, and $E[e|X] = 0$, then e is independent of X.

False, for a counter example, assume $X = (X_1, ..., X_k)'$ and $U = (U_1, ..., U_k)'$ where $X_1, ..., X_k$ and $U_1, ..., U_k$ are independently distributed standard normal. Let $e = X'U$. Thus, conditional on X, e is distributed $N(0, X'X)$, so $E[e|X] = 0$, but X and e are not independent.

2.13 If $Y = X'\beta + e$, and $E[Xe] = 0$, then $E[e|X] = 0$.

False, for a counter example, assume $X = (X_1, ..., X_k)'$ where $X_1, ..., X_k \sim N(0, 1)$ and e is a degenerate random variable equal to 1. Notice that $E[Xe] = E[X] = 0$ and $E[e|X] = E[e] = 1$.

2.14 If $Y = X'\beta + e$, and $E[e|X] = 0$, and $E[e^2|X] = \sigma^2$, then e is independent of X.

False, for a counter example, assume $X = (X_1, ..., X_k)'$ where $X_i \sim N(0, 1)$ and $Z = (Z_1, ...Z_k)$ where $Z_i \sim N(1, \sigma^2/x_i^2)$. Thus, $E[Z_i|X_i] = 1$ and $\text{var}[Z_i|X_i] = \sigma^2/x_i^2$. Define $Y := X'Z$. Notice that $E[Y|X] = E[X'Z|X] = X'E[Z|X] = \sum_{i=1}^n X_i$, so $\beta = (1, ..., 1)'$. Define $e := Y - E[Y|X] = Y - \sum_{i=1}^n X_i = X'(Z - 1)$. Notice that $E[e|X] = E[Y - E[Y|X]|X] = 0$ and $E[e^2|X] = E[X'(Z - 1)(Z - 1)'X|X] = X'E[(Z - 1)(Z - 1)'|X]X = X'E[(Z - 1)(Z - 1)'|X]X = X'\text{var}[Z]X = X'(X'\sigma^2 X)^{-1}X = \sigma^2$. However, X and e are not independent.

[2]Hint: $P[Y = j] = \frac{\exp(-\lambda)(\lambda)^j}{j!}$, then $E[Y] = \lambda$ and $\text{var}[Y] = \lambda$.

7. 2.16 Let X and Y have the joint density $f(x,y) = \frac{3}{2}(x^2+y^2)$ on $0 \leq x \leq 1, 0 \leq y \leq 1$. Compute the coefficients of the best linear predictor $Y = \alpha + \beta X + e$. Compute the conditional expectation $m(x) = E[Y|X=x]$. Are the best linear predictor and conditional expectation different?

Best Linear Predictor (BLP):

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 \frac{3}{2}(x^2+y^2)dy = \frac{3}{2}\left[x^2y + \frac{y^3}{3}\right]_{y=0}^1 = \frac{3}{2}x^2 + \frac{1}{2}$$

$$E[X] = \int_0^1 x\left(\frac{3}{2}x^2 + \frac{1}{2}\right)dx = \left[\frac{3}{8}x^4 + \frac{1}{4}x^2\right]_0^1 = \frac{5}{8}$$

$$E[X^2] = \int_0^1 x^2\left(\frac{3}{2}x^2 + \frac{1}{2}\right)dx = \left[\frac{3}{10}x^5 + \frac{1}{6}x^3\right]_0^1 = \frac{7}{15}$$

Since the joint distribution is symmetric, $E[Y] = \frac{5}{8}$ and $E[Y^2] = \frac{7}{15}$.

$$\begin{aligned}
E[XY] &= \int_0^1 \int_0^1 xy\frac{3}{2}(x^2+y^2)dxdy \\
&= \frac{3}{2}\int_0^1 \int_0^1 x^3y + xy^3 dxdy \\
&= \frac{3}{2}\int_0^1 \left[\frac{x^4y}{4} + \frac{x^2y^3}{2}\right]_{x=0}^1 dy \\
&= \frac{3}{2}\int_0^1 \frac{y}{4} + \frac{y^3}{2}dy \\
&= \frac{3}{2}\left[\frac{y^2}{8} + \frac{y^4}{8}\right]_0^1 \\
&= \frac{3}{8}
\end{aligned}$$

$$\begin{aligned}
S(\alpha,\beta) &= E[(Y-\alpha-\beta X)^2] \\
&= E[Y^2 - \alpha Y - \beta XY - \alpha Y + \alpha^2 + \alpha\beta X - \beta XY + \alpha\beta X + \beta^2 X^2] \\
&= E[Y^2] - \alpha E[Y] - \beta E[XY] - \alpha E[Y] + \alpha^2 + \alpha\beta E[X] - \beta E[XY] + \alpha\beta E[X] + \beta^2 E[X^2] \\
&= \frac{7}{15} - \frac{5}{8}\alpha - \frac{3}{8}\beta - \frac{5}{8}\alpha + \alpha^2 + \frac{5}{8}\alpha\beta - \frac{3}{8}\beta + \frac{5}{8}\alpha\beta + \frac{7}{15}\beta^2 \\
&= \frac{7}{15} - \frac{5}{4}\alpha - \frac{3}{4}\beta + \alpha^2 + \frac{5}{4}\alpha\beta + \frac{7}{15}\beta^2
\end{aligned}$$

FOC $[\alpha]$:

$$0 = -\frac{5}{4} + 2\alpha + \frac{5}{4}\beta \implies \beta = 1 - \frac{8}{5}\alpha$$

FOC [β]:

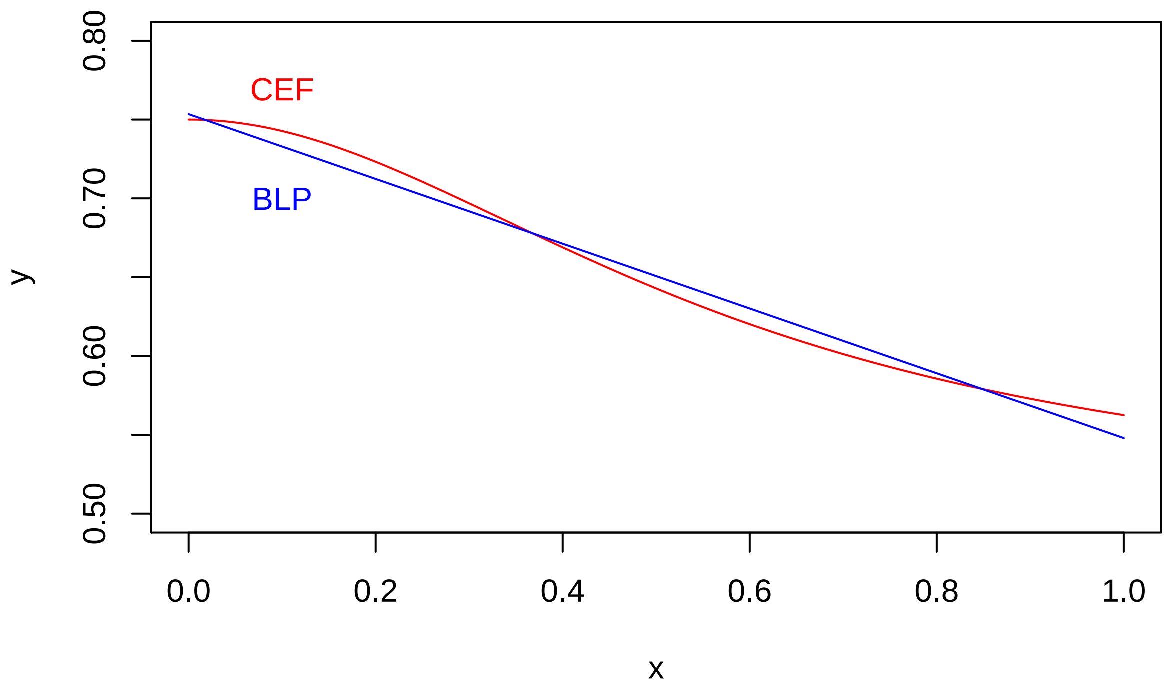
$$0 = -\frac{3}{4} + \frac{5}{4}\alpha + \frac{14}{15}\beta \implies 45 = 75\alpha + 56\beta$$

$$45 = 75\alpha + 56\left(1 - \frac{8}{5}\alpha\right) \implies \alpha = \frac{55}{73} \implies \beta = \frac{-15}{73}$$

Conditional Expectation Function (CEF):

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{3}{2}(x^2+y^2)}{\frac{3}{2}x^2 + \frac{1}{2}} = \frac{3x^2+3y^2}{3x^2+1}$$

$$m(x) = E[Y|X=x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy = \int_0^1 \frac{3x^2y+3y^3}{3x^2+1}dy = \left[\frac{\frac{3}{2}x^2y^2 + \frac{3}{4}y^4}{3x^2+1}\right]_{y=0}^{1} = \frac{6x^2+3}{12x^2+4}$$



8. 4.1 - 4.6

4.1 For some integer k, set $\mu_k = E[Y^k]$.

(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 for μ_k.

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

(c) Calculate the variance of $\hat{\mu}_k$, say $\text{var}[\hat{\mu}_k]$. What assumption is needed for $\text{var}[\hat{\mu}_k]$ to be finite?

We need to assume that $|\mu_{2k}| < \infty$ for $\text{var}[\hat{\mu}_k]$ to be finite:

$$\text{var}[\hat{\mu}_k] = \text{var}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^k\right] = \frac{1}{n^2}\sum_{i=1}^{n} \text{var}[Y_i^k] = \frac{1}{n^2}\sum_{i=1}^{n}(E[Y_i^{2k}] - E[Y_i^k]^2) = \frac{1}{n^2}\sum_{i=1}^{n}(\mu_{2k} - \mu_k^2) = \frac{\mu_{2k} - \mu_k^2}{n}$$

(d) Propose an estimator of $\text{var}[\hat{\mu}_k]$.

$$\frac{\hat{\mu}_{2k} - \hat{\mu}_k^2}{n} = \frac{n^{-1}\sum_{i=1}^{n} Y_i^{2k} - (n^{-1}\sum_{i=1}^{n} Y_i^k)^2}{n}$$

4.2 Calculate $E[(\bar{Y}-\mu)^3]$, the skewness of $\bar{y}$. Under what conditions is it zero?

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

The skewness of $\bar{y}$ is zero if the skewness of Y_i is zero ($E[(Y_i-\mu)^3]=0$). The skewness of $\bar{y}$ approaches zero as n gets large.

4.3 Explain the difference between $\bar{Y}$ and μ. Explain the difference between $n^{-1}\sum_{i=1}^{n}X_iX_i'$ and $E[X_iX_i']$.

The difference between $\bar{Y}$ and μ is $\bar{Y}$ is a statement about a sample and μ is a statement about a population. Namely, $\bar{Y}$ is the sample mean and μ is the population mean. Similarly, $n^{-1}\sum_{i=1}^{n}X_iX_i'$ is the sample variance and $E[X_iX_i']$ is the population variance.

4.4 True or False. If $Y_i = X_i\beta + e_i, X_i \in \mathbb{R}, E[e_i|X_i] = 0$, and $\hat{e}_i$ is the OLS residual from the regression of Y_i on X_i, then $\sum_{i=1}^n X_i^2 \hat{e}_i = 0$.

False. Counter example with simulated data:

```
beta <- 2
x <- runif(n = 100)
e <- rnorm(n=100)
y <- x * beta + e
beta_hat <- as.numeric((y %*% x)/(x %*% x))
print(beta_hat)
```

```
## [1] 2.418011
```

```
e_hat <- y - x * beta_hat
print(sum(x^2 %*% e_hat))
```

```
## [1] 0.09543305
```

4.5 Prove (4.15) and (4.16).

(4.15) $E[\hat{\beta}|X] = \beta$

Since $E[Y|X] = E[X\beta + e|X] = E[X\beta|X] + E[e|x] = X\beta$,

$$E[\hat{\beta}|X] = E[(X'X)^{-1}X'Y|X] = (X'X)^{-1}X'E[Y|X] = (X'X)^{-1}X'X\beta = \beta$$

(4.16) $\text{var}[\hat{\beta}|X] = (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$

Since $\text{var}[Y|X] = \text{var}[X\beta + e|X] = \text{var}[e|X] = \Omega$,

$$\begin{aligned}
\text{var}[\hat{\beta}|X] &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X] \\
&= E[((X'X)^{-1}X'e)((X'X)^{-1}X'e)'|X] \\
&= E[(X'X)^{-1}(X'ee'X)(X'X)^{-1}|X] \\
&= (X'X)^{-1}(X'E[ee'|X]X)(X'X)^{-1} \\
&= (X'X)^{-1}(X'\text{var}[e|X]X)(X'X)^{-1} \\
&= (X'X)^{-1}(X'\Omega X)(X'X)^{-1}
\end{aligned}$$

4.6 Prove Generalized Gauss-Markov Theorem (Theorem 4.5): In the linear regression model (Assumption 4.2) and $\Omega > 0$, if $\tilde{\beta}$ is a linear unbiased estimator of β then $\text{var}[\tilde{\beta}|X] \geq (X'\Omega^{-1}X)^{-1}$.

Let $\tilde{\beta}$ be a linear unbiased estimator. Thus, $\tilde{\beta} = A'y$ for some A that is $n \times k$ where $A'X = I_k$. The variance of $\tilde{\beta}$ is $\text{var}[\tilde{\beta}|X] = \text{var}[A'y|X] = A'\text{var}[y|X]A = A'\text{var}[e|X]A = A'\Omega A$. Defining $C = A - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}$:[3]

$$\begin{aligned}
A'\Omega A &= (C + \Omega^{-1}X(X'\Omega^{-1}X)^{-1})'\Omega(C + \Omega^{-1}X(X'\Omega^{-1}X)^{-1}) \\
&= C'\Omega C + C'\Omega\Omega^{-1}X(X'\Omega^{-1}X)^{-1} + (\Omega^{-1}X(X'\Omega^{-1}X)^{-1})'\Omega C + (\Omega^{-1}X(X'\Omega^{-1}X)^{-1})'\Omega(\Omega^{-1}X(X'\Omega^{-1}X)^{-1}) \\
&= C'\Omega C + C'X(X'\Omega^{-1}X)^{-1} + (X(X'\Omega^{-1}X)^{-1})'C + (\Omega^{-1}X(X'\Omega^{-1}X)^{-1})'(X(X'\Omega^{-1}X)^{-1}) \\
&= (\Omega^{1/2}C')(\Omega^{1/2}C) + (X'\Omega^{-1}X)^{-1}
\end{aligned}$$

Since $(\Omega^{1/2}C')(\Omega^{1/2}C) \geq 0 \implies \text{var}[\tilde{\beta}|X] \geq (X'\Omega^{-1}X)^{-1}$.

[3] Notice that $X'C = X'(A - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}) = I - I = 0$.