GEE derivation stuff

Saturday, November 25, 2023 17:16

Consider data:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma^2 & R & 0 \\ R & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}\right)$$

where y_1 and y_2 are correlated

Want to find weights $(\alpha_1, \alpha_2, \alpha_3) = W$ such that $\hat{\mu}$ is unbiased estimator

know: $E(y_1 - \mu) = 0$
$E(y_2 - \mu) = 0$
$E(y_3 - \mu) = 0$

Then $\alpha_1 (y_1 - \hat{\mu}) + \alpha_2 (y_2 - \hat{\mu}) + \alpha_3 (y_3 - \hat{\mu}) = 0$

$$\hat{\mu} = \frac{\alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3}{\alpha_1 + \alpha_2 + \alpha_3}$$

$$w_i = \frac{\alpha_i}{\alpha_1 + \alpha_2 + \alpha_3}$$

Regression w/ heteroskedastic errors:

$y = X\beta + \epsilon$

$\epsilon \sim N(0, V)$ $\quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$

Then, residual covariance matrix is

$$V = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & v_n \end{bmatrix}$$

Then, define W as the inverse:

$$W = \begin{bmatrix} v_1^{-1/2} & & & 0 \\ \vdots & v_2^{-1/2} & & \\ & & \ddots & \\ 0 & & & v_3^{-1/2} \end{bmatrix}$$

aka $diag(W) = (v_1^{-1/2}, v_2^{-1/2}, \ldots, v_n^{-1/2})$

Then $WW = V^{-1}$

Then, from $y^* = Wy$ we get:

$$Cov(y^*) = Cov(Wy)$$
$$W\, cov(y^*)\, W = WVW$$
$$= (WV^{1/2})(V^{1/2}W)$$
$$= I \cdot I$$
$$= I$$

Then, $Var(y_i^*) = Var(w_i y_i)$
$= w_i^2 v_i$
$= v_i^{-1} v_i$
$= 1$

diagonals equal to 1

$E(y^*) = E[WX\beta + W\epsilon]$
$= WX\beta + 0$
$= WX\beta$

Weighted Least Squares (when you have heteroskedastic errors)

Let $X^* = WX$

$y^* = X^* \beta + \epsilon^*$ $\quad \epsilon^* \sim N(0, I)$

$Wy = WX\beta + W\epsilon$

Throwback: OLS: $\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$ — same thing but with weights

$$\hat{\beta}_{WLS} = (X^{*\prime} X^*)^{-1} X^{*\prime} y^*$$
$$= ((WX)'(WX))^{-1} (WX)'(Wy)$$
$$= [X'W'WX]^{-1} X'W'Wy$$
$$= [X'V^{-1}X]^{-1} X'V^{-1}y$$

$$Cov(\hat{\beta}_{WLS}) = Cov([X'V^{-1}X]^{-1} X'V^{-1}y)$$
$$= [X'V^{-1}X]^{-1} X'V^{-1}\, Cov(y)\, [[X'V^{-1}X]^{-1} X'V^{-1}]'$$
$$= [X'V^{-1}X]^{-1} X'\cancel{V^{-1}}\, \cancel{V}\, [V^{-1}X[X'V^{-1}X]^{-1}]$$
$$= [X'V^{-1}X]^{-1} \cancel{X'V^{-1}X[X'V^{-1}X]^{-1}}$$
$$= [X'V^{-1}X]^{-1}$$

Example: Fixed Effect Meta Analysis

Let $\hat{\beta}_i$ be the effect size from study i with estimated variance $\hat{\sigma}_i^2$. We will assume the model

$$\hat{\beta}_i = \mu + \epsilon_i \qquad \epsilon_i \sim N(0, \hat{\sigma}_i^2).$$

This can be expressed as

$$y = [\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n]'$$
$$X = [1, 1, \ldots, 1]' \quad \text{(i.e., intercept only)}$$
$$V = diag(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_n^2)$$

Therefore
- $V^{-1} = diag(1/\hat{\sigma}_1^2, 1/\hat{\sigma}_2^2, \ldots, 1/\hat{\sigma}_n^2)$
- $X'V^{-1} = [1/\hat{\sigma}_1^2, 1/\hat{\sigma}_2^2, \ldots, 1/\hat{\sigma}_n^2]'$
- $X'V^{-1}X = \sum_{i=1}^n 1/\hat{\sigma}_i^2$
- $X'V^{-1}y = \sum_{i=1}^n \hat{\beta}_i/\hat{\sigma}_i^2$

all the weights sum to 1

inverse variance weight

$$\mu_{wls} = (X'V^{-1}X)^{-1} X'V^{-1}y = \frac{\sum_{i=1}^n \hat{\beta}_i/\hat{\sigma}_i^2}{\sum_{i=1}^n 1/\hat{\sigma}_i^2}$$

... the inverse-variance weighted average!

Generalized Least Squares

V is non-diagonal (so there's both heteroskedasticity & correlation between obs)

$\hat{\beta}_{GLS} = [X'V^{-1}X]^{-1} X'V^{-1}y$

$Cov(\hat{\beta}_{GLS}) = [X'V^{-1}X]^{-1}$

If we know V, the weighting scheme will give us the smallest variances among the unbiased estimators

If homoscedastic but still corr. between obs:

$V = \underbrace{\sigma^2}_{\text{common residual variance}} R$ $\quad R = \begin{bmatrix} 1 & & \\ \rho_{12} & 1 & \\ \rho_{13} & \cdots & 1 \end{bmatrix}$

$\hat{\beta}_{GLS} = (X'R^{-1}X)^{-1} X'R^{-1}y$ — to derive: use $\sigma^{-2}R^{-1} = V^{-1}$

NAIVE $[Cov(\hat{\beta}_{GLS}) = \sigma^2 (X'R^{-1}X)^{-1}$

Robust Regression

Don't know V. $\quad y = X\beta + \epsilon$

$\hat{\beta}_{OLS} = (X'X)^{-1} X'y$ still unbiased & consistent

$$\widehat{Cov}(\hat{\beta}_{OLS})_{robust} = \underbrace{(X'X)^{-1}X'}_{\text{Bread}} \underbrace{([y - X\hat{\beta}_{OLS}][y - X\hat{\beta}_{OLS}]')}_{\text{Meat}} \underbrace{X(X'X)^{-1}}_{\text{Bread}}$$

$$\underbrace{\hat{\epsilon} \cdot \hat{\epsilon}'}_{\hat{V}}$$

Ideal conditions:
- balanced data
- little missing — converges to $\widehat{Cov}(\hat{\beta}_{OLS})$ for $\epsilon \sim N(0, V)$
- # obs per subject << # subjects

Tradeoffs in modeling

When assumptions are true: → lower mean squared error
- stronger assumptions = more accurate estimates. → fewer type II error
- weaker assumptions = less accurate if there is more structure to exploit.

When assumptions are false:
- stronger assumptions = incorrect inference. → more type I error
- weaker assumptions = more robust.

For example, we see this in non-parametric tests, which are generally less powerful when parametric assumptions hold.

E.g. two-sample t test more powerful than Wilcoxon rank sum when data is normally distributed, but if it is not, t-test ↑ type 1 error

ON WHY NO OLS IN THIS CASE (heteroscedasticity)

Weighted Least Squares

The weighted least squares estimate is unbiased (for any V).

$$E(\hat{\beta}_{wls}) = E(X'V^{-1}X)^{-1}X'V^{-1}y)$$
$$= (X'V^{-1}X)^{-1}X'V^{-1}E(y)$$
$$= (X'V^{-1}X)^{-1}X'V^{-1}X\beta = \beta.$$

What if we used the regular ordinary least squares estimate?

$$E(\hat{\beta}_{ols}) = E[(X'X)^{-1}X'y]$$
$$= (X'X)^{-1}X'E(y)$$
$$= (X'X)^{-1}X'X\beta = \beta. \quad \text{Still unbiased!}$$

BUT....

Note on OLS estimator

The OLS estimator of β is unbiased. It is also consistent

→ true generating model i.e. if $y = X\beta + \epsilon$, $\epsilon \sim N(0, V)$

$$\lim_{n\to\infty} P\{|\hat{\beta}_{ols} - \beta| > \epsilon\} = 0.$$

$\hat{\beta}_{OLS}$ will converge to the true β

So what is the problem? (why r we worrying)

① inference: p-values can be wrong

② Efficiency: can define an estimator w/ lower variance

And, $Cov(\hat{\beta}_{OLS})$ doesn't hold anymore for weighted.

$$Cov(\hat{\beta}_{OLS}) = Cov((X'X)^{-1}X'y)$$
$$= (X'X)^{-1}X'\, Cov(y)\, X(X'X)^{-1}$$
$$= (X'X)^{-1}X'\, V\, X(X'X)^{-1}$$
$$\neq (X'X)^{-1}\hat{\sigma}^2_{OLS}$$

only works if V is the same everywhere.

Two issues with OLS when heteroscedasticity

The OLS estimate of the SE of $\hat{\beta}_{ols}$ can lead to invalid inference.

Under the null, $\hat{\beta}_k/\widehat{SE}_{ols}(\hat{\beta}_k)$ does not follow a t distribution.

Moreover, the true variance of $\hat{\beta}_{ols}$ is larger than $\hat{\beta}_{wls}$ – OLS is inefficient and WLS is more accurate.

Summarize:
1. OLS inference is invalid when the errors are not iid.
2. We can do a better job of estimating the coefficients with WLS.