Correcting for heteroskedasticity

January 11, 2014

Allowing non-scalar covariance matrix of residuals

- Instead of $\mathrm{Var}(Y \mid X) = \sigma^2 I$, $\mathrm{Var}(Y \mid X) = \Omega$.
- Then the maximum likelihood estimator for β, which is (like OLS) unbiased, is the **generalized least squares** (GLS) estimator

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.$$

- Problem: When $\Omega = \sigma^2 I$, the unknown σ^2's cancel out of this formula. But with any other Ω matrix, there is no cancellation. And we generally do not know Ω.

Estimating Ω

- Just treating all elements of Ω as unknown parameters and using a flat prior will not be helpful.

- Ω has n^2 elements, and symmetry just reduces the number to $(n^2+n)/2$. For $n > 1$, this exceeds n, and usually trying to estimate more parameters than you have data points does not work, unless you are willing to rely on a proper prior.

- So far we are considering i.i.d. cases, though, which implies that Ω will be diagonal, though not scalar.

Weighted least squares

If we knew the diagonal elements of Ω, $\sigma_i^2, i = 1, \ldots, n$, the GLS formula could be written as

$$(\sum_i X_i' X_i / \sigma_i^2)^{-1} \sum_i X_i' y_i / \sigma_i^2 \ .$$

One way to calculate this is to multiply each row of the X matrix and of the y column vector by the corresponding $1/\sigma_i$ (not $1/\sigma_i^2$), then apply the usual OLS formula. That is, use OLS on data weighted by the inverse of the standard errors of the residuals.

Instead of $(n^2 + n)/2$ unknown parameters in Ω, we now have just n — the σ_i^2's. But that's still too many to estimate from a sample of size n.

Heteroskedasticity that does not depend on X doesn't matter, asymptotically

If $\varepsilon_i \mid \{X, \nu^2\} \sim N(0, \nu^2)$, with ν^2 itself random and independent of X, then the distribution of ε_i, integrating out the unobserved and unknown ν^2, is

$$\int p(\nu^2) \frac{1}{\nu\sqrt{2\pi}} e^{-\frac{\varepsilon_i^2}{\nu^2}} d\nu \,.$$

This will still have mean zero, and if the distribution of ν is not too fat-tailed, finite variance. So the usual SNLM distribution theory for OLS is still justified as a large sample approximation. The scalar covariance matrix assumption is actually correct in this case.

How to estimate σ_i^2

- The i.i.d. assumption tells us the distribution of ε_i can depend ony on X_i, not on X_j for $j \neq i$. So only the diagonal of Ω is non-zero.
- We need a model, with fewer than n parameters, for $E[u_i^2 \mid X_i]$.
- One approach: Form the predicted values from an initial OLS estimate,

$$\hat{y} = X\hat{\beta}_{OLS}\ , \text{ estimate} \tag{1}$$

$$\hat{\varepsilon}_i^2 = (y_i - \hat{y}_i)^2 = \alpha_0 + \alpha_1 \hat{y}_i + \xi_i\ , \tag{2}$$

then use $\hat{\alpha}_0 + \hat{\alpha}_1 \hat{y}_i$ in place of σ_i^2 in the GLS formula.

Another approach

Form $\Omega(\alpha_0, \alpha_1, X)$ by putting $\alpha_0 + X_i\alpha_1$ on the diagonal of Ω, then maximize likelihood or form a posterior mean, using the full likelihood in all the unknown parameters α_0,α_1, and β.

$$(2\pi)^{-nk/2}\prod_{i=1}^{n}(\alpha_0 + X_i\alpha_1)^{-1/2} \cdot \exp\left(-\sum_{i=1}^{n}\frac{(y_i - X_i\beta)^2}{2(\alpha_0 + X_i\alpha_1)}\right) .$$

Of course one does not have to stick to a linear model for $E[u_i^2 \mid X_i]$. It could also be a polynomial, or one could make $|u_i|$ linear function of X_i, for example.