ECOM40006/90013 Econometrics 3

Week 7 Extras: Solutions

Question 1: Ordinary Least Squares

(a) To derive the score for minimizing the sum of squared errors, it would help to first figure out what the gradient of $(y_i - x_i'\beta)^2$ is. This would be

$$\begin{aligned}\frac{\partial(y_i - x_i'\beta)^2}{\partial\beta} &= \frac{\partial(y_i - x_i'\beta)^2}{\partial(y_i - x_i'\beta)}\frac{\partial(y_i - x_i\beta)}{\partial x_i'\beta}\frac{\partial x_i'\beta}{\partial\beta} \\ &= (y_i - x_i'\beta) \times (-1) \times x_i \\ &= -2x_i(y_i - x_i'\beta)\end{aligned}$$

Note that while $y_i - x_i'\beta$ is a scalar, we move x_i to the left because it is not conformable for multiplication on the right. In this case, the score can then be derived as

$$\begin{aligned}s_n(\beta) &= \sum_{i=1}^{n}\frac{\partial(y_i - x_i'\beta)^2}{\partial\beta} \\ &= -2\sum_{i=1}^{n} x_i(y_i - x_i'\beta).\end{aligned}$$

(b) The exact notation for the first order condition satisfies

$$\sum_{i=1}^{n}\frac{\partial(y_i - x_i'\beta)^2}{\partial\beta}\bigg|_{\beta=\hat\beta} = 0.$$

The left hand side is often abbreviated as $s_n(\hat\beta)$ in score notation for maximum likelihood, which we will cover soon. Generally, the problem with going straight to this notation is that we first have to take a gradient, and then proceed to evaluate it at $\beta = \hat\beta$. That's not easy to conceptualize in one go, so informally we often just leave everything as β and then only evaluate it as $\hat\beta$ at the very end.

More pragmatically, it's only really recommended to write out this full notation if and only if a question specifically asks you to do so.

Going to the actual answer itself, observe that the residuals are denoted as

$$\hat u_i = y_i - x_i'\hat\beta,$$

and the derivative of the conditional mean function is

$$\begin{aligned}\left.\frac{\partial \mathbb{E}(y_i|x_i)}{\partial \beta}\right|_{\beta=\hat{\beta}} &= \left.\frac{\partial x_i'\beta}{\partial \beta}\right|_{\beta=\hat{\beta}} \\ &= x_i\big|_{\beta=\hat{\beta}} \\ &= x_i.\end{aligned}$$

Therefore the score evaluated at $\beta = \hat{\beta}$ gives

$$s_n(\hat{\beta}) = \sum_{i=1}^{n} \left.\frac{\partial x_i'\beta}{\partial \beta}\right|_{\beta=\hat{\beta}} (y_i - x_i\hat{\beta}) = \sum_{i=1}^{n} \text{deriv. of mean} \times \text{residuals}$$

(c) The FOC implies the OLS estimator takes the form

$$\begin{aligned} \sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}) &= 0 \\ \implies \sum_{i=1}^{n} x_i y_i - \left[\sum_{i=1}^{n} x_i x_i'\right]\hat{\beta} &= 0 \\ \implies \left[\sum_{i=1}^{n} x_i x_i'\right]\hat{\beta} &= \sum_{i=1}^{n} x_i y_i \\ \implies \left[\sum_{i=1}^{n} x_i x_i'\right]^{-1}\left[\sum_{i=1}^{n} x_i x_i'\right]\hat{\beta} &= \left[\sum_{i=1}^{n} x_i x_i'\right]^{-1}\sum_{i=1}^{n} x_i y_i \\ \implies \hat{\beta} &= \left[\sum_{i=1}^{n} x_i x_i'\right]^{-1}\sum_{i=1}^{n} x_i y_i. \end{aligned}$$

We'll use this again in later questions. Note that this is simply the individual vector notation for what we already know as the OLS estimator: $\hat{\beta} = (X'X)^{-1}X'y$.

Question 2: Weighted Least Squares

(a) The only difference between the assumptions in this question and in Question 1 is that the error term is heteroskedastic in some way. The derivation of the OLS estimator does not really need anything from the variance of the error term. So the derivation of the OLS estimator proceeds with exactly the same steps and the estimator itself remains exactly the same, i.e.

$$\hat{\beta} = \left[\sum_{i=1}^{n} x_i x_i'\right]^{-1}\sum_{i=1}^{n} x_i y_i.$$

(b) The conditional variance is computed directly as

$$\begin{aligned}\mathrm{Var}(u_i^*|x_i) &= \mathrm{Var}\left(\frac{u_i}{\sqrt{f(x_i)}}\bigg|x_i\right)\\ &= \frac{1}{f(x_i)}\mathrm{Var}(u_i|x_i)\\ &= \frac{f(x_i)}{f(x_i)}\\ &= 1\end{aligned}$$

where we use the fact that functions of x_i behave as constants when conditioned on x_i in the second line, allowing us to use the rule $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$ by treating $1/\sqrt{f(x_i)}$ as if it was a constant.

The result is a constant that does not depend on x_i, so the conditional variance of u_i^* does not exhibit heteroskedasticity at all. In fact, what this means is that if we were to regress y_i^* on x_i^*, we would end up with an OLS estimator where its standard errors do not suffer from heteroskedasticity.

(c) Substituting in y_i^*, x_i^* and u_i^* gives the modified OLS estimator

$$\begin{aligned}\hat{\beta} &= \left[\sum_{i=1}^{n} x_i^* x_i^{*\prime}\right]^{-1}\sum_{i=1}^{n} x_i^* y_i^*\\ &= \left[\sum_{i=1}^{n} \frac{x_i x_i'}{f(x_i)}\right]^{-1}\sum_{i=1}^{n} \frac{x_i y_i}{f(x_i)}.\end{aligned}$$

This specific estimator is known as the *weighted least squares* (WLS) estimator.

(d) The derivation is literally identical to how we got the score for the OLS estimator, but we replace y_i, x_i and u_i with y_i^*, x_i^* and u_i^* respectively. This means that the score can straight away be written as

$$s_n(\beta) = \sum_{i=1}^{n} x_i^*(y_i^* - x_i^{*\prime}\beta)$$

Evaluating this at $\hat{\beta}$ and substituting in our known expressions gives

$$\begin{aligned}s_n(\hat{\beta}) &= \sum_{i=1}^{n} \frac{x_i}{\sqrt{f(x_i)}}\left(\frac{y_i}{\sqrt{f(x_i)}} - \frac{x_i'}{\sqrt{f(x_i)}}\hat{\beta}\right)\\ &= \sum_{i=1}^{n} x_i \times \frac{y_i - x_i'\hat{\beta}}{f(x_i)}\\ &= \sum_{i=1}^{n} \text{deriv. of mean} \times \frac{\text{residuals}}{\text{variance weighting}}.\end{aligned}$$

So the moral of the story here is that when we weight the data using the conditional variance, we get an estimator that (i) doesn't have heteroskedasticity and (ii) has FOCs that look like the one that we just got above. It's on this basis that we can apply this intuition to places like maximum likelihood as well.

Question 3: The Feasible GLS Estimator

(a) Provided that the form of heteroskedasticity is known – i.e. $\sigma_i^2 = f(x_i)$ where $f(x_i)$ is known, we can define a diagonal matrix

$$\Omega^{-1} = \text{diag}\left(\frac{1}{f(x_1)}, \frac{1}{f(x_2)}, \ldots, \frac{1}{f(x_N)}\right)$$

such that $\Omega^{-1} = \Omega^{-1/2}\Omega^{-1/2}$ (the matrix $\Omega^{-1/2}$ is simply the matrix Ω^{-1} but with all diagonal elements square rooted). The GLS estimator can be obtained by pre-multiplying both sides of the linear model by $\Omega^{-1/2}$, namely:

$$y = X\beta + u \implies \Omega^{-1/2}y = \Omega^{-1/2}X\beta + \Omega^{-1/2}u \implies y^* = X^*\beta + u^*,$$

with definitions of y^*, X^* and u^* analogous to previous derivations in tutorial extras notes. OLS estimation on this transformed linear model gives the desired GLS estimator of

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

(b) The conditional variance of the FGLS estimator can be taken directly, observing that

$$\begin{aligned}\hat{\beta}_{FGLS} &= (X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}(X\beta + u) \\ &= \beta + (X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}u\end{aligned}$$

where the first expression, β, is considered a constant and is ignored in variance calculations. This gives us

$$\begin{aligned}\Omega_{FGLS} = \text{Var}(\hat{\beta}_{FGLS}|X) &= \text{Var}(\beta + (X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}u|X) \\ &= \text{Var}((X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}u|X) \\ &= (X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}\text{Var}(u|X)\hat{S}^{-1}X(X'\hat{S}^{-1}X)^{-1} \\ &= (X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}\Omega\hat{S}^{-1}X(X'\hat{S}^{-1}X)^{-1}\end{aligned}$$

Notice that this looks similar to something that we will find out soon...

(c) Using the variance decomposition formula we can write

$$\begin{aligned}\text{Var}(X'\hat{S}^{-1}u)^{-1} &= \mathbb{E}(\text{Var}(X'\hat{S}^{-1}u|X)) + \text{Var}(\mathbb{E}(X'\hat{S}^{-1}u|X)) \\ &= \mathbb{E}(X'\hat{S}^{-1}\text{Var}(u|X)\hat{S}^{-1}X) + \text{Var}(X'\hat{S}^{-1}\underbrace{\mathbb{E}(u|X)}_{=0}) \\ &= \mathbb{E}(X'\hat{S}^{-1}\Omega\hat{S}^{-1}X)\end{aligned}$$

as required.

(d) For the asymptotic distribution of the FGLS estimator, observe from before that we can write

$$\begin{aligned}
\hat{\beta}_{FGLS} &= \beta + (X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}u \\
\implies \hat{\beta}_{FGLS} - \beta &= (X'\hat{S}^{-1}X)^{-1}X'\hat{S}^{-1}u \times \frac{N}{N} \\
&= \left(\frac{1}{N}X'\hat{S}^{-1}X\right)^{-1} \frac{1}{N}X'\hat{S}^{-1}u \\
\implies \sqrt{N}(\hat{\beta}_{FGLS} - \beta) &= \left(\frac{1}{N}X'\hat{S}^{-1}X\right)^{-1} \frac{1}{\sqrt{N}}X'\hat{S}^{-1}u \\
&\xrightarrow{d} Q^{-1}N(0, P) \\
&= N(0, Q^{-1}PQ^{-1}),
\end{aligned}$$

as required.

(e) If the variance estimator $\hat{S}$ is consistent, then observe that we can write

$$P = \text{Var}(X'\hat{S}^{-1}u) = \mathbb{E}(X'\hat{S}^{-1}\Omega\hat{S}^{-1}X).$$

Provided that $\hat{S} \xrightarrow{p} \Omega$, one has

$$X'\hat{S}^{-1}\Omega\hat{S}^{-1}X \xrightarrow{p} X'\hat{S}^{-1}S\hat{S}^{-1}X = X'\Omega^{-1}X$$

and that $Q = \mathbb{E}(X'\hat{S}^{-1}X)$ where $X'\hat{S}^{-1}X \xrightarrow{p} X'\Omega^{-1}X$. Putting these two together means that $P = Q$ so that asymptotically we can write

$$\sqrt{N}(\hat{\beta}_{FGLS} - \beta) \xrightarrow{d} N(0, Q^{-1}) \text{ or } N\left(0, (X'\Omega^{-1}X)^{-1}\right).$$