(5.1) $f(x,y) = \begin{cases} \dfrac{3(x^2+y)}{11} & \text{, for } 0 \leq x \leq 2 \\ & \phantom{\text{, for }} 0 \leq y \leq 1 \\ 0 & \text{, otherwise} \end{cases}$

(a) for $0 \leq x \leq 2$, find the CEF of Y given X

$$E(Y|x) = \int_{-\infty}^{\infty} y \cdot g_2(y|x) \cdot dy$$

$$= \int_{-\infty}^{\infty} y \cdot \frac{f(x,y)}{f_1(x)} \cdot dy$$

$$f_1(x) = \int_0^1 \frac{3(x^2+y)}{11} \cdot dy$$ (Ch. 4).

$$f_1(x) = \frac{3(2x^2+1)}{22}$$ for $0 \leq x \leq 2$

$$E(Y|X) = \int_0^2 \frac{y\,(3)\,(x^2+y)\,22}{11\,(3)\,(2x^2+1)} \cdot dy$$

$$= \int_0^1 \frac{2yx^2 + 2y^2}{(2x^2+1)} \cdot dy$$

$$= \left[\frac{1}{(2x^2+1)} \left(y^2x^2 + \frac{2}{3}y^3 \right) \right]_0^1 = \frac{3x^2+2}{6x^2+3}$$

(b) • $E(X) = \int_{-\infty}^{\infty} x \cdot f_1(x) \cdot dx$

$$= \int_0^2 [x \cdot 3(2x^2+1)]/22 \cdot dx.$$

$$= \frac{3}{22} \int_0^2 2x^3 + x \cdot dx = \frac{3}{22} \left[\frac{x^4}{2} + \frac{x^2}{2} \right]_0^2$$

$$= \frac{3}{22(2)} [16+4] = \frac{\cancel{20}^{\cancel{10}\,5}(3)}{\cancel{22}_{11}(\cancel{2})} = \frac{15}{11} = 1.36$$

• $E(Y)$

$$f_1(y) = \int_{-\infty}^{\infty} f(x,y) \cdot dx$$

$$= \int_0^2 3(x^2+y)/11 \cdot dx = \frac{3}{11} \left[\frac{x^3}{3} + xy \right]_0^2$$

$$= \frac{8+6y}{11}$$

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f_1(y) \cdot dy$$

$$= \int_0^1 y \cdot (8+6y)/11 \cdot dy = \frac{1}{11} \left[\frac{8y^2}{2} + \frac{\cancel{6}^{\,2} y^3}{\cancel{3}} \right]_0^1$$

$$= \frac{8+4}{22} = \frac{12}{22} = \frac{6}{11} = 0.55$$

- $E(x^2) = \int_{-\infty}^{\infty} x^2 . f_1(x) . dx$

$$= \int_0^2 x^2 . \left[\frac{3(2x^2+1)}{22}\right] . dx = \frac{3}{22}\left[\frac{2x^5}{5} + \frac{x^3}{3}\right]_0^2$$

$$= \frac{8(3)}{22}\left[\frac{24+5}{15}\right] = \frac{4}{55}(29) = 2.11$$

- $E(Y^2) = \int_{-\infty}^{\infty} y^2 . f_1(y) . dy$

$$= \int_0^1 y^2 (8+6y)/11 . dy = \frac{1}{11}\left[\frac{8y^3}{3} + \frac{\cancel{6}^{3} y^4}{\cancel{4}_{2}}\right]_0^1$$

$$= \frac{16+9}{66} = 0.38$$

- $E(XY) = E_X[X E(Y|X)]$

$$= E_X\left[X . \frac{3x^2+2}{6x^2+3}\right] = \int_{-\infty}^{\infty} h(x) . f_1(x) . dx$$

$$= \int_0^2 x . \left(\frac{3x^2+2}{6x^2+3}\right)\left(\frac{3}{22}(2x^2+1)\right) . dx$$

$$= \frac{\cancel{3}}{22}\int_0^2 \frac{(3x^3+2x)\cancel{(2x^2+1)}}{\cancel{3}\cancel{(2x^2+1)}} . dx$$

$$= \frac{1}{22}\left[\frac{3x^4+4x^2}{4}\right]_0^2 = \frac{1}{22}[16] = 0.73$$

- $V(X) = E(X^2) - E^2(X)$
  $= 2.11 - (1.36)^2 = 0.26$

- $V(Y) = E(Y^2) - E^2(Y)$
  $= 0.38 - (0.55)^2 = 0.08$

- $C(X,Y) = E(XY) - E(X)E(Y)$
  $= (0.73) - (1.36)(0.55) = -0.018$

(c) Find the BLP of Y given X

$$\beta = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{-0.018}{0.26} = -0.07$$

$$\alpha = \mu_y - \beta\mu_x = 0.55 + 0.07(1.36) = 0.65$$

$$E^*(Y|X) = 0.65 - 0.07X$$

(d) From the graph, BLP seems to be a good approximation of CEF.

(e) It does not seem so.

(5.6) The r.v X and Y are jointly distributed. Let $\varepsilon = Y - E(Y|X)$ and $U = Y - E^*(Y|X)$, where $E(Y|X)$ is the CEF and $E^*(Y|X)$ is the BLP. Determine whether the following is true or false: $C(\varepsilon, U) = V(\varepsilon)$

$$
\begin{aligned}
C(\varepsilon, U) &= E(\varepsilon U) - E(\varepsilon)E(U) \\
&= E\left[(Y - E(Y|X))(Y - E^*(Y|X))\right] - E(Y - E(Y|X)).E(Y - E^*(Y|X)) \\
&= E\left[Y^2 - Y E^*(Y|X) - Y E(Y|X) + E(Y|X)E^*(Y|X)\right] - \\
&\quad - (E(Y) - E(Y|X))(E(Y) - E^*(Y|X)) \\
&= E(Y^2) - \cancel{E(Y)E^*(Y|X)} - \cancel{E(Y)E(Y|X)} + \cancel{E(Y|X)E^*(Y|X)} \\
&\quad - E^2(Y) + \cancel{E(Y)E^*(Y|X)} + \cancel{E(Y)E(Y|X)} - \cancel{E(Y|X)E^*(Y|X)} \\
&= E(Y^2) - E^2(Y) \qquad -(1)
\end{aligned}
$$

$$
\begin{aligned}
V(\varepsilon) &= E\left[(\varepsilon - E(\varepsilon))^2\right] \\
&= E\left[(Y - E(Y|X) - E(Y - E(Y|X))^2\right] \\
&= E\left[(Y - \cancel{E(Y|X)} - E(Y) - \cancel{E(Y|X)})^2\right] \\
&= E\left[Y^2 + E^2(Y) - 2Y E(Y)\right] \\
&= E(Y^2) + E^2(Y) - 2E^2(Y) \\
&= E(Y^2) - E^2(Y) \qquad -(2)
\end{aligned}
$$

As $(1) = (2)$, $C(\varepsilon, U) = V(\varepsilon)$

(5.8) In a bivariate population, let us define the best proportional predictor of Y given X as the ray through the origin, $E^{**}(Y|X) = \gamma X$ with γ being the value for c that minimizes $E(U^2)$, where now $U = Y - cX$

(a) Show that $\gamma = E(XY)/E(X^2)$

$U = Y - \gamma X$

Taking first-order condition & equating it to 0:

$$\frac{\partial E(U^2)}{\partial \gamma} = \frac{\partial}{\partial \gamma}\left[E(Y - \gamma X)^2\right]$$

$$= \frac{\partial}{\partial \gamma}\left[E(Y^2 + \gamma^2 X^2 - 2Y\gamma X)\right]$$

$$= E\left[X^2(2)\gamma - 2YX\right]$$

$$= 2\gamma E(X^2) - 2E[XY] = 0$$

(γ is a constant)

$$\Rightarrow \gamma = \frac{E[XY]}{E[X^2]}$$

$\square$

(b) Is $E^{**}(Y|X)$ an unbiased predictor? Explain.

A predictor is unbiased if the expected forecast error is zero.

$$E(\varepsilon) = E(Y - \gamma X)$$
$$= E(Y) - \frac{E[XY]}{E[X^2]} E[X]$$

$$E(\varepsilon) = E[Y - E^{**}(Y|X)]$$
$$= E[Y] - E[E^{**}(Y|X)]$$
$$= E[Y] - E[Y]$$
$$= 0$$

(Is this legal?)

So $E(\varepsilon) = 0$, the predictor is unbiased.

(c) Let $U = Y - \gamma X$. Does $C(X, U) = 0$?

$$C(X,U) = E\{[X - E[X]][U - E[U]]\}$$
$$= E\{[X - E[X]][Y - \gamma X - E[Y] - \gamma E[X]]\}$$
$$= E\{XY - \gamma X^2 - XE[Y] - \gamma XE[X] - YE[X]$$
$$+ \gamma XE[X] + E[X]E[Y] + \gamma E^2[X]\}$$
$$= E[XY] - \gamma E[X^2] - \cancel{E[X]}E[Y] - \cancel{\gamma E^2[X]}$$
$$- E[Y]E[X] + \gamma \cancel{E^2[X]} + E[X]\cancel{E[Y]} + \gamma E^2[X]$$

$$= E[XY] - E[Y]E[X] - \gamma E[X^2] + \gamma E^2[X]$$
$$= \cancel{E[XY]} - E[Y]E[X] - \cancel{E(XY)} + \frac{E(XY)}{E(X^2)} E^2[X]$$
$$= \frac{E(XY)\, E^2[X]}{E(X^2)} - E[Y]E[X]$$
$$= E[X] \left[\underbrace{\frac{E(XY)\, E[X]}{E(X^2)} - E[Y]}_{=0} \right]$$
$$= 0$$

[from part (b)]
so $E(\varepsilon) = 0$

(d) Find the minimized value of $E(U^2)$.

$$E(U^2) = V[Y - \gamma X] \qquad (\text{As } E(U) = 0)$$

$$= V(Y) + \gamma^2 V(X) - 2\gamma C(X,Y) \quad (\text{Can stop here?})$$

$$= V(Y) + \frac{E^2(XY)}{E^2(X^2)} V(X) - \frac{2E(XY)}{E(X^2)} C(X,Y)$$

$$= V(Y) + \frac{E^2(XY)}{E^2(X^2)}\left[E(X^2) - E^2(X)\right] - \frac{2E(XY)}{E(X^2)}\left[E(XY) - E(X)E(Y)\right]$$

$$= V(Y) + \frac{E^2(XY)}{E(X^2)} - \frac{E^2(XY)E^2(X)}{E^2(X^2)} - \frac{2E^2(XY)}{E(X^2)} + \frac{2\left[E(XY)E(X)E(Y)\right]}{E(X^2)}$$

$$= V(Y) - \frac{E^2(XY)}{E(X^2)} - \frac{E^2(XY)E^2(X)}{E^2(X^2)} + \frac{2E(XY)E(X)E(Y)}{E(X^2)}$$

$$= V(Y) - \gamma E(XY) - \gamma^2 E^2(X) + 2\gamma E(X)E(Y)$$

(e) Compare this $E(U^2)$ with those that result when the marginal expectation is used, and when the BLP is used.

- Using Marginal Expectation:-

Let

$V = Y - E(Y)$

$E(U^2) = E[(Y - E(Y))^2]$

$= V[Y - E(Y)]$

$= V(Y)$

As we already know from the text that $\epsilon = 0$, so $E(U) = 0$

- Using BLP:-

$E^*(Y|X) = \alpha + \beta X$

Let $U = Y - (\alpha + \beta X)$

$E(U^2) = V[Y - (\alpha + \beta X)]$ (Same as above)

$= V(Y) - \beta^2 V(X) - 2\beta C(X, Y)$

$= V(Y) - \beta^2 V(X)$ (On plugging value of $\beta = \frac{\sigma_{xy}}{\sigma_x^2}$ from the text)

Econometrics PS1

Tanya Sethi

(Q1) Joint distⁿ of (y, x)

	$Y=1$	$Y=2$	$Y=3$	Totals
$x=0$	0.10	0.40	0.15	0.65
$x=1$	0.05	0.15	0.15	0.35
Totals	0.15	0.55	0.30	1.00

(a) Compute the best linear predictor of Y given X. Compare the best linear predictor to the conditional expectation function.

• BLP of Y given X : $\beta = \frac{\sigma_{XY}}{\sigma_X^2}$; $\alpha = \mu_Y - \beta\mu_X$

$\sigma_{XY} = E(XY) - E(X)E(Y)$

$E(X) = 0(0.65) + 1(0.35) = 0.35$

$E(Y) = 1(0.15) + 2(0.55) + 3(0.30) = 2.15$

$$E(XY) = 1(0.05) + 2(0.15) + 3(0.15) = 0.8$$

$$\sigma_{XY} = 0.8 - 0.35(2.15) = 0.0475$$

$$\sigma_X^2 = E(X^2) - E^2(X)$$

$$E(X^2) = 0.35 - (0.35)^2 = 0.2275$$

$$\beta = \frac{0.0475}{0.2275} \simeq 0.21$$

$$\alpha = E(Y) - \beta E(X) = 2.15 - 0.21(0.35)$$

$$\alpha = 2.07$$

$$BLP = \alpha + \beta X = 2.07 + 0.21X$$

- Conditional Expectation Function

$$CEF = \begin{cases} (0.10 + 2(0.40) + 3(0.15))/0.65 = 2.07 & , \text{if } x = 0 \\ 0.05 + 2(0.15) + 3(0.15)/0.35 = 2.28 & , \text{if } x = 1 \end{cases}$$

(b) Compute the best linear predictor of X given Y. Compare the best linear predictor to the conditional expectation function of X given Y.

BLP of X given Y: $\beta = \dfrac{\sigma_{xy}}{\sigma_y^2}$; $\alpha = \mu_x - \beta\mu_y$

$\sigma_{xy} = 0.0475$

$\sigma_Y^2 = E(Y^2) - E^2(Y)$

$E(Y^2) = 1(0.15) + 4(0.55) + 9(0.30) = 5.05$

$\sigma_Y^2 = 5.05 - 4.6225 = 0.4275$

$\beta = 0.112$

$\alpha = 0.35 - 0.112(2.15) = 0.11$

$BLP = 0.11 + 0.112Y$

$$CEF(x|y) = \begin{cases} 0.05/0.15, & \text{if } y = 1 \\ 0.15/0.55, & \text{if } y = 2 \\ 0.15/0.3, & \text{if } y = 3 \end{cases}$$

(c) Discuss the following statement: "Private schools improve test scores. After all, the estimated effect of private school enrollment is to increase the probability of scoring 3 from 3/13 to 3/7.

No. The CEF shows that expectation of a student in private school is 0.5 if they get a 3 score?

(5) Suppose that $Z \sim N(42, 2500)$, $W \sim N(0, 500)$, that Z & W are independent, and that $X = Z + W$. Calculate the conditional expectation function $E(Z|X)$. How do you know that the CEF is linear?

As $Z \sim N(\cdot)$ & $W \sim N(\cdot)$; $X \sim N(\cdot)$ by the linear function result.

$$E(Z|X) = \int_{-\infty}^{\infty} z \cdot f(z|x) \cdot dz$$

$$= \int_{-\infty}^{\infty} z \cdot \frac{f(z,x)}{f_X(x)} \cdot dz \quad (*)$$

Now $X = Z + W$ where Z & W are independent

Let $A_1 \sim N(0,1)$ & $A_2 \sim N(0,1)$ & they are independent

$\Rightarrow Z = 42 + 2500 A_1$

$W = 0 + 500 A_2$

$$X = 42 + 2500 A_1 + 500 A_2$$

For given $Z = z_1$

$X = z_1 + 500\,A_2$

$\Rightarrow \quad X \mid z_1 \sim N(\cdot)$ with $E(X \mid z_1) = z_1$ &

$V(X \mid z_1) = (500)^2$

$$h(X \mid z_1) = \frac{f(A_2)}{500}$$

$$A_2 = (X - 42 - 2500\,A_1)/500$$

$$f(z, x) = f(z)\,f(a_2)/500$$

$$f(z) = \frac{e^{\{-[(z-\mu)/\sigma]^2/2\}}}{\sqrt{2\pi\sigma^2}} = \frac{e^{(-a_1^2/2)}}{\sqrt{2\pi(2500)^2}}$$

$$f(a_2) = \frac{e^{(-a_2^2/2)}}{\sqrt{2\pi}}$$

$$f(z, x) = \frac{e^{[-(a_1^2 + a_2^2)/2]}\,[2501]}{500(2500)\sqrt{2\pi}}$$

I can't figure out how to compute f_X as it is based on both A_1 & A_2.

However, once I compute it, I can plug it back in $(*)$ to compute $E(Z|X)$.