

# Section 5: GLS and SUR

Jeffrey Greenbaum

February 16, 2007

## Contents

<b>1</b>	<b>Section Preamble</b>	<b>1</b>
<b>2</b>	<b>GLS</b>	<b>2</b>
2.1	The GLS Estimator . . . . .	2
2.2	Relative Efficiency . . . . .	3
2.3	Exercises . . . . .	5
2.3.1	2004 Exam, Question 1A . . . . .	5
2.3.2	Relative Efficiency of GLS to OLS . . . . .	5
<b>3</b>	<b>Robust OLS Estimation</b>	<b>6</b>
3.1	OLS Asymptotics . . . . .	6
3.2	Robust Covariance Matrix Estimator . . . . .	7
<b>4</b>	<b>Feasible GLS Alternatives: SUR</b>	<b>7</b>
4.1	Motivation and Examples . . . . .	7
4.2	SUR Model . . . . .	8
4.3	Exercises . . . . .	9
4.3.1	Goldberger 30.1 . . . . .	9
4.3.2	Goldberger 30.2 . . . . .	10
4.3.3	Goldberger 30.3 . . . . .	11

## 1 Section Preamble

In this section, we relax the hypothesis that  $Var(\varepsilon|X) = \sigma^2 I$ , which is equivalent to relaxing  $Var(y|X) = \sigma^2 I$ . Recall that this assumption combines two assumptions. First, the variance of the errors are conditionally homoskedastic:  $Var(\varepsilon_i|X) = \sigma^2 \forall i$ . In other words, the elements along the diagonal of the variance-covariance matrix of  $\varepsilon$  all equal  $\sigma^2$ . Second,  $Cov(\varepsilon_i, \varepsilon_j|X) = 0 \forall i \neq j$ . That is, all non-diagonal elements of the variance-covariance matrix of  $\varepsilon$  are zero.

Instead, we assume that such variance-covariance matrix exists, and it has a general form of  $Var(y|X) = \Sigma = \sigma^2\Omega$ .  $\Omega$  is assumed to be nonsingular and by construction symmetric and positive definite. Goldberger proves that the value of  $\sigma^2$  is not needed for efficient estimation.

Recall that we are assuming that the regressors are non-stochastic. However, if regressors are not non-stochastic, then we can obtain equivalent calculations by conditioning on them. We will proceed in terms of the conditional because of its wide use in econometrics. Assuming the regressors to be non-stochastic is unrealistic because economics is generally a nonexperimental science that relies on random sampling from a population distribution.

As usual, we will ask the two questions related to relaxing an hypothesis:

1. Where did we use this hypothesis? What changes without it?

We have used this hypothesis in the proof of the Gauss-Markov Theorem. Specifically, we have used it to show that  $\hat{\beta}_{OLS}$  is the most efficient of the class of linear unbiased estimators. Although  $\hat{\beta}_{OLS}$  remains consistent and unbiased, we will show that it is no longer efficient because the variance changes. As a result,  $\hat{\beta}_{OLS}$  is no longer BLUE. Moreover, it has an inconsistent covariance matrix estimate, so estimating standard errors becomes problematic.

2. How can we remedy that?

If  $\Omega$  is unknown, then the first option is to proceed with OLS because  $\hat{\beta}_{OLS}$  is still unbiased. The standard errors, however, must be corrected using robust estimation.

The second alternative is to use a new estimator that is now most efficient amongst the class of linear, unbiased estimators, known as Aitken's Generalized Least Squares, or GLS. GLS is ideal if we know  $\Omega$ . However, if we do not assume any structure for  $\Omega$ , then it is not plausible to get a pivotal estimate because there are more elements of  $\Omega$  to estimate than data points. Yet if we can make a simplifying assumption to the general form of  $\Omega$ , then we can use Feasible GLS to estimate  $\Omega$ , which is at least as efficient as OLS

## 2 GLS

In this section we derive the GLS estimator as the best linear unbiased estimator if  $Var(y|X) = \sigma^2\Omega$ .

### 2.1 The GLS Estimator

If  $\Omega$  is known, then the best linear unbiased estimator, known as the GLS estimator, can be derived by a transformation to the classical linear regression model and computing the least squares estimate. Because  $\Omega$  is positive definite, the Cholesky decomposition provides that there exists a nonsingular  $\Omega^{1/2}$  such that  $\Omega = \Omega^{1/2'}\Omega^{1/2} = \Omega^{1/2}\Omega^{1/2}$ . In this section we transform the linear model by multiplying through by  $\Omega^{-1/2}$  and keep the remaining classical regression assumptions. We confirm that this model satisfies the classical linear regression assumptions.

Thus, the transformed model is:  $\Omega^{-1/2}y = \Omega^{-1/2}X + \Omega^{-1/2}\varepsilon$ .

Full Rank Regressors -

We still assume that  $\text{rank}(X) = K$ . As Ruud proves on p.855, because  $\Omega^{-1/2}$  is nonsingular,  $\text{rank}(\Omega^{-1/2}X) = K$ .

Nonstochastic Regressors -

We still assume that  $X$  is nonstochastic and since  $\Omega^{-1/2}$  is treated as given,  $\Omega^{-1/2}X$  is nonstochastic. Note that since  $\Omega^{-1/2}$  is nonsingular,  $X$  and  $\Omega^{-1/2}X$  contain the same information, and equivalently, we can condition on either.

Linear Expectation -

$$\begin{aligned} E(\Omega^{-1/2}\varepsilon|\Omega^{-1/2}X) &= E(\Omega^{-1/2}\varepsilon|X) \\ &= \Omega^{-1/2}E(\varepsilon|X) = \Omega^{-1/2}0 = 0 \end{aligned}$$

Scalar Covariance Matrix -

$$\begin{aligned} \text{Var}(\Omega^{-1/2}\varepsilon|\Omega^{-1/2}X) &= \text{Var}(\Omega^{-1/2}\varepsilon|X) \\ &= \Omega^{-1/2}\text{Var}(\varepsilon|X)\Omega^{-1/2'} \\ &= \Omega^{-1/2}(\sigma^2\Omega)\Omega^{-1/2} = \sigma^2I \end{aligned}$$

Therefore by the Gauss-Markov Theorem, the least squares estimate of this model is BLUE:

$$\begin{aligned} \hat{\beta}_{GLS} &= ((\Omega^{-1/2}X)'(\Omega^{-1/2}X))^{-1}(\Omega^{-1/2}X)'(\Omega^{-1/2}y) \\ &= (X'\Omega^{-1/2}\Omega^{-1/2}X)^{-1}X'\Omega^{-1/2}\Omega^{-1/2}y \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \end{aligned}$$

Note that  $\hat{\beta}_{GLS} = \hat{\beta}_{OLS}$  if  $\text{Var}(y|X) = \sigma^2I$  as expected from substitution of  $I$  into the model.

## 2.2 Relative Efficiency

Now we will confirm that no other linear unbiased estimator is more efficient than  $\hat{\beta}_{GLS}$  as the Gauss-Markov Theorem suggests. The purpose of this confirmation is to validate that the specific transformation made produces a least squares estimator that is BLUE.

To show that  $\hat{\beta}_{GLS}$  is BLUE for any non-singular  $\Omega$ , we need to show that  $\hat{\beta}_{GLS}$  is relatively efficient to any other linear unbiased estimate of  $\beta$ , which we call  $\tilde{\beta}$ . Note that this derivation resembles that of the Gauss-Markov Theorem.

Recall that  $\hat{\beta}_{GLS}$  is efficient relative to  $\tilde{\beta}$  if and only if:

$$\text{Var}(\tilde{\beta}|X) - \text{Var}(\hat{\beta}_{GLS}|X) \text{ is positive semi-definite}$$

First, let's confirm that  $\hat{\beta}_{GLS}$  is in the class of linear and unbiased estimators.

1. Let  $A = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$  so  $A$  is not a function of  $y$ . Then  $\hat{\beta}_{GLS} = Ay$  is linear in  $y$ .
2.  $\hat{\beta}_{GLS}$  is conditionally unbiased:

$$\begin{aligned} E(\hat{\beta}_{GLS}|X) &= E((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y|X) \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(y|X) \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta = \beta \end{aligned}$$

$\tilde{\beta}$  is a linear and unbiased estimator of  $\beta$  if:

1.  $\tilde{\beta} = Ay$ .
2.  $E(\tilde{\beta}|X) = \beta$ .

Combining these two statements:

$$\begin{aligned} E(\tilde{\beta}|X) = \beta &\iff E(Ay|X) = \beta \iff AE(y|X) = \beta \\ &\iff AX\beta = \beta \iff AX = I \implies X'A' = I \quad \text{as well.} \end{aligned}$$

Let's now take the conditional variance of both estimators to evaluate the relative efficiency claim:

$$\begin{aligned} Var(\hat{\beta}_{GLS}|X) &= Var((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y|X) \\ &= ((X'\Omega^{-1}X)^{-1}X'\Omega^{-1})Var(y|X)(X'\Omega^{-1}X)^{-1}X'\Omega^{-1})' \\ &= ((X'\Omega^{-1}X)^{-1}X'\Omega^{-1})(\sigma^2\Omega)(\Omega^{-1}X(X'\Omega^{-1}X)^{-1}) \\ &= \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1} \\ Var(\tilde{\beta}|X) &= Var(Ay|X) = AVar(y|X)A' = \sigma^2A\Omega A' \end{aligned}$$

We thus want to know whether  $\sigma^2(A\Omega A') - \sigma^2(X'\Omega^{-1}X)^{-1}$  is positive semi-definite.  $\sigma^2 > 0$  so it is equivalent to factor it out and check if  $A\Omega A' - (X'\Omega^{-1}X)^{-1}$  is positive semi-definite.

We will prove that this difference is positive semi-definite by making use of the property:

For any  $A$  and  $B$  that are invertible,  $A - B$  is positive semi-definite if and only if  $B^{-1} - A^{-1}$  is positive semi-definite (Amemiya, p. 461, Property 17).

Using this property, we can then check whether  $X'\Omega^{-1}X - (A\Omega A')^{-1}$  is positive semi-definite:

$$\begin{aligned} X'\Omega^{-1}X - (A\Omega A')^{-1} &= X'\Omega^{-1/2}\Omega^{-1/2}X - X'A'(A\Omega^{-1/2}\Omega^{-1/2}A')^{-1}AX \\ &= X'\Omega^{-1/2}I\Omega^{-1/2}X - X'\Omega^{-1/2}\Omega^{1/2}A'(A\Omega^{-1/2}\Omega^{-1/2}A')^{-1}A\Omega^{1/2}\Omega^{-1/2}X \\ &= X'\Omega^{-1/2}(I - \Omega^{1/2}A'(A\Omega^{1/2}\Omega^{1/2}A')^{-1}A\Omega^{1/2})\Omega^{-1/2}X \\ &= Z'(I - W(W'W)^{-1}W')Z \\ &= Z'(I - P)Z \end{aligned}$$

where  $Z = \Omega^{-1/2}X$ ,  $W = \Omega^{1/2}A'$ , and  $I - P$  is the projection onto  $Col(\Omega^{1/2}A')^\perp$ .

Projection matrices are idempotent and symmetric.

As a result,  $Z'(I - P)Z = Z'(I - P)(I - P)Z = ((I - P)Z)'((I - P)Z) = \|(I - P)Z\|^2$ .

This norm must have a nonnegative length, and so  $Z'(I - P)Z$  must be positive semi-definite.

## 2.3 Exercises

Professor Powell has used versions of questions from Goldberger in previous exams in the True/False section, especially those pertaining to the topics in GLS that we will cover this week and next. The first question in this section comes from Professor Powell's exam in 2004, which is in the spirit of Goldberger 27.1 The second is in the spirit of the derivation about relative efficiency and is instructive even though it has not appeared in any of the last 5 exams nor in Goldberger.

### 2.3.1 2004 Exam, Question 1A

Question: True/False/Explain. If the Generalized Regression models holds – that is,  $E(y|X) = X\beta$ ,  $Var(y|X) = \sigma^2\Omega$ , and  $X$  full rank with probability one – then the covariance matrix between Aitken's Generalized LS estimator of  $\hat{\beta}_{GLS}$  (with known  $\Omega$  matrix) and the classical LS estimator  $\hat{\beta}_{LS}$  is equal to the variance matrix of the LS estimator.

Answer: False.

$$\begin{aligned} Cov(\hat{\beta}_{GLS}, \hat{\beta}_{LS}|X) &= Cov((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, (X'X)^{-1}X'y|X) \\ &= ((X'\Omega^{-1}X)^{-1}X'\Omega^{-1})Cov(y, y|X)((X'X)^{-1}X')' \\ &= \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega X(X'X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1} = Var(\hat{\beta}_{GLS}|X) \end{aligned}$$

The correct statement would be that the covariance of the GLS and the LS estimators is equal to the variance of the \*GLS\* estimator.

### 2.3.2 Relative Efficiency of GLS to OLS

Question: True/False/Explain.  $\hat{\beta}_{GLS}$  is efficient relative to  $\hat{\beta}_{OLS}$ .

Answer: We expect this statement to be true because  $\hat{\beta}_{OLS}$  is linear and unbiased, and indeed, it is.

Recall we prove this statement by showing that  $Var(\hat{\beta}_{OLS}) - Var(\hat{\beta}_{GLS})$  is positive semi-definite.

$$\begin{aligned} Var(\hat{\beta}_{OLS}|X) &= Var((X'X)^{-1}X'y|X) \\ &= ((X'X)^{-1}X')Var(y|X)((X'X)^{-1}X')' \\ &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

Then this question reduces to showing that  $\sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} - \sigma^2(X'\Omega^{-1}X)^{-1}$  is positive semi-definite.  $\sigma^2$  can be factored out and does not affect the positive semi-definiteness of this difference because it is positive. Accordingly, using the property from Amemiya, we can check the positive semi-definiteness of:

$$\begin{aligned} & (X'\Omega^{-1}X) - ((X'X)^{-1}(X'\Omega X)(X'X)^{-1})^{-1} \\ &= (X'\Omega^{-1/2}\Omega^{-1/2}X) - (X'\Omega^{-1/2}\Omega^{1/2}X)(X'\Omega^{1/2}\Omega^{1/2}X)^{-1}(X'\Omega^{1/2}\Omega^{-1/2}X) \\ &= X'\Omega^{-1/2}(I - \Omega^{1/2}X(X'\Omega^{1/2}\Omega^{1/2}X)^{-1}X'\Omega^{1/2})\Omega^{-1/2}X \\ &= X'\Omega^{-1/2}(I - P_{\Omega^{1/2}X})\Omega^{-1/2}X \end{aligned}$$

This expression is positive semi-definite since the projection matrix reduces it to a norm.

### 3 Robust OLS Estimation

Since  $\hat{\beta}_{GLS}$  reduces to  $\hat{\beta}_{OLS}$  if  $Var(\varepsilon|X) = \sigma^2I$ , why don't we use  $\hat{\beta}_{GLS}$  all of the time? Although we have been taking  $\Omega$  as given, in practice we do not know the value of its elements since we do not observe  $\varepsilon$ . We could try to estimate its elements using our  $N$  data points. However, we cannot easily obtain a consistent estimate since  $\Omega$  has  $\frac{N(N+1)}{2}$  parameters, which is larger than the sample size. In this section, we consider the notion in which we cannot make any claims about the structure of  $\Omega$ . We then use  $\hat{\beta}_{OLS}$  because it produces an unbiased estimate, and then we present a method that can be used to correct for the standard errors that are not consistently estimated. In the subsequent section and next week, we discuss Feasible GLS methods in which we can estimate  $\Omega$  because we make assumptions about the structure of  $\Omega$  that reduce the problem to no more than  $N$  elements of  $\Omega$  to estimate.

#### 3.1 OLS Asymptotics

Recall the derivation for limiting distribution for  $\hat{\beta}_{OLS}$ :

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{OLS} - \beta) &= \left(\frac{X'X}{n}\right) (\sqrt{n}) \left(\frac{X'\varepsilon}{n}\right) \\ &\rightarrow_d N(0, E(X'X)^{-1}Var(X'\varepsilon)E(X'X)^{-1}) \end{aligned}$$

Upon relaxing the variance-covariance assumption to  $Var(\varepsilon|X) = \sigma^2\Omega$ ,

$$Var(X'\varepsilon) = plim_{n \rightarrow \infty} \frac{E(X'\varepsilon\varepsilon'X)}{n} = plim_{n \rightarrow \infty} \frac{\sigma^2(X'\Omega X)}{n}$$

Rearranging the limiting distribution expression further yields:

$$\frac{\sqrt{n}(\hat{\beta}_{OLS} - \beta)}{\sigma \sqrt{\left(\frac{X'X}{n}\right)^{-1} \left(\frac{X'\Omega X}{n}\right) \left(\frac{X'X}{n}\right)^{-1}}} \rightarrow_d N(0, 1)$$

Thus, a consistent estimator of  $Var(\hat{\beta}_{OLS})$  is  $\frac{1}{n} \left(\frac{X'X}{n}\right)^{-1} \left(\frac{\sigma^2 X'\Omega X}{n}\right) \left(\frac{X'X}{n}\right)^{-1}$ .

### 3.2 Robust Covariance Matrix Estimator

In practice, we do not know  $\Sigma = \sigma^2\Omega$ , so we must estimate it. This substitution is not trivial, however, since  $\Omega$  cannot be estimated consistently without further assumptions. Instead, Eicker (1967) and White (1980) established a method to consistently estimate  $plim\left(\sigma^2 \left(\frac{X'\Omega X}{n}\right)\right)$ .

There is large evidence, though, that the Eicker-White estimator does not behave very well in finite samples. In other words, a very large sample is necessary to be convinced by the hypothesis testing results obtained with OLS estimation and Eicker-White variance correction. We will return to this robust estimation next week when discussing heteroskedasticity and autocorrelation.

## 4 Feasible GLS Alternatives: SUR

An alternative to correcting the  $\hat{\beta}_{OLS}$  standard errors is to use the unbiased, efficient GLS estimate and to make assumptions that permit consistent estimation of  $\Omega$ . This approach is possible by arguing that  $\Omega$  has a specific structure. Often, the least squares residuals are used to establish  $\hat{\Omega}$ . Then substitute  $\hat{\Omega}$  for  $\Omega$  into  $\hat{\beta}_{GLS}$  to compute  $\hat{\beta}_{FGLS}$ . Because  $\hat{\Omega}$  is a consistent estimator of  $\Omega$ ,  $\hat{\beta}_{GLS}$  and  $\hat{\beta}_{FGLS}$  have the same asymptotic distributions. The first set-up we consider that lends itself to Feasible GLS estimation is Zellner's Seemingly Unrelated Regressions (SUR).

### 4.1 Motivation and Examples

SUR is least squares estimation on a system of equations where each individual equation,  $j$ , is stacked by each individual,  $i$ . The system then contains at least two distinct dependent variables, and each individual should be represented in each equation of the system. The important requirement is that the errors associated with each individual's equations are correlated. However, they are not correlated across individuals.

For example, suppose you would like to study factors associated with better GRE scores. It is conceivable that at least one factor that helps an individual do well on the math section also helps for the verbal and writing sections. Then there is likely to be correlation between the errors in the equation for the math score, the equation for the verbal score, and the equation for the writing score within an individual. However, ignoring possible neighborhood or peer effects, it is conceivable that the error terms within the math scores are not correlated across individuals.

Suppose, for a second example, we are interested in studying factors associated with students from UC schools attending Ph.D. programs in the humanities, social sciences, and natural sciences. We estimate three equations, one for each category of majors. It is reasonable that there is correlation in the errors for the three within a UC. However after controlling for observable factors such

as labor market conditions, we can assume that the error terms within a major are not correlated across universities. For instance, consider some unknown and not understood reason that explains why academia is cool at Berkeley but not in San Diego. Groups of friends at Berkeley may apply to graduate school regardless of their major whereas at San Diego, groups of friends will spend time at the beach and applying for jobs. Groups of friends need not be restricted to academic majors.

## 4.2 SUR Model

If we can argue that the data satisfies the SUR framework, then it satisfies the following model:

$$\begin{aligned} y_{ij} &= x'_{ij}\beta_j + \epsilon_{ij} & i = 1, \dots, N & \quad j = 1, \dots, M \\ y_j &= X_j\beta_j + \epsilon_j \end{aligned}$$

where  $i$  tracks individuals,  $j$  tracks the different categories of dependent variables.

$y_j$  is the  $N \times 1$  vector obtained by stacking the  $y_{ij}$  for a fixed  $j$ .

$X_j$  is the  $N \times K_j$  matrix obtained by stacking the row vectors  $x'_{ij}$  for a fixed  $j$  and is indexed by  $K_j$ , reflecting that the same explanatory variables are not necessarily relevant for each equation. It follows that  $\beta_j$  is a  $K_j \times 1$  vector

Each equation in terms of  $j$  satisfies the classical assumptions with one additional assumption about the relationship between each one. Thus, the assumptions of the model are:

- 1)  $E(y_j|X_j) = X_j\beta$
- 2)  $V(y_j|X_j) = \sigma_{jj}I_N$
- 2')  $Cov(y_j, y_k|X_j, X_k) = \sigma_{jk}I_N$
- 3)  $X_j$  are nonstochastic and full rank with probability 1

Assumptions 1, 2, and 3 mirror the classical linear assumptions, where assumption 2 states that for each category  $j$ , each error term has the same conditional variance of  $\sigma_{jj}$ . Assumption 2' is a modification. It says that there is only correlation between the errors within an individual. For categories  $j$  and  $k$ , all individuals have equal correlation of  $\sigma_{jk}$  for their own error terms.

Stacking once more over  $j$  yields the general representation of  $y = X\beta + \epsilon$ , in which  $y$  is the  $NM \times 1$  vector obtained by stacking over  $y_j$ , and  $X$  is a  $NM \times \sum_{j=1}^M K_j$  block-diagonal matrix, with each block being a  $X_j$  matrix. This representation is necessary so that in the matrix multiplication of  $X\beta$  we can back out each equation in terms of  $j$ .

$Var(y|X)$  requires use of the Kronecker product representation. By assumptions 2 and 2',

$$V(y|X) = \begin{pmatrix} \sigma_{11}I_N & \sigma_{12}I_N & \dots & \sigma_{1M}I_N \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{M1}I_N & \cdot & \cdot & \sigma_{MM}I_N \end{pmatrix} = \Sigma \otimes I_N$$

Substituting this variance into  $\beta_{OLS}$  and  $\beta_{GLS}$  of the matrix form of the linear model yields:

$$\begin{aligned}\hat{\beta}_{OLS} &= (X'X)^{-1}X'y \\ \hat{\beta}_{GLS} &= (X'(\Sigma \otimes I_N)^{-1}X)^{-1}X'(\Sigma \otimes I_N)^{-1}y\end{aligned}$$

The conditional variances of each are:

$$\begin{aligned}Var(\hat{\beta}_{OLS}|X) &= ((X'X)^{-1}X')Var(y|X)((X'X)^{-1}X')' = (X'X)^{-1}X'(\Sigma \otimes I_N)X(X'X)^{-1} \\ Var(\hat{\beta}_{GLS}|X) &= (X'(\Sigma \otimes I_N)^{-1}X)^{-1}X'(\Sigma \otimes I_N)^{-1}(\Sigma \otimes I_N)(\Sigma \otimes I_N)^{-1}X(X'(\Sigma \otimes I_N)^{-1}X)^{-1} \\ &= (X'(\Sigma \otimes I_N)^{-1}X)^{-1}\end{aligned}$$

Professor Powell derives in his lectures notes two cases in which GLS reduces to OLS with the SUR model:

- a)  $\Sigma$  diagonal, and thus there is no seeming since the equations are indeed unrelated.
- b)  $X_j = X_0$  for each  $j$ , where each equation has the same explanatory variables.

### 4.3 Exercises

A version of Goldberger 30.1 appeared in both the 2002 and 2005 exams. A version of Goldberger 30.2 appeared in 2003. Thus, this section presents solutions to 30.1, 30.2, and 30.3 in Goldberger.

#### 4.3.1 Goldberger 30.1

Question: True or False? In the SUR model, if the explanatory variables in the two equations are identical, then the LS residuals from the two equations are uncorrelated with each other.

Answer: The statement is false unless  $\sigma_{12} = 0$ , thereby making the equations unrelated.

$$\text{Let } \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \quad \text{where } Var(y|X) = \Sigma = \begin{pmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{21}I & \sigma_{22}I \end{pmatrix}$$

Suppose  $X_1 = X_2 = X$ .

Then using OLS,  $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y_1 = (X'X)^{-1}X'y_1$  and  $\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y_2 = (X'X)^{-1}X'y_2$ .

The residual vector from the first equation is  $e_1 = y_1 - X_1\hat{\beta}_1 = Iy_1 - X(X'X)^{-1}X'y_1 = (I - P_X)y_1$  where  $P_X = X(X'X)^{-1}X'$  is a projection matrix so  $(I - P_X)$  is a projection matrix.

Similarly for the second equation,  $e_2 = y_2 - X_2\hat{\beta}_2 = Iy_2 - X(X'X)^{-1}X'y_2 = (I - P_X)y_2$ .

$$\begin{aligned}Cov(e_1, e_2|X) &= Cov((I - P_X)y_1, (I - P_X)y_2|X) \\ &= (I - P_X)Cov(y_1, y_2|X)(I - P_X)' \\ &= (I - P_X)\sigma_{12}I(I - P_X) \\ &= \sigma_{12}(I - P_X)(I - P_X) = \sigma_{12}(I - P_X) \neq 0\end{aligned}$$

### 4.3.2 Goldberger 30.2

Question: True or False? 1. In the SUR Model, if the explanatory variables in the two equations are orthogonal to each other, then the LS coefficient estimates for the two equations are uncorrelated with each other. 2. The GLS estimate reduces to the LS estimate.

Answer: The first statement is true, the second statement is false.

$$1. \text{ Let } \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \quad \text{where } \text{Var}(y|X) = \Sigma = \begin{pmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{21}I & \sigma_{22}I \end{pmatrix}$$

Using OLS,  $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y_1$  and  $\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y_2$ .

If the explanatory variables in the two equations are orthogonal to each other, then  $X_1'X_2 = 0$ .

$$\begin{aligned} \text{Cov}(\hat{\beta}_1, \hat{\beta}_2|X) &= ((X_1'X_1)^{-1}X_1')\text{Cov}(y_1, y_2|X)((X_2'X_2)^{-1}X_2')' \\ &= (X_1'X_1)^{-1}X_1'\sigma_{12}I(X_2(X_2'X_2)^{-1}) \\ &= \sigma_{12}(X_1'X_1)^{-1}X_1'X_2(X_2'X_2)^{-1} \\ &= \sigma_{12}(X_1'X_1)^{-1}(0)(X_2'X_2)^{-1} = 0 \end{aligned}$$

Thus, it is true that the covariance of OLS estimators  $\hat{\beta}_1$  and  $\hat{\beta}_2$  is zero.

2. (Note Professor Powell added this part to Goldberger 30.2 in the 2003 exam.)

$$\begin{aligned} \hat{\beta}_{GLS} &= \left( \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}' \begin{pmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{21}I & \sigma_{22}I \end{pmatrix} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}' \begin{pmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{21}I & \sigma_{22}I \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \left( \begin{pmatrix} \sigma_{11}X_1' & \sigma_{12}X_1' \\ \sigma_{12}X_2' & \sigma_{22}X_2' \end{pmatrix} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} \sigma_{11}X_1' & \sigma_{12}X_1' \\ \sigma_{12}X_2' & \sigma_{22}X_2' \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11}X_1'X_1 & \sigma_{12}X_1'X_2 \\ \sigma_{12}X_2'X_1 & \sigma_{22}X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{11}X_1'y_1 + \sigma_{12}X_1'y_2 \\ \sigma_{12}X_2'y_1 + \sigma_{22}X_2'y_2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11}X_1'X_1 & 0 \\ 0 & \sigma_{22}X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{11}X_1'y_1 + \sigma_{12}X_1'y_2 \\ \sigma_{12}X_2'y_1 + \sigma_{22}X_2'y_2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sigma_{11}}(X_1'X_1)^{-1} & 0 \\ 0 & \frac{1}{\sigma_{22}}(X_2'X_2)^{-1} \end{pmatrix} \begin{pmatrix} \sigma_{11}X_1'y_1 + \sigma_{12}X_1'y_2 \\ \sigma_{12}X_2'y_1 + \sigma_{22}X_2'y_2 \end{pmatrix} \\ &= \begin{pmatrix} (X_1'X_1)^{-1}X_1'y_1 + \frac{\sigma_{12}}{\sigma_{11}}(X_1'X_1)^{-1}X_1'y_2 \\ \frac{\sigma_{21}}{\sigma_{22}}(X_2'X_2)^{-1}X_2'y_1 + (X_2'X_2)^{-1}X_2'y_2 \end{pmatrix} \\ &\neq \begin{pmatrix} (X_1'X_1)^{-1}X_1'y_1 \\ (X_2'X_2)^{-1}X_2'y_2 \end{pmatrix} \\ &= \hat{\beta}_{OLS} \end{aligned}$$

Thus,  $\hat{\beta}_{GLS}$  does not reduce to  $\hat{\beta}_{OLS}$  in this case.

### 4.3.3 Goldberger 30.3

Question: Suppose that  $E(y_1) = x_1\beta_1$ ,  $E(y_2) = x_2\beta_2$ ,  $V(y_1) = 4I$ ,  $V(y_2) = 5I$ , and  $C(y_1, y_2) = 2I$ . Here  $y_1, y_2, x_1$ , and  $x_2$  are  $n \times 1$ , with  $x_1'x_1 = 5$ ,  $x_2'x_2 = 6$ ,  $x_1'x_2 = 3$ . Calculate the variances of the OLS and GLS estimators.

Answer:

$$\text{Let } \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \quad \text{where } \text{Var}(y|X) = \Sigma = \begin{pmatrix} 4I & 2I \\ 2I & 5I \end{pmatrix}$$

OLS Variance -

Recall that  $\text{Var}(\beta_{OLS}|X) = \text{Var}((X'X)^{-1}X'y|X) = (X'X)^{-1}X'\text{Var}(y|X)X(X'X)^{-1}$ :

$$\begin{aligned} (X'X)^{-1} &= \left( \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}' \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right)^{-1} = \begin{pmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 5 & 0 \\ 0 & 6 \end{pmatrix}^{-1} = \begin{pmatrix} 1/5 & 0 \\ 0 & 1/6 \end{pmatrix} \\ X'\Sigma X &= \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}' \begin{pmatrix} 4I & 2I \\ 2I & 5I \end{pmatrix} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} = \begin{pmatrix} 4X_1' & 2X_1' \\ 2X_2' & 5X_2' \end{pmatrix} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \\ &= \begin{pmatrix} 4X_1'X_1 & 2X_1'X_2 \\ 2X_2'X_1 & 5X_2'X_2 \end{pmatrix} = \begin{pmatrix} 20 & 6 \\ 6 & 30 \end{pmatrix} \\ (X'X)^{-1}X'\Sigma X(X'X)^{-1} &= \begin{pmatrix} 1/5 & 0 \\ 0 & 1/6 \end{pmatrix} \begin{pmatrix} 20 & 6 \\ 6 & 30 \end{pmatrix} \begin{pmatrix} 1/5 & 0 \\ 0 & 1/6 \end{pmatrix} \\ &= \begin{pmatrix} 4/5 & 1/5 \\ 1/5 & 5/6 \end{pmatrix} \end{aligned}$$

GLS Variance -

Recall that  $\text{Var}(\hat{\beta}_{GLS}|X) = (X'(\Sigma \otimes I_N)^{-1}X)^{-1}$ :

$$\begin{aligned} (\Sigma \otimes I_N)^{-1} &= \begin{pmatrix} 4I & 2I \\ 2I & 5I \end{pmatrix}^{-1} = \frac{1}{16} \begin{pmatrix} 5I & -2I \\ -2I & 4I \end{pmatrix} \\ (X'(\Sigma \otimes I_N)^{-1}X)^{-1} &= \left[ \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}' \left( \frac{1}{16} \begin{pmatrix} 5I & -2I \\ -2I & 4I \end{pmatrix} \right) \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right]^{-1} \\ &= \left( \frac{1}{16} \begin{pmatrix} 5X_1'X_1 & -2X_1'X_2 \\ -2X_2'X_1 & 4X_2'X_2 \end{pmatrix} \right)^{-1} = \left( \frac{1}{16} \begin{pmatrix} 25 & -6 \\ -6 & 24 \end{pmatrix} \right)^{-1} = \begin{pmatrix} \frac{32}{47} & \frac{8}{47} \\ \frac{8}{47} & \frac{100}{141} \end{pmatrix} \end{aligned}$$

Note that the difference between the OLS and GLS variances is positive semidefinite.