GLS vs. OLS in SUR model
(Goldberger, ch. 30)

Model:

$y_1 = x_1\beta_1 + \varepsilon_1$

$y_2 = x_2\beta_2 + \varepsilon_2$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$y = X\beta + \epsilon$

where $X = \begin{bmatrix} x_1 & \\ & x_2 \end{bmatrix}$ 2nx(k1+k2), $(\Sigma \otimes I) = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{21}I & \sigma_{22}I \end{bmatrix}$ 2nx2n,

$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ 2nx1, $\epsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$ 2nx1, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, (k1+k2)x1

Two estimators (good vs. evil):

$\widehat{\beta}_{gls} = (X'(\Sigma \otimes I)^{-1}X)^{-1}X'(\Sigma \otimes I)^{-1}y$

$\widehat{\beta}_{ols} = (X'X)^{-1}X'y$ (equation by equation)

Two variances:

$V(\widehat{\beta}_{gls}|X) = (X'(\Sigma \otimes I)^{-1}X)^{-1}$

$$V(\widehat{\beta}_{ols}|X) = \begin{bmatrix} \sigma_{11}(x_1'x_1)^{-1} & \\ & \sigma_{22}(x_2'x_2)^{-1} \end{bmatrix}$$

The showdown: which variance is larger?

Let’s look at one of the estimators (the result applies symmetrically to the other)

$V(\widehat{\beta}_{2,gls}|X) = (X'(\Sigma \otimes I)^{-1}X)^{-1}_{2,2}$

$$X'(\Sigma \otimes I)^{-1}X = \begin{bmatrix} \sigma^{11}x_1'x_1 & \sigma^{12}x_1'x_2 \\ \sigma^{12}x_2'x_1 & \sigma^{22}x_2'x_2 \end{bmatrix}$$

where $\sigma^{22} = \dfrac{\sigma_{11}}{\sigma_{11}\sigma_{22} - (\sigma_{12})^2}$, $\sigma^{11} = \dfrac{\sigma_{22}}{\sigma_{11}\sigma_{22} - (\sigma_{12})^2}$, $\sigma^{12} = \dfrac{-\sigma_{12}}{\sigma_{11}\sigma_{22} - (\sigma_{12})^2}$

The 2,2 block is V($\widehat{\beta}_{2,gls}|X$)
By submatrix of inverse thm (see golberger, p.191)

$$\begin{aligned} \text{V}(\widehat{\beta}_{2,gls}|X) &= [\sigma^{22}x_2'x_2 - \sigma^{12}x_2'x_1(\sigma^{11}x_1'x_1)^{-1}\sigma^{12}x_1'x_2]^{-1} \\ &= [\sigma^{22}x_2'x_2 - \frac{(\sigma^{12})^2}{\sigma^{11}}x_2'x_1(x_1'x_1)^{-1}x_1'x_2]^{-1} \\ &= [\sigma^{22}x_2'x_2 - \frac{(\sigma^{12})^2}{\sigma^{11}}x_2^{*\prime}x_2^*]^{-1} \end{aligned}$$

where $x_2^* = I - x_1(x_1'x_1)^{-1}x_1'x_2$
Note: x_2^* are simply the estimated residuals from a regression of x_2 on x_1.

[Proof:
$x_2^{*\prime}x_2^* = x_2'M_1'M_1x_2$
where $M_1 = I - x_1(x_1'x_1)^{-1}x_1' = I - w_1$
$M_1'M_1 = (I - w_1)'(I - w_1) = (w_1' - I)(I - w_1) = w_1'w_1$
$w_1'w_1 = x_1(x_1'x_1)^{-1}x_1'x_1(x_1'x_1)^{-1}x_1' = w_1$($w_1$is idempotent)
Thus, $x_2^{*\prime}x_2^* = x_2'w_1x_2 = x_2'x_1(x_1'x_1)^{-1}x_1'x_2$]

Manipulation of the other terms (σ^{22}, etc.) leads to...

$$\text{V}(\widehat{\beta}_{2,gls}|X) = \sigma_{22}[x_2'x_2 + \frac{\rho^2}{1-\rho^2}x_2^{*\prime}x_2^*]^{-1}$$

Compare this to the variance of ols for beta 2:

$$\text{V}(\widehat{\beta}_{2,ols}|X) = \sigma_{22}(x_2'x_2)^{-1}$$

Note: V($\widehat{\beta}_{2,gls}|X$) $\leq$ V($\widehat{\beta}_{2,ols}|X$) if $x_2'x_2 + \dfrac{\rho^2}{1-\rho^2}x_2^{*\prime}x_2^* \geq x_2'x_2$

Thus, the larger the $\dfrac{\rho^2}{1-\rho^2}x_2^{*\prime}x_2^*$ term, the larger is the efficiency gain (V($\widehat{\beta}_{2,gls}|X$) –V($\widehat{\beta}_{2,ols}|X$)) of the GLS estimator.

Cases (from lecture notes):
1) No correlation across equations: $\sigma_{12} = 0$.
This implies $\rho^2 = 0$, thus there is no efficiency gain to GLS.

2) Identical xs in each equation: $x_2 = x_1 = x_0$. (cov($x_{i2}, x_{i1}) = 1$) for all i).
This implies $x_2^* = 0$ is zero, thus there is no efficiency gain to GLS.
$x_2^* = (I - x_1(x_1'x_1)^{-1}x_1')x_2 = (I - x_0(x_0'x_0)^{-1}x_0')x_0 = x_0 - x_0(x_0'x_0)^{-1}x_0'x_0 = 0.$

Note: you can also see 1) and 2) by showing that in both cases $\widehat{\beta}_{2,gls} = \widehat{\beta}_{2,ols}$.

3) xs are orthogonal: cov($x_{i2}, x_{i1}) = 0$ for all i.

This maximizes the sum of squared residuals ($x_2^{*\prime} x_2^*$) from the regression of x2 on x1.
Conditional on ρ, this case gives the highest efficiency gain to GLS.

The moral of the story:
GLS is best when the xs are uncorrelated with each other and there is high correlation in cross equation errors.