1 Regression Model

We will talk about the conditional mean model and the linear projection model.

Notation: in this note, y is a scale random variable, and x is a $K \times 1$ random vector.

1.1 Conditional Expectation Model

A regression model can be written as $y = m(x) + \epsilon$, where $m(x) = E[y|x]$ is called the *conditional mean function*, and $\epsilon = y - m(x)$ is called the *regression error*.

The error term ϵ satisfies the following properties.

- $E[\epsilon|x] = 0$,
- $E[\epsilon] = 0$,
- $E[h(x)\epsilon] = 0$, where h is a function of x.

The last one means that ϵ is uncorrelated with any function of x.

The conditional expectation function is of interest, because it is the best prediction of y under the *mean squared error* (MSE).[1]

Among all the functions $g(X)$, the conditional mean function $m(x)$ minimizes the MSE.

Proof. We take a guess-and-verify approach.

$$E\left[(y - g(x))^2\right] = E\left[(y - m(x))^2\right] + 2E\left[(y - m(x))(m(x) - g(x))\right] + E\left[(m(x) - g(x))^2\right].$$

The first term is irrelevant to $g(x)$. The second term is $2E[\epsilon(m(x) - g(x))] = 0$, which is again irrelevant of $g(x)$. The third term is minimized at $g(x) = m(x)$. □

1.2 Linear Projection Model

As discussed in the previous section, we are interested in the conditional mean function $m(x)$. However, $m(x)$ is a complex function depending on the joint distribution of (y, x). A special case is $m(x) = x'\beta$, that is, the conditional mean function is a linear function of x.[2] It is true if (y, x) follows a joint normal distribution. Even if $m(x) \neq x'\beta$, the linear $x'\beta$ is still useful as an approximation, as will be clear soon. Therefore, we may write the linear regression model, or the linear projection model, as

$$\begin{aligned} y &= x'\beta + e & (1) \\ E[xe] &= 0, & (2) \end{aligned}$$

where e is called the *projection error*. Eq.(2) implies that, if a constant is included in x, we have $E[e] = 0$ and moreover, $\text{cov}(x, e) = E[xe] = 0$.

[1] The quadratic loss function is between y and a prediction $g(x)$ is defined as $L(y, g(x)) = (y - g(x))^2$, and its expectation $R(y, g(x)) = E\left[(y - g(x))^2\right]$ is called the MSE.

[2] The linear function is not as restrictive as one might thought. It can be used to generate some nonlinear (in random variables) effect. For example, if $y = x_1\beta_2 + x_2\beta_2 + x_1x_2\beta_3 + e$, then $\frac{\partial}{\partial x_1} m(x_1, x_2) = \beta_1 + x_2\beta_3$, which is nonlinear in x_1, while it is still linear in the parameter β.

The coefficient β in the linear projection model has a straightforward closed-form. Multiplying x on both sides and taking expectation, we have $E[xy] = E[xx']\beta$. If $E[xx']$ is invertible, we explicitly solve

$$\beta = \left(E\left[xx'\right]\right)^{-1} E\left[xy\right]. \tag{3}$$

Even if $m(x) \neq x'\beta$, we are interested in β as is the *linear* minimizer of the MSE. That is,

$$\beta = \arg\min_{\beta \in \mathbb{R}^K} E\left[\left(y - x'\beta\right)^2\right]. \tag{4}$$

Proof. We look for such a β that minimizes $E\left[\left(y - x'\beta\right)^2\right]$. Set the first order condition to zero, $2E\left[x\left(y - x'\beta\right)\right] = 0$. We solve $\beta = \left(E\left[xx'\right]\right)^{-1} E\left[xy\right]$. □

In the meantime, $x'\beta$ is also the best *linear* approximation to $m(x)$.

Proof. If we replace y in (4) by $m(x)$, we solve the minimizer as

$$\left(E\left[xx'\right]\right)^{-1} E\left[xm\left(x\right)\right] = \left(E\left[xx'\right]\right)^{-1} E\left[E\left[xy|x\right]\right] = \left(E\left[xx'\right]\right)^{-1} E\left[xy\right] = \beta.$$

Therefore β is also the linear minimizer of $E\left[\left(m\left(x\right) - x'\beta\right)^2\right]$, the best linear approximation to $m(x)$ under MSE. □

1.2.1 Subvector Regression

Sometimes we are interested in a subvector of β, but not the entire vector β. For example, when we include a constant and some variables in x, we are often more interested in the slope coefficients (those associated with the random variables), as they represent the effect of these explanatory factors. In such a regression

$$y = \beta_1 + x'\beta_2 + e,$$

we take an expectation to get $E\left[y\right] = \beta_1 + E\left[x\right]'\beta_2$. Differentiate the two equations,

$$y - E\left[y\right] = \left(x - E\left[x\right]\right)'\beta_2,$$

so that

$$\beta_2 = \left(E\left[\left(x - E\left[x\right]\right)\left(x - E\left[x\right]\right)'\right]\right)^{-1} E\left[\left(x - E\left[x\right]\right)\left(y - E\left[y\right]\right)\right] = \left(var\left(x\right)\right)^{-1}\left(cov\left(x, y\right)\right),$$

where for two random vectors x and y (a scalar is a 1×1 vector), the variance and covariance are

$$\begin{aligned} \mathrm{var}\left(x\right) &= E\left[\left(x - E\left[x\right]\right)\left(x - E\left[x\right]\right)'\right] \\ \mathrm{cov}\left(x, y\right) &= E\left[\left(x - E\left[x\right]\right)\left(y - E\left[y\right]\right)\right], \end{aligned}$$

respectively. This is a special case of the subvector regression.

Don't read the part below. It is not ready yet.

To discuss the general case, we need to know the formula of the partitioned inverse, a fact from linear algebra. If $Q=\begin{pmatrix}Q_{11} & Q_{12}\\ Q_{21} & Q_{22}\end{pmatrix}$ is a symmetric and positive definite matrix, then

$$Q^{-1}=\begin{pmatrix}\left(Q_{11}-Q_{12}Q_{22}Q_{21}\right)^{-1} & -\left(Q_{11}-Q_{12}Q_{22}Q_{21}\right)^{-1}Q_{12}Q_{22}^{-1}\\ -\left(Q_{22}-Q_{21}Q_{11}Q_{12}\right)^{-1}Q_{21}Q_{11}^{-1} & \left(Q_{22}-Q_{21}Q_{11}Q_{12}\right)^{-1}\end{pmatrix}.$$

Let $A_{11\cdot2}=E\left[x_1x_1'\right]-E\left[x_1x_2'\right]\left(E\left[x_2x_2'\right]\right)^{-1}E\left[x_2x_1'\right]$, and $A_{1y\cdot2}=E\left[x_1y\right]-E\left[x_1x_2'\right]\left(E\left[x_2x_2'\right]\right)^{-1}E\left[x_2y\right]$ then $\beta_1=A_{11\cdot2}^{-1}A_{1y\cdot2}$. Why is this useful?

Let x_1 be a scalar and x_2 be a vector (with constant). We first run a regression

$$x_1=x_2'\gamma+u$$

so that $u=x_1-x_2'\gamma=x_1-x_2'\left(E\left[x_2x_2'\right]\right)^{-1}E\left[x_2x_1'\right]$. We then run a regression of y on u with a constant, so that

$$\beta_u=\frac{cov(u,y)}{var(u)}.$$

The nominator $cov\left(u,y\right)=E\left[x_1y\right]-E\left[x_1x_2'\right]\left(E\left[x_2x_2'\right]\right)^{-1}E\left[x_1y\right]$. The denominator $var\left(u\right)=E\left[u^2\right]=A_{11\cdot2}$.

1.3 Omitted Variable Bias

Long regression

$$y=x_1'\beta_1+x_2'\beta_2+\epsilon$$

and short regression

$$y=x_1'\gamma+u.$$

To discuss how to sign the bias, we first demean all the variables, which is equivalent as if we project out the effect of the constant.

$$\begin{aligned}\tilde{y}= & \quad\tilde{x}_1'\beta_1+\tilde{x}_2'\beta_2+\epsilon\\ \tilde{y}= & \qquad\qquad\quad\tilde{x}_1'\gamma+u\end{aligned}$$

where tilde denotes the demeaned variable. Now the cross moment equals to the covariance.