Quantitative Methods in Economics
Linear Predictor

Maximilian Kasy

Harvard University, fall 2016

Roadmap, Part I

1. **Linear predictors and least squares regression**
2. Conditional expectations
3. Some functional forms for linear regression
4. Regression with controls and residual regression
5. Panel data and generalized least squares

Takeaways for these slides

- Best linear predictor minimizes average squared prediction error (population concept)
- Least squares regression minimizes mean squared residual (sample analog)
- First order condition for minimization $\leftrightarrow$ orthogonality conditions
- Geometric interpretation
- Omitted variables

A population prediction problem

- Suppose there are two random variables X and Y.
- You want to predict Y using a linear function of X.
- What's the best predictor, if you know the joint distribution of X and Y?

Some definitions

- Best linear Predictor:

$$\hat{Y} = \beta_0 + \beta_1 X, \quad \min_{\beta_0, \beta_1} E(Y - \hat{Y})^2$$

- Inner Product:

$$\langle Y, X \rangle = E(YX)$$

- Norm:

$$||Y|| = \langle Y, Y \rangle^{1/2}, \quad \min_{\beta_0, \beta_1} ||Y - \hat{Y}||^2$$

Questions for you

Recall the definition of an inner product.

- What properties does it have?
- What does that imply for a norm?

Questions for you

- Write the objective function as an inner product.
- Find the first order conditions for the minimization problem.

The solution as an orthogonal projection

- Orthogonal ($\perp$) Projection

$$\langle Y - \hat{Y}, 1\rangle = 0, \quad \langle Y - \hat{Y}, X\rangle = 0$$

- Substitute:

$$\langle Y - \beta_0 1 - \beta_1 X, 1\rangle = \langle Y, 1\rangle - \beta_0\langle 1, 1\rangle - \beta_1\langle X, 1\rangle = 0,$$
$$\langle Y - \beta_0 1 - \beta_1 X, X\rangle = \langle Y, X\rangle - \beta_0\langle 1, X\rangle - \beta_1\langle X, X\rangle = 0$$

- Use definition of $\langle\cdot,\cdot\rangle$:

$$E(Y) - \beta_0 - \beta_1 E(X) = 0, \quad E(YX) - \beta_0 E(X) - \beta_1 E(X^2) = 0$$

Questions for you

Express β_1 in terms of Covariances / Variances.

- Solution:

$$\beta_0 = E(Y) - \beta_1 E(X), \quad \beta_1 = \frac{E(YX) - E(Y)E(X)}{E(X^2) - E(X)E(X)} = \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)}$$

- Notation:

$$E^*(Y \mid 1, X) = \beta_0 + \beta_1 X$$

Least squares fit

- So far, we assumed that the joint distribution of X and Y is known – in particular their variance and covariance.
- What do we get if we take the joint distribution in a sample, rather than in the population?
- Data:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad x_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Questions for you

Define a sample minimization problem analogous to the population problem we just considered.

- Sample prediction problem:

$$\hat{y}_i = b_0 + b_1 x_i$$

$$\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- Inner Product:

$$\langle y, x \rangle = \frac{1}{n} \sum_{i=1}^{n} y_i x_i$$

- Minimum Norm:

$$\min_{b_0, b_1} ||y - b_0 x_0 - b_1 x||^2$$

Questions for you

First order condition?

- As before: Orthogonality conditions

$$\langle y - \hat{y}, x_0 \rangle = 0, \quad \langle y - \hat{y}, x \rangle = 0$$

- Use definition of $\langle \cdot, \cdot \rangle$:

$$\bar{y} - b_0 - b_1 \bar{x} = 0, \quad \overline{yx} - b_0 \bar{x} - b_1 \overline{x^2} = 0,$$

- where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad \overline{yx} = \frac{1}{n} \sum_{i=1}^{n} y_i x_i, \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

- Solve:

$$b_0 = \bar{y} - b_1 \bar{x}, \quad b_1 = \frac{\overline{yx} - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}\bar{x}}$$

Goodness of fit

- Back to the population problem.
- How well does X predict Y?
- Average squared prediction error:

$$||Y - E^*(Y|1,X)||^2 \leq ||Y - E^*(Y|1)||^2 = ||Y - E(Y)||^2$$

Questions for you

Why does the inequality hold?

- Goodness of fit:

$$1 - R^2_{\text{pop}} = \frac{||Y - E^*(Y|1,X)||^2}{||Y - E(Y)||^2}, \quad 0 \leq R^2_{\text{pop}} \leq 1$$

- Sample Analog:

$$R^2 = 1 - \frac{||y - (\hat{y} \,|\, 1, x)||^2}{||y - \bar{y}||^2}, \quad 0 \leq R^2 \leq 1$$

A clasic empirical example

- Jacob Mincer, *Schooling, Experience and Earnings*, 1974, Table 5.1;
- 1 in 1000 sample, 1960 census; 1959 annual earnings; $n = 31093$
- y = log(earnings), s = years of schooling
- Fitted regression:

$$\hat{y} = 7.58 + .070s, \quad R^2 = .067$$

Omitted variables

- What if we are actually interested in a different prediction problem?
- How do earnings vary as we vary education, holding constant predetermined IQ?
- Y = earnings, X_1 = education, X_2 = IQ

$$E^*(Y \mid 1, X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (Long)$$

$$E^*(Y \mid 1, X_1) = \alpha_0 + \alpha_1 X_1 \qquad (Short)$$

$$E^*(X_2 \mid 1, X_1) = \gamma_0 + \gamma_1 X_1 \qquad (Aux)$$

- Define prediction error:

$$U \equiv Y - E^*(Y \mid 1, X_1, X_2),$$

- so that

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U \quad \text{with} \quad U \perp 1, \quad U \perp X_1, \quad U \perp X_2$$

Questions for you

- Substitute this formula for Y in the short regression $E^*(Y \mid 1, X_1)$.
- Use the fact that the best linear predictor is linear in Y to split this into parts.
- Use the fact that U has to be orthogonal to 1 and X_1.
- Derive a formula for α in terms of β and γ.

Solution

- Substitute:

$$\begin{aligned}E^*(Y \mid 1, X_1) &= \beta_0 + \beta_1 X_1 + \beta_2 E^*(X_2 \mid 1, X_1) + E^*(U \mid 1, X_1) \\ &= \beta_0 + \beta_1 X_1 + \beta_2(\gamma_0 + \gamma_1 X_1) \\ &= (\beta_0 + \beta_2 \gamma_0) + (\beta_1 + \beta_2 \gamma_1) X_1 .\end{aligned}$$

- Omitted Variable Bias Formula:

$$\alpha_1 = \beta_1 + \beta_2 \gamma_1$$

Sample analog

- Least squares version:

$$\begin{aligned} \hat{y}_i \,|\, 1, x_{i1}, x_{i2} &= b_0 + b_1 x_{i1} + b_2 x_{i2}, && (Long) \\ \hat{y}_i \,|\, 1, x_{i1} &= a_0 + a_1 x_{i1}, && (Short) \\ \hat{x}_{i2} \,|\, 1, x_{i1} &= c_0 + c_1 x_{i1}. && (Aux) \end{aligned}$$

- Omitted Variable Bias Formula for sample:

$$a_1 = b_1 + b_2 c_1$$

Empirical example continued

- Griliches and Mason, "Education, Income, and Ability," *Journal of Political Economy*, 1972;
- CPS subsample of veterans, age 21–34 in 1964; y = log of usual weekly earnings; ST = total years of schooling; SB = schooling before army; SI = schooling after army ($ST = SB + SI$); $AFQT$ = armed forces qualification test (percentile)
- Table 3:

$$\hat{y} = .0508 ST + \dots$$

$$\hat{y} = .0433 ST + .00150 AFQT + \dots$$

$$\hat{y} = .0502 SB + .0528 SI + \dots$$

$$\hat{y} = .0418 SB + .0475 SI + .00154 AFQT + \dots$$