2002.6 Regression - Matrix Treatment

Now it gets good

Matrix Notation for Regressions

We can write a regression model in matrix form as follows:

$$Y = X\beta + u$$

Here, Y is an $N \times 1$ vector of outcomes. X is an $N \times k$ vector of covariates. u is a $N \times 1$ vector of errors.

We approximate this with

$$Y = X\hat{\beta} + \hat{u}$$

OK, so what does this look like inside? Let's do a small example, with $n = 5$ and $k = 3$:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \\ x_{4,1} & x_{4,2} & x_{4,3} \\ x_{5,1} & x_{5,2} & x_{5,3} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{bmatrix}$$

Let's multiply out only the first row of $X\beta$:

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \hat{\beta}_1 x_{1,1} + \hat{\beta}_2 x_{1,2} + \hat{\beta}_3 x_{1,3}$$

We're going to end up with a very familiar system of linear equations: $y_i = \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \hat{\beta}_3 x_{3,i} + \hat{u}_i$. Note here that the *row* indicators for X became *columns* of our system.

If you're worried where the intercept went, worry not. $x_{i,1}$ is usually 1 by convention, which makes $\hat{\beta}_1$ our constant term.

Regression Estimator

Three important matrix differentiation results:

1. $\frac{d}{dx}a'x = \frac{d}{dx}x'a = a$
2. $\frac{d}{dx'}a'x = a'$
3. $\frac{d}{dx}x'Ax = 2Ax.$

See Matrix Calculus for detail; and recall Matrix Properties.

We want to minimize the sum of $\hat{u}_i$. Fortunately, we can represent that with $\hat{u} \cdot \hat{u}$ - a dot product - or equivalently $\hat{u}'\hat{u}$. So:

$$\begin{aligned}\hat{u}'\hat{u} &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\ &= (y' - \hat{\beta}'X')(y - X\hat{\beta}) \\ &= y'y - 2y'X\hat{\beta} - \hat{\beta}'X'X\hat{\beta}\end{aligned}$$

Now we can start using the matrix-calculus results to differentiate the minimand with respect to $\hat{\beta}$. The first term drops away. For the second term, note that $y \times X$ is a $1 \times k$ vector, so we can use the first result: the derivative is $-X'y$. The third term: $X'X$ is symmetric. So this is the quadratic form (result 3): $2X'X\hat{\beta}$.

This gives us the first-order condition to solve:

$$\begin{aligned}0 &= -2X'y - 2X'X\hat{\beta} \\ \hat{\beta} &= (X'X)^{-1}X'y\end{aligned}$$

The requirement that $X'X$ be invertible means that X must be invertible - i.e. that it have no collinearity.