Seemingly Unrelated Regresssions (SURE)

We consider a sample $i = 1, ..., N$ of (say) individuals. We are modeling two outcomes (in general, K), say

$$h = Z\beta^1 + v$$

and

$$c = W\beta^2 + u\,.$$

Here h is $N \times 1$, c is $N \times 1$, Z is $N \times P$, and W is $N \times Q$. We assume X and W are exogenous and the usual conditions for OLS holds, but the observations across individual are independent but $var(e_i, u_i) = \Sigma$. We can estimate these equations one-by-one using OLS, but this is not efficient if the e_i and u_i are correlated (i.e., Σ is not diagonal). The efficient estimator is GLS after having stacked the observations suitably. This is in principle trivial but to derive the formulas one has to keep carefully track of how things stack up. (I got confused by the way Bruce Hansen does it and I think it is not the best way, so here goes:)

Consider the stacked system

$$y = X\beta + e\,,$$

where $y' = (h', c')$, $e' = (v', u')$ and X is block-diagonal with Z and W on the diagonal:

$$X = \begin{pmatrix} Z & 0 \\ 0 & W \end{pmatrix}.$$

Of course, the OLS estimator is $(X'X)^{-1}X'Y$, which gives the same coefficients as OLS for each equation. But this is not efficient as it ignores the correlation. So what is the variance covariance matrix for Ee'? If the variance in Σ are σ_v^2 and σ_u^2 and the covariance is σ_{uv}, then

$$Ee' = \begin{pmatrix} \sigma_v^2 & 0 & 0 & ... & 0 & \sigma_{uv} & 0 & 0 & ... & 0 \\ 0 & \sigma_v^2 & 0 & ... & 0 & 0 & \sigma_{uv} & 0 & ... & 0 \\ 0 & 0 & \vdots & ... & 0 & 0 & 0 & \vdots & ... & 0 \\ 0 & 0 & 0 & ... & \sigma_v^2 & 0 & 0 & 0 & ... & \sigma_{uv} \\ \sigma_{uv} & 0 & 0 & ... & 0 & \sigma_u^2 & 0 & 0 & ... & 0 \\ 0 & \sigma_{uv} & 0 & ... & 0 & 0 & \sigma_u^2 & 0 & ... & 0 \\ 0 & 0 & \vdots & ... & 0 & 0 & 0 & \vdots & ... & 0 \\ 0 & 0 & 0 & ... & \sigma_{uv} & 0 & 0 & 0 & ... & \sigma_u^2 \end{pmatrix} .$$

Take a bit of time to understand this. On the diagonal is for person i first the h terms and then the c terms. The covariance in the first row is in column $N+1$ because there is where the second variable for observation $i = 1$ is located. For the second row, that happens in column $N+2$ etc. However, a much more compact way of writing this is

$$Ee' = \Sigma \otimes I_N ,$$

where I_N is the identity matrix of order N. The GLS-estimator is then

$$\hat{\beta} = [X'(\Sigma \otimes I_N)^{-1}X]^{-1}X'(\Sigma \otimes I_N)^{-1}Y .$$

Nothing is new here, except the work in figuring out the structure of the variance matric. The GLS estimator is more efficient, and allows you easily test restrictions across the β^1 and β^2 parameters, so I have sometimes used this because I wanted to test even if my dataset was so big that efficiency was not an issue.. The variance of β is given by the GLS formula $var(\beta) = [X'(\Sigma \otimes I_N)^{-1}X]^{-1}$. There is nothing in the setup here that does not instantly generalize to K equations, so the formulas are valid for any number of equations. What if we didn't observe, say, c, for some individuals? We could still do the stacked GLS, but you would not get the nice formula with the Kronecker product.

If the regressors are the same in the two equations, things simplify a lot. Let us call the joint regressor Z. Now we have

$$X = I_2 \otimes Z \,,$$

which would have same form with I_K instead of I_2 if we have K equation. We then get the GLS formula

$$\hat{\beta} = [(I_2 \otimes Z')(\Sigma \otimes I_N)^{-1}(I_2 \otimes Z)]^{-1}[(I_2 \otimes Z')(\Sigma \otimes I_N)^{-1}]Y \,.$$

using the multiplication formula for Kronecker products we get

$$\hat{\beta} = [\Sigma^{-1} \otimes Z'Z]^{-1}(\Sigma^{-1} \otimes Z')\,Y \,.$$

or

$$\hat{\beta} = [\Sigma \otimes (Z'Z)^{-1}](\Sigma^{-1} \otimes Z')\,Y \,.$$

or

$$\hat{\beta} = [I_2 \otimes (Z'Z)^{-1}Z']\,Y \,.$$

Look at this. $I_2 \otimes (Z'Z)^{-1}Z'$ is block-diagonal with $(Z'Z)^{-1}Z'$ on the diagonal, so we get

$$\begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1}Z'h \\ (Z'Z)^{-1}Z'c \end{pmatrix} ,$$

that is, OLS equation by equation. (Which you could have logically deduced when you saw the variance matrix cancelling out.) Note that the variance of β is

$$\mathrm{Var}(\beta) = [I_2 \otimes (Z'Z)^{-1}Z']\mathrm{Var}(Y)[I_2 \otimes (Z'Z)^{-1}Z]' \,,$$

or

$$\mathrm{Var}(\beta) = [I_2 \otimes (Z'Z)^{-1}Z'](\Sigma \otimes I_N)[I_2 \otimes (Z'Z)^{-1}Z']' = \Sigma \otimes (Z'Z)^{-1}Z'Z\,(Z'Z)^{-1} \,,$$

or

$$\mathrm{Var}(\beta) = \Sigma \otimes (Z'Z)^{-1} ,$$

so unless Σ is diagonal, you still need to estimate it if you want to test across the equations. You can see that the t-stats for, say, β_1 is the same as you get from equation-by-equation OLS estimation.