ECON 6200

Problem Set 1

Gabe Sekeres

February 11, 2025

n.b. I'm using capitals without subscripts to denote matrices and capitals with subscripts to denote vectors. I refuse to use bold on principle.

1. Consider the projection $\hat{Y} = \hat{\alpha} + \hat{\beta}X$.

   (a) Recall that the OLS estimator is defined as

$$(\hat{\alpha}, \hat{\beta}) \equiv \underset{(a,b)}{\text{argmin}}(Y - a - bX)^2 \equiv \underset{(a,b)}{\text{argmin}} \sum_{i=1}^{n}(Y_i - a - bX_i)^2$$

We have the first order conditions

$$0 = -2\sum_{i=1}^{n}(Y_i - a - bX_i) \qquad (a)$$

$$0 = -2\sum_{i=1}^{n}(Y_i - a - bX_i)X_i \qquad (b)$$

Multiplying by $\frac{1}{n}$, the first condition becomes

$$\bar{Y} - a - b\bar{X} = 0 \Longrightarrow \hat{\alpha} = \bar{Y} - b\bar{X}$$

Substituting back into the second condition, we get

$$0 = \sum_{i=1}^{n}(Y_i - \bar{Y} + b\bar{X} - bX_i)X_i$$

$$0 = \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) + b\sum_{X_i} X_i(\bar{X} - X_i)$$

$$\text{so} \qquad \hat{\beta} = \frac{\sum_{i=1}^{n} X_i(Y_i - \bar{Y})}{\sum_{i=1}^{n} X_i(X_i - \bar{X})}$$

Expanding the numerator and denominator, we get:

$$\hat{\beta} = \frac{\sum(X_i - \bar{X} + \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X} + \bar{X})(X_i - \bar{X})} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y}) + \bar{X}\sum(Y_i - \bar{X})}{\sum(X_i - \bar{X})^2 + \bar{X}\sum(X_i - \bar{X})}$$

and since $\sum(X_i - \bar{X}) = \sum(Y_i - \bar{Y}) = 0$, we have that

$$\hat{\beta} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

   (b) Recall that the sample correlation coefficient between X and Y is defined by

$$R_{XY} = \frac{s_{xy}^2}{s_x^2 s_y^2} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

and R^2 is

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

Substituting our expression for $\hat{Y}$, recalling that $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$, we get

$$R^2 = \frac{\sum(\bar{Y} - \hat{\beta}\bar{X} + \hat{\beta}X_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\hat{\beta}^2 \sum(X_i - \bar{X})^2}{\sum(Y_i - \bar{Y})^2}$$

and using the expression for $\hat{\beta}$ from part (a), we have that this simplifies to

$$R^2 = \left(\frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}\right)^2 \frac{\sum(X_i - \bar{X})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\left(\sum(X_i - \bar{X})(Y_i - \bar{Y})\right)^2}{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2} = (R_{XY})^2$$

(c) Consider the projection $\hat{X} = \hat{\gamma} + \hat{\delta}Y$. This is exactly the same as in part (a), where we can show that

$$\hat{\delta} = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sum(Y_i - \bar{Y})^2}$$

This has the same numerator (since multiplication is commutative) as the above projection coefficient, but normalized to the variance of Y instead of X. Moreover, by repeating the process in part (b), we can see directly that in this regression, the R^2 is

$$R^2 = (R_{YX})^2 = \frac{\left(\sum(Y_i - \bar{Y})(X_i - \bar{X})\right)^2}{\sum(Y_i - \bar{Y})^2 \sum(X_i - \bar{X})^2} = (R_{XY})^2$$

So the R^2 for the projection of Y onto X is the same as for the projection of X onto Y.

2. Rank-rank regression (the dog ate Jörg's data)

(a) The regression we ran was the projection $\hat{X} = \hat{\alpha} + \frac{3}{5}Y$. Since in this case the exact means are known, we have that $\bar{X} = \bar{Y} = 500$. Recall that the OLS estimator is defined as

$$(\hat{\alpha}, \hat{\beta}) \equiv \underset{(a,b)}{\operatorname{argmin}} \sum_{i=1}^{n} (X_i - a - bY_i)^2$$

which admits the first order condition on $\hat{\alpha}$ of

$$0 = -2\sum_{i=1}^{n}(X_i - a - bY_i) \underbrace{\Longrightarrow}_{\cdot\frac{1}{n}} \hat{\alpha} = \bar{X} - b\bar{Y} \Longrightarrow \hat{\alpha} = 500 - \frac{3}{5} \cdot 500 = 200$$

(b) We have that

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{X}_i - \bar{X})^2}{\sum(X_i - \bar{X})^2}$$

Recalling that $\hat{X} = 200 + \frac{3}{5}Y$, that $\bar{X} = \bar{Y}$, and that $\bar{Y} = 200 + \frac{3}{5}\bar{Y}$, we can convert this to

$$R^2 = \frac{\sum(200 + 3/5 \cdot Y_i - \bar{Y})^2}{\sum(X_i - \bar{X})^2} = \frac{9}{15} \cdot \frac{\sum(Y_i - \bar{Y})^2}{\sum(X_i - \bar{X})^2} = \frac{9}{15} \cdot \frac{\mathrm{Var}(Y)}{\mathrm{Var}(X)} = \frac{9}{15}$$

where the last equality follows because X and Y are just reorderings of each other, so they have the same variance.

(c) As we saw in Problem 1, the R^2 for each regression is the same, so it is $\frac{9}{15}$ in both. Similarly from Problem 1, the estimated coefficients in the two directions have the relationship with R^2 such that $\hat{\beta}_{XY} \cdot \hat{\beta}_{YX} = R^2$. Since we know that $\hat{\beta}_{XY} = \frac{3}{5}$ and we know that $R^2 = \frac{9}{15}$, we know that the estimated coefficient for the correctly specified regression is also $\frac{3}{5}$. Finally, we can estimate the intercept using the same first order condition:

$$0 = -2\sum_{i=1}^{n}(Y_i - a - bX_i) \implies \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 500 - \frac{3}{5} \cdot 500 = 200$$

3. Consider the projection onto a categorical variable versus a set of indicator variables

(a) The projection onto a categorical variable is well-defined only if (i) the relationship between Y and the levels of X is explicitly linear, (ii) if levels of X are correctly defined *in order* of their effect on Y, and is well-defined without a constant only if (i) and (ii) hold and additionally as long as $Y = 0$ whenever $X = 0$. This extremely restrictive set of conditions will basically never be met in practice. On the contrary, the second projection is always well-defined, as long as the various categories are mutually exclusive and have at least slightly differential effects on Y, and as long as we omit one level of X (as is standard in this case). The constant is, however, necessary in this case.

(b) Observe that the projection is a projection from $\mathbb{R}^n$ into $\mathbb{R}^m + \hat{\alpha}$, where $\hat{\alpha}$ is the constant, as Z consists definitionally of an orthonormal basis for $\mathbb{R}^m$. Consider $\hat{Y}_i$, the ith component of the projection. We will have that $Z_j = 0$ for all $j \neq i$, and $Z_i = 1$. Thus, $\hat{Y}_i = \hat{\alpha} + \hat{\beta}_i Z_i$. Since we definitionally have that $\hat{\alpha} = \bar{Y}$, we can say that $\hat{Y}_i - \bar{Y} = \hat{\beta}_i Z_i$, so defining $n_i = \sum_{j=1}^{n} Z_i$, we have that

$$\hat{\beta}_i = \frac{\hat{Y}_i - \bar{Y}}{n_i} = \frac{1}{n_i}\sum_{j:X_j=i} Y_i = \mathbb{E}[Y \mid X = i]$$

which is the conditional sample average.

(c) Note that if there is a meaningful, precise linear relationship between Y and X, and the categories of X are correctly ordered, then all of that variation could be replicated by a regression on Z, and would recover the same coefficients. Thus, R^2 for the second regression will always be (weakly) higher than the first. It will almost always be strictly higher, as the second regression can also capture relationships that are non-linear in the categories of X. The R^2 for the two regressions will be the same if and only if the conditions from (a) hold and there is a precise linear relationship.

4. Consider projecting Y on a constant, X, and (possibly) X^2

(a) The population projection coefficient $\hat{\beta}$ is defined if and only if the matrix of covariates is nonsingular. Defining

$$M := \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} \mathbb{1} & X \end{bmatrix}$$

this condition becomes that $M'M$ is invertible, for which it is necessary and sufficient that there is some variance in X.

(b) For the population projection coefficient of Y on (X, X^2) to be defined, we need that the matrix $M'M$ is nonsingular, where $M = \begin{bmatrix} \mathbb{1} & X & X^2 \end{bmatrix}$. For a simple case where this fails, consider the case where $X_i^2 = X_i$ for all X, which could be the case when $X_i \in \{0, 1\}$ for all i. In this case, the column space of M would be (directly) linearly dependent, so $M'M$ would be singular.

(c) We have that $\tilde{Y} = \hat{a} + \hat{b}X$ and that $\hat{Y} = \hat{\alpha} + \hat{\beta}X + \hat{\gamma}X^2$. From Frisch-Waugh-Lovell, a sufficient condition such that $\hat{b} = \hat{\beta}$ is that the first order condition for regressing Y on X is the same as the first order condition for regressing Y on the residuals of a regression of X on X^2. That will be the case if X^2 explains precisely none of the variation in X – basically, if they are orthogonal. Consider the following example: if

$$X = \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases}$$

then the matrix $M'M$ is nonsingular, so the population projection coefficient is well-defined. However, since $X_i^2 = 1$ for all X_i, the residuals of the regression of X on X^2 are precisely X, so by Frisch-Waugh-Lovell $\hat{b} = \hat{\beta}$.