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 Function

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

The error term e satisfies three properties.

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

The last properties means that e is uncorrelated with any function of x.

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

Among all the functions $g(X)$, the conditional mean function $E[y|X]$ minimizes MSE.

Proof. Because the minimization is on a functional space, it is difficult to implement. 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 $g(x) = m(x)$. $\square$

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 that $m(x) = x'\beta$, that is, the conditional mean function is a linear function of x. It is true if (y, x) follows a joint normal distribution.

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,

$$y = x_1\beta_2 + x_2\beta_2 + x_1x_2\beta_3 + e.$$

Then $\frac{\partial}{\partial x_1} m(x_1, x_2) = \beta_1 + x_2\beta_3$, which is nonlinear in x_1. When we talk about the linear function here, we only require that the function is linear in the parameter.

We are particularly interested in the particular β^* such that the *projection error* e is orthogonal to x. Such a model is called a *linear projection model*.

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

When it is clear in the context that we are talking about the linear projection model, we often drop out the star on β for a concise notation.

Eq.(2) implies that, if a constant is included in x, we have $E[e] = 0$. With no constant, this is not necessarily true. Moreover, when $E[e] = 0$, we have $\text{cov}(x, e) = E[xe] = 0$ so that x and e are uncorrelated. However, in general we cannot claim $cov(h(x), e)$ for a general function $h(\cdot)$.

The coefficient β^* in the linear projection model has a straightforward closed-form. When we multiply x on both sides and note $E[xe] = 0$, we have $E[xy] = E[xx']\beta^*$. If $E[xx']$ is invertible, we explicitly solve

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

We are interested in β^*, as is the *linear* minimizer of the MSE. That is,

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

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

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

Proof. If we replace y in (4) by $m(x)$, we have

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

Therefore β^* is also the minimizer of $E\left[(m(x) - x'\beta)^2\right]$, 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 of 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 indicates the effect of these explanatory factors. This such a regression

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

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

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

so that

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

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

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

respectively. This is a special case of a 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

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

Let $A_{11\cdot 2} = E[x_1 x_1'] - E[x_1 x_2'](E[x_2 x_2'])^{-1} E[x_2 x_1']$, and $A_{1y\cdot 2} = E[x_1 y] - E[x_1 x_2'](E[x_2 x_2'])^{-1} E[x_2 y]$ then $\beta_1 = A_{11\cdot 2}^{-1} A_{1y\cdot 2}$. 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 $u = x_1 - x_2'\gamma = x_1 - x_2'(E[x_2 x_2'])^{-1} E[x_2 x_1']$. We then run a regression of y on u with a constant, so that

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

The nominator $cov(u,y) = E[x_1 y] - E[x_1 x_2'](E[x_2 x_2'])^{-1} E[x_1 y]$. The denominator $var(u) = E\left[u^2\right] = A_{11\cdot 2}$.

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.

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

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