LECTURE NOTE 3

RESIDUAL REGRESSION, OMITTED VARIABLES, AND A MATRIX VERSION

1. RESIDUAL REGRESSION

Consider the linear predictor with a general list of K predictor variables (plus a constant):

$$E^*(Y \mid 1, X_1, \ldots, X_K) = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K. \tag{1}$$

We are going to develop a formula for a single coefficient, which, for convenience, will be β_K. Our result will use the linear predictor of X_K given the other predictor variables:

$$E^*(X_K \mid 1, X_1, \ldots, X_{K-1}) = \gamma_0 + \gamma_1 X_1 + \cdots + \gamma_{K-1} X_{K-1}.$$

Define $\tilde{X}_K$ as the residual (prediction error) from this linear predictor:

$$\tilde{X}_K = X_K - E^*(X_K \mid 1, X_1, \ldots, X_{K-1}).$$

The result is that β_K is the coefficient on $\tilde{X}_K$ in the linear predictor of Y given just $\tilde{X}_K$:

Claim 1. $E^*(Y \mid \tilde{X}_K) = \beta_K \tilde{X}_K$ with $\beta_K = E(Y\tilde{X}_K)/E(\tilde{X}_K^2)$.

Proof. Substitute

$$X_K = \gamma_0 + \gamma_1 X_1 + \cdots + \gamma_{K-1} X_{K-1} + \tilde{X}_K$$

into (1) to obtain

$$\begin{aligned} E^*(Y \mid 1, X_1, \ldots, X_K) &= \beta_0 + \beta_1 X_1 + \cdots + \beta_{K-1} X_{K-1} \\ &\quad + \beta_K(\gamma_0 + \gamma_1 X_1 + \cdots + \gamma_{K-1} X_{K-1} + \tilde{X}_K) \\ &= \tilde{\beta}_0 + \tilde{\beta}_1 X_1 + \cdots + \tilde{\beta}_{K-1} X_{K-1} + \beta_K \tilde{X}_K, \end{aligned} \tag{2}$$

with

$$\tilde{\beta}_j = \beta_j + \beta_K \gamma_j \qquad (j = 0, 1, \ldots, K-1). \tag{3}$$

The residual from predicting Y must be orthogonal to $1, X_1, \ldots, X_K$. Since $\tilde{X}_K$ is a linear combination of $1, X_1, \ldots, X_K$, we must have $\tilde{X}_k$ orthogonal to $Y - E^*(Y \,|\, 1, X_1, \ldots, X_K)$:

$$\langle Y - \tilde{\beta}_0 - \tilde{\beta}_1 X_1 - \cdots - \tilde{\beta}_{K-1} X_{K-1} - \beta_K \tilde{X}_K, \tilde{X}_K \rangle = 0. \tag{4}$$

Since $\tilde{X}_K$ is the residual from a prediction based on $1, X_1 \ldots, X_{K-1}$, it is orthogonal to those variables, and (4) reduces to

$$\langle Y - \beta_K \tilde{X}_K, \tilde{X}_K \rangle = \langle Y, \tilde{X}_K \rangle - \beta_K \langle \tilde{X}_K, \tilde{X}_K \rangle = 0.$$

So $\beta_K \tilde{X}_K$ is the orthogonal projection of Y on $\tilde{X}_K$ and

$$\beta_K = \langle Y, \tilde{X}_K \rangle / \langle \tilde{X}_K, \tilde{X}_K \rangle = E(Y \tilde{X}_K)/E(\tilde{X}_K^2). \quad \diamond$$

This population result has a sample counterpart. The data are in the matrices

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad x_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix} \qquad (j = 0, 1, \ldots, K).$$

Consider the least-squares fit using the K predictor variables (and a constant):

$$\hat{y}_i \,|\, 1, x_1, \ldots, x_K = b_0 + b_1 x_{i1} + \ldots + b_K x_{iK}.$$

Claim 2 provides a formula for a single coefficient, such as b_K. The result uses the least-squares fit of x_K on the other predictor variables:

$$\hat{x}_{iK} \,|\, 1, x_1, \ldots, x_{K-1} = c_0 + c_1 x_{i1} + \cdots + c_{K-1} x_{i,K-1}.$$

Define $\tilde{x}_K$ as the residual from this least-squares fit:

$$\tilde{x}_{iK} = x_{iK} - (\hat{x}_{iK} \,|\, 1, x_1, \ldots, x_{K-1}).$$

Then b_K is the coefficient on $\tilde{x}_K$ in the least squares fit of y on just $\tilde{x}_K$:

Claim 2. $(\hat{y}_i \,|\, \tilde{x}_K) = b_k \tilde{x}_{iK}$ with $b_K = \frac{1}{n}\sum_{i=1}^n y_i \tilde{x}_{iK} \Big/ \frac{1}{n}\sum_{i=1}^n \tilde{x}_{iK}^2$.

The proof is the same as for claim 1, with the least-squares inner product $\langle y, x_j \rangle = \sum_{i=1}^n y_i x_{ij}/n$ replacing the linear predictor (or mean-square) inner product $\langle Y, X_j \rangle = E(YX_j)$.

2. OMITTED VARIABLES

This section derives the general version, with K predictor variables, of the omitted variable formula in claim 1 of Note 1. We shall use the notation (and part of the argument) from the residual regression result in Section 1. The short linear predictor is

$$E^*(Y \,|\, 1, X_1, \ldots, X_{K-1}) = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_{K-1} X_{K-1}.$$

Claim 3. $\alpha_j = \beta_j + \beta_K \gamma_j \qquad (j = 0, 1, \ldots, K-1)$.

Proof. Let U denote the following prediction error:

$$U \equiv Y - E^*(Y \,|\, 1, X_1, \ldots, X_K).$$

Use equation (2) to write

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 X_1 + \cdots + \tilde{\beta}_{K-1} X_{K-1} + \beta_K \tilde{X}_K + U,$$

with (from (3)) $\tilde{\beta}_j = \beta_j + \beta_K \gamma_j$. Note that for $j = 0, 1, \ldots, K-1$,

$$\langle Y - \tilde{\beta}_0 - \tilde{\beta}_1 X_1 - \cdots - \tilde{\beta}_{K-1} X_{K-1}, X_j \rangle = \langle \beta_K \tilde{X}_K + U, X_j \rangle = 0.$$

These orthogonality conditions characterize the short linear predictor, and so $\alpha_j = \tilde{\beta}_j$. $\diamond$

The sample counterpart of this result uses the short least-squares fit:

$$\hat{y}_i \,|\, 1, x_1, \ldots, x_{K-1} = a_0 + a_1 x_{i1} + \cdots + a_{K-1} x_{i,K-1}.$$

Claim 4. $a_j = b_j + b_K c_j \qquad (j = 0, 1, \ldots, K-1)$.

This least-squares version of the omitted variable bias formula is a computational identity, which can be checked on a data set using a least-squares computer program.

3. MATRIX VERSION OF LINEAR PREDICTOR AND LEAST-SQUARES FIT

Set up the following $(K+1) \times 1$ matrices:

$$X = \begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_K \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}.$$

The linear predictor coefficients β_j are determined by the following orthogonality conditions:

$$\langle Y - \beta_0 - \beta_1 X_1 - \cdots - \beta_K X_K, X_j \rangle = 0 \qquad (j = 0, 1, \ldots, K).$$

So

$$E[(Y - X'\beta)X_j] = E[X_j(Y - X'\beta)] = 0 \qquad (j = 0, 1, \ldots, K).$$

We can write all the orthogonality conditions together as

$$E[X(Y - X'\beta)] = 0.$$

This gives the following system of linear equations:

$$E(XY) - E(XX')\beta = 0,$$

which has the solution

$$\beta = [E(XX')]^{-1} E(XY)$$

(provided that the $(K+1) \times (K+1)$ matrix $E(XX')$ is nonsingular).

For the least-squares fit, set up the $(K+1) \times 1$ matrices

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad x_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix} \quad (j = 0, 1, \ldots, K), \quad b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_K \end{pmatrix},$$

and the $n \times (K+1)$ matrix

$$x = \begin{pmatrix} x_0 & x_1 & \dots & x_K \end{pmatrix} = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1K} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & \dots & x_{nK} \end{pmatrix}.$$

The least-squares coefficients b_j are determined by the following orthogonality conditions:

$$\langle y - b_0 x_0 - b_1 x_1 - \cdots - b_K x_K, x_j \rangle = 0 \qquad (j = 0, 1, \dots, K).$$

So

$$(y - xb)' x_j = x_j'(y - xb) = 0 \qquad (j = 0, 1, \dots, K).$$

We can write all the orthogonality conditions together as

$$\begin{pmatrix} x_0' \\ x_1' \\ \vdots \\ x_K' \end{pmatrix} (y - xb) = x'(y - xb) = 0.$$

This gives the following system of linear equations:

$$x'y - x'xb = 0,$$

which has the solution

$$b = (x'x)^{-1} x'y$$

(provided that the $(K+1) \times (K+1)$ matrix $x'x$ is nonsingular).