Math 455 Class 27 March 27

Chapter 8 : Problems with the error.

8.1 Generalized Least Squares and

8.2 Weighted Least Squares : a special case of Generalized Least Squares.

The standard assumptions of linear regression require the variance - covariance matrix of the errors $Var(\varepsilon)$ to be $\sigma^2 I$, that is :

$$Var(\varepsilon_i) = \sigma^2 \text{ and } Cov(\varepsilon_i, \varepsilon_j) = 0.$$

Generalized least Squares replaces this assumption with.

$$Var(\varepsilon) = \sigma^2 \Sigma \quad \text{where } \Sigma \text{ is known.}$$

Note : Σ must be symmetric and positive definite.

Special Case : Weighted Least squares.

$$Var(\varepsilon) = \begin{pmatrix} 1/w_1 & & O \\ & \ddots & \\ O & & 1/w_n \end{pmatrix}$$

w_i are the "weights".

These methods require that we know, or at least be able to estimate, the Σ in $Var(\varepsilon) = \sigma^2 \Sigma$.

Alternative: Heteroskedasticity - Consistent (HC) standard errors (frequently used in econometrics).

Not treated in Faraway.

Note: many other techniques popular in econometrics are also not treated, e.g. instrumental variables

---

The big insight of 8.1 & 8.2:

The GLS problem can be reduced to the OLS problem (which we have already solved) by linear algebra:

Since Σ is positive definite, $\Sigma = SS^T$, where S is triangular. (Cholesky Decomp.)

Reference: Axler, Linear Algebra Done Right (4^{th} ed)
Thm 7.63, p. 267

How do we do this? Answer: Faraway p. 113.

Our problem: (in matrix notation):

$$Y = X\beta + \varepsilon \quad \text{where} \quad Var(\varepsilon) = \sigma^2 \Sigma.$$

Estimate $\hat{\beta}$.

To solve, transform by multiplying by S^{-1}.

$$S^{-1} Y = S^{-1} X \beta + S^{-1} \varepsilon$$

Rewrite:

$$Y' = X'\beta + \varepsilon'$$

where $Y' = S^{-1} Y$, $X' = S^{-1} X$, $\varepsilon' = S^{-1} \varepsilon$

can all be computed, because we know Σ, and can find S such that $SS^T = \Sigma$.

This ε' satisfies

$$Var(\varepsilon') = Var(S^{-1}\varepsilon) = S^{-1} Var(\varepsilon) (S^{-1})^T$$

see, e.g. Seber & Lee Ch. 1

$$= S^{-1} \sigma^2 SS^T (S^T)^{-1} = S^{-1}(\sigma^2 I) S$$

$$= \sigma^2 I S^{-1} S = \sigma^2 I$$

Note: we used a fact from linear algebra

A scalar multiple of I commutes with any matrix.

This is <u>not</u> true of a matrix $\begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}$.

We know the solution to $Y' = X'\beta + \varepsilon'$,

because this now satisfies the axioms of linear regression.

Ordinarily we write: $\hat{\beta} = (X^T X)^{-1} X^T Y$.

Here we put primes on everything and translate:

$$\hat{\beta} = \left((X')^T X'\right)^{-1} (X')^T Y'$$

$$\hat{\beta} = \left((S^{-1}X)^T S^{-1}X\right)^{-1} (S^{-1}X)^T S^{-1}Y$$

$$= \left(X^T (S^{-1})^T S^{-1} X\right)^{-1} X^T (S^{-1})^T S^{-1} Y$$

$$= \left(X^T \Sigma^{-1} X\right)^{-1} X^T \Sigma^{-1} Y$$

since $(S^{-1})^T S^{-1}$
$= (SS^T)^{-1}$
$= \Sigma^{-1}$

Note: Knowledge of Σ changes our estimate $\hat{\beta}$, not just the standard errors of the $\hat{\beta}_i$.

This is distinct from the HC methods of econometrics, which does NOT change the OLS estimate $\hat{\beta}$.

The special case is Weighted Least Squares: Σ is diagonal.

Some special cases in which WLS might be appropriate:

(1) We observe $Var(\varepsilon_i) \propto X_i$

(is proportional to.)

after making a plot of $|\hat{\varepsilon}_i|$ vs X_i.

Weights: $\frac{1}{X_i}$.

(2) Y_i are the averages of n_i observations.

Then $Var(Y_i) = Var(\varepsilon_i) = \frac{\sigma^2}{n_i}$.

Weights: n_i

(3) $Var(Y_i)$ is (somehow) known

$$W_i = \frac{1}{Var(Y_i)} \quad \text{(Context-Dependent)}$$

Remark on Linear Algebra: if you only remember one fact from Linear Algebra it should be:

Symmetric Matrices can be Diagonalized.

Positive semidefinite: A satisfies $x^T A x \geq 0$.

Define $\langle x, y \rangle_A$ by $\langle x, y \rangle_A = x^T A y$.

This is bilinear. But need not satisfy other axioms of inner product. For example, the ordinary inner product satisfies $\langle x, x \rangle = \|x\|^2$ and $\langle x, x \rangle = 0$ iff $x = 0$.

If we make the weaker requirement that

$\langle x, x \rangle_A \overset{def}{=} x^T A x \geq 0$, this is positive semi-def.