Set 40 Least-squares regression line

(a) Derive expressions for a and b that give the minimal value of the two-argument function defined by $g(a,b) = \mathbb{E}[(a + bX - Y)^2]$. (Hint: a is called the intercept and b the slope of the least-squares regression line, which is $y = a + bx$.)

$$g(a,b) = \mathbb{E}[(a + bX - Y)^2]$$

$$g(a,b) = \sum_{i=1}^{n} (a + bx_i - y_i)^2$$

Take the partial derivative with respect to a and find its min value

$$\sum_{i=1}^{n} (a + bx_i - y_i) = 0$$

$$\sum_{i=1}^{n} a + \sum_{i=1}^{n} bx_i = \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$$

$$\frac{1}{n}\sum_{i=1}^{n} y_i = a + \frac{b}{n}\sum_{i=1}^{n} x_i$$

$$\mathbb{E}[Y] = a + \mathbb{E}[X]$$

$$a = \mathbb{E}[Y] - b\mathbb{E}[X]$$

$$a = \bar{Y} - b\bar{X}$$

Take the partial derivative with respect to b and find its min value

$$g(a,b) = \sum_{i=1}^{n} (\bar{Y} - b\bar{X} + bx_i - y_i)^2$$

$$g(a,b) = \sum_{i=1}^{n} \left((\bar{Y} - y_i) + b(x_i - \bar{X})\right)^2$$

$$\sum_{i=1}^{n}[(\bar{Y}-y_i)+b(x_i-\bar{X})](x_i-\bar{X})=0$$

$$\sum_{i=1}^{n}(\bar{Y}-y_i)(x_i-\bar{X})+b(x_i-\bar{X})^2=0$$

$$-\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)+b\sum_{i=1}^{n}(\bar{X}-x_i)^2=0$$

$$b\sum_{i=1}^{n}(\bar{X}-x_i)^2=\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)$$

$$b=\frac{\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)}{\sum_{i=1}^{n}(\bar{X}-x_i)^2}$$

Finally, substitute b in to find a

$$a=\bar{Y}-b\bar{X}$$

$$a=\bar{Y}-\frac{\sum_{i=1}^{n}(\bar{Y}-y_i)(\bar{X}-x_i)}{\sum_{i=1}^{n}(\bar{X}-x_i)^2}\bar{X}$$

(b) Construct empirical estimators $\hat{a}$ and $\hat{b}$ for a and b, respectively, and then compute the estimators and draw the corresponding least-squares regression line using the following data:

Year(i)	1	2	3	4	5	6	7	8	9	10	11	12
Market(x_i)	0.15	0.13	0.07	0.12	-0.04	0.31	0.23	0.31	0.02	-0.07	0.07	0.02
Fund(y_i)	-0.05	0.05	0.01	0.25	0.04	0.15	0.40	0.29	0.33	-0.03	0.02	-0.02

$$\bar{X}=\frac{15+13+7+12-4+31+23+31+2-7+7+2}{100*12}=0.11$$

$$\bar{Y}=\frac{-5+5+1+25+4+15+40+29+33-3+2-2}{100*12}=0.12$$

$$\begin{aligned}\mathrm{cov}(X,Y)=&[(12+5)*(11-15)+(12-5)*(11-13)+(12-1)*(11-7)\\&+(12-25)*(11-12)+(12-4)*(11+4)+(12-15)*(11-31)\\&+(12-40)*(11-23)+(12-29)*(11-31)+(12-33)*(11-2)\\&+(12+3)*(11+7)+(12-2)*(11-7)+(12+2)*(11-2)]\div 100\\&=0.1078\end{aligned}$$

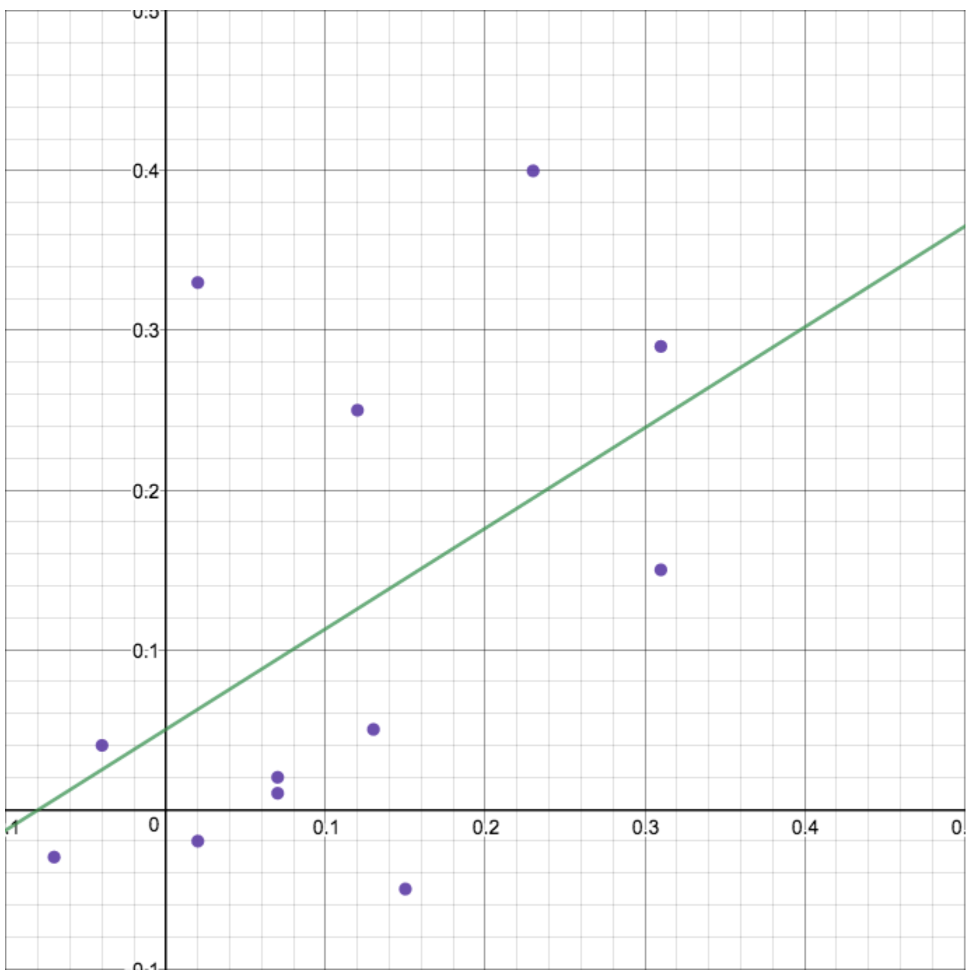
$$\text{var}(X) = [(11-15)^2 + (11-13)^2 + (11-7)^2 + (11-12)^2 + (11+4)^2 + (11-31)^2 + (11-23)^2 + (11-31)^2 + (11-2)^2 + (11+7)^2 + (11-7)^2 + (11-2)^2] \div 100^2 = 0.1708$$

Therefore,

$$b = \frac{\text{cov}(X,Y)}{\text{var}(X,Y)} = \frac{0.1078}{0.1708} \approx 0.63$$

$$a = \bar{Y} - \frac{\text{cov}(X,Y)}{\text{var}(X,Y)}\bar{X} = 0.12 - \frac{0.1078}{0.1708}0.11 \approx 0.05$$

$$y \approx 0.05 + 0.63x$$



(c) Prove the equation

$$\frac{\mathrm{cov}(X,Y)}{\sigma_X^2} = \mathrm{corr}(X,Y)\frac{\sigma_Y}{\sigma_X}$$

where $\mathrm{cov}(X,Y)$ is the covariance between X and Y, and $\mathrm{corr}(X,Y)$ is the correlation between X and Y. (Note: The left-hand side of equation is the famous "beta" that every financial portfolio manager knows and uses on the daily basis, with X being the market return and Y the fund return.)

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Therefore,

$$\frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}\frac{\sigma_Y}{\sigma_X}$$

$$\frac{\mathrm{cov}(X,Y)}{\sigma_X}\frac{1}{\sigma_X}$$

$$\frac{\mathrm{cov}(X,Y)}{\sigma_X^2}$$

(d) Prove that the mean squared error $MSE := \mathbb{E}\left[\left(\hat{Y} - Y\right)^2\right]$ between the least squares predictor $\hat{Y} = a + bX$ of the response variable Y is equal to $\sigma_Y^2(1-\rho^2)$ where σ_Y^2 is the variance of Y and $\rho = \mathrm{corr}(X,Y)$ is the correlation between X and Y.

$$\mathbb{E}\left[\left(\hat{Y} - Y\right)^2\right]$$

$$\hat{Y} = a + bX$$

$$\hat{Y} = \bar{Y} - \frac{\mathrm{cov}(X,Y)}{\sigma_X}\bar{X} + \frac{\mathrm{cov}(X,Y)}{\sigma_X}X$$

$$\mathbb{E}\left[\left(\mathbb{E}[Y] - \mathbb{E}[X]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + X\frac{\mathrm{cov}(X,Y)}{\sigma_X} - Y\right)^2\right]$$

$$\mathbb{E}\left[\mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + 2X\mathbb{E}[Y]\frac{\mathrm{cov}(X,Y)}{\sigma_X} - 2Y\mathbb{E}[Y] + \mathbb{E}[X]^2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2} \right.$$
$$- 2X\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2} + 2Y\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + X^2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2} - 2XY\frac{\mathrm{cov}(X,Y)}{\sigma_X}$$
$$\left. + Y^2\right]$$

$$\mathbb{E}[\mathbb{E}[Y]^2] + \mathbb{E}\left[-2\mathbb{E}[X]\mathbb{E}[Y]\frac{\mathrm{cov}(X,Y)}{\sigma_X}\right] + \mathbb{E}\left[2X\mathbb{E}[Y]\frac{\mathrm{cov}(X,Y)}{\sigma_X}\right] + \mathbb{E}[-2Y\mathbb{E}[Y]]$$
$$+ \mathbb{E}\left[\mathbb{E}[X]^2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}\right] + \mathbb{E}\left[-2X\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}\right] + \mathbb{E}\left[2Y\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)}{\sigma_X}\right]$$
$$+ \mathbb{E}\left[X^2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}\right] + \mathbb{E}\left[-2XY\frac{\mathrm{cov}(X,Y)}{\sigma_X}\right] + \mathbb{E}[Y^2]$$

$$\mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + 2\mathbb{E}[X]\mathbb{E}[Y]\frac{\mathrm{cov}(X,Y)}{\sigma_X} - 2\mathbb{E}[Y]\mathbb{E}[Y] + \mathbb{E}[X]^2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}$$
$$- 2\mathbb{E}[X]\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2} + 2\mathbb{E}[Y]\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + \mathbb{E}[X^2]\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}$$
$$- 2\mathbb{E}[XY]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + \mathbb{E}[Y^2]$$

$$\mathbb{E}[Y]^2 - 2\mathbb{E}[Y]^2 - \mathbb{E}[X]^2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2} + 2\mathbb{E}[Y]\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + \mathbb{E}[X^2]\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}$$
$$- 2\mathbb{E}[XY]\frac{\mathrm{cov}(X,Y)}{\sigma_X} + \mathbb{E}[Y^2]$$

$$\mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + \mathbb{E}[X^2]\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2} - \mathbb{E}[X]^2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2} - 2\mathbb{E}[XY]\frac{\mathrm{cov}(X,Y)}{\sigma_X}$$
$$+ 2\mathbb{E}[Y]\mathbb{E}[X]\frac{\mathrm{cov}(X,Y)}{\sigma_X}$$

$$\mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + (\mathbb{E}[X^2] - \mathbb{E}[X]^2)\left(\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}\right) - 2\left(\frac{\mathrm{cov}(X,Y)}{\sigma_X}\right)(\mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X])$$

$$\sigma_Y^2 + \sigma_X^2\left(\frac{\mathrm{cov}(X,Y)^2}{\sigma_X^2}\right) - 2\left(\frac{\mathrm{cov}(X,Y)}{\sigma_X}\right)\mathrm{cov}(X,Y)$$

$$\sigma_Y^2 + \mathrm{cov}(X,Y)^2 - 2\frac{\mathrm{cov}(X,Y)^2}{\sigma_X}$$

$$\sigma_Y^2 + \operatorname{cov}(X,Y)^2\left(1 - \frac{2}{\sigma_X}\right)$$

$$\sigma_Y^2 + \operatorname{cov}(X,Y)^2\left(\frac{\sigma_X}{\sigma_X} - \frac{2}{\sigma_X}\right)$$

$$\sigma_Y^2 + \operatorname{cov}(X,Y)^2\left(\frac{\sigma_X - 2}{\sigma_X}\right)$$

$$\sigma_Y^2 + \operatorname{cov}(X,Y)^2\left(\frac{\sigma_X - 2}{\sigma_X}\right)\left(\frac{\sigma_X}{\sigma_X}\right)$$

$$\sigma_Y^2 + \operatorname{cov}(X,Y)^2\sigma_X\left(\frac{\sigma_X - 2}{\sigma_X^2}\right)\left(\frac{\sigma_Y^2}{\sigma_Y^2}\right)$$

$$\sigma_Y^2\left(1 + \operatorname{cov}(X,Y)^2\sigma_X\left(\frac{\sigma_X - 2}{\sigma_X^2\sigma_Y^2}\right)\right)$$

$$\sigma_Y^2\left(1 + \sigma_X(\sigma_X - 2)\left(\frac{\operatorname{cov}(X,Y)^2}{\sigma_X^2\sigma_Y^2}\right)\right)$$

$$\sigma_Y^2(1 + (\sigma_X^2 - 2\sigma_X)\rho^2)$$

Doesn’t work…