Ch. 8 Nonspherical Disturbances

(July 25, 2017)

1 Introduction

This chapter will assume that the full ideal conditions hold except that the covariance matrix of the disturbances , i.e. $E(\boldsymbol{\varepsilon\varepsilon'}) = \sigma^2\mathbf{\Sigma}$, where $\mathbf{\Sigma}$ is not the identity matrix. In particular, $\mathbf{\Sigma}$ may be nondiagonal and/or have unequal diagonal elements.

Two cases we shall consider in details are *heteroscedasticity* and *autocorrelation.*

(a). Disturbances are heteroscedastic when they have different variance. Heteroscedasticity usually arises in cross-section data where the scale of the dependent variable and the explanatory power of the model tend to vary across observations. The disturbances are still assumed to be uncorrelated across observation, so $\sigma^2\mathbf{\Sigma}$ would be[1]

$$\sigma^2\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & . & . & . & 0 \\ 0 & \omega_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \omega_N \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & . & . & . & 0 \\ 0 & \sigma_2^2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \sigma_N^2 \end{bmatrix}.$$

(b). Autocorrelation is usually found in time-series data. Economic time-series often display a "memory" in that variation around the regression function is not inde-

[1] In most cases disturbances are heteroscedastic when we are dealing with cross-sectional sample, so we use "sub-i" here.

pendent from one period to the next. Time series data are usually homoscedasticity, so $\sigma^2\mathbf{\Sigma}$ would be

$$\sigma^2\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & . & . & . & \rho_{T-1} \\ \rho_1 & 1 & . & . & . & \rho_{T-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \rho_{T-1} & \rho_{T-2} & . & . & . & 1 \end{bmatrix}.$$

$\boxed{\text{Example}}$.
See Figure 8.1 on p.159 (Ch. 8) and Figure 19.1 on p.627 (Ch. 19) of Greene 6th edition. ■

In recent studies, *panel data* sets, constituting of cross sections observed at several points in time, have exhibited both characteristics. The next three chapter examines in details specific types of generalized regression models.

Our earlier results for the classical mode will have to be modified. We first consider the consequence of the more general model for the ordinary least squares estimators.

2 Efficient Estimators

It is useful to begin with considering cases in which $\mathbf{\Sigma}$ is a known, symmetric and positive definite matrix. This assumption will occasionally be true, but in most models, $\mathbf{\Sigma}$ will contain unknown parameters that must also be estimated.

Example.

Assume that the variance of disturbances in each sample ($i = 1, 2, ..., N$) is different but is just the product of a unknown overall variance σ^2 and the "second" explanatory variable (say) X_2, i.e. $\sigma_i^2 = \sigma^2 X_{2i}$, then

$$\sigma^2\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 X_{21} & 0 & . & . & . & 0 \\ 0 & \sigma^2 X_{22} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \sigma^2 X_{2N} \end{bmatrix} = \sigma^2 \begin{bmatrix} X_{21} & 0 & . & . & . & 0 \\ 0 & X_{22} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & X_{2N} \end{bmatrix},$$

therefore, we have a "known" $\mathbf{\Sigma}$,

$$\mathbf{\Sigma} = \begin{bmatrix} X_{21} & 0 & . & . & . & 0 \\ 0 & X_{22} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & X_{2N} \end{bmatrix}$$

once upon we have this sample of size N. ■

2.1 Generalized Least Square (GLS) Estimators

2.1.1 The Slopes and Variance Estimators

If $\mathbf{\Sigma}$ is a (known) positive symmetric matrix, it can be factored into

$$\mathbf{\Sigma} = \mathbf{C\Lambda C}',$$

where the column of $\mathbf{C}$ are the eigenvectors of $\mathbf{\Sigma}$ and the eigenvalues of $\mathbf{\Sigma}$ are arrayed in the diagonal matrix $\mathbf{\Lambda}$. Let $\mathbf{\Lambda}^{1/2}$ be the diagonal matrix with ith diagonal element

$\sqrt{\lambda_i}$. Also let $\mathbf{P}' = \mathbf{C}\mathbf{\Lambda}^{-1/2}$, so

$$\mathbf{\Sigma}^{-1} = \mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{C}' = \mathbf{C}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{C}' = \mathbf{P}'\mathbf{P}.$$

Theorem.

Suppose that the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ satisfies the ideal conditions except that $\mathbf{\Sigma}$ is not the identity matrix. Suppose that

$$\lim_{T\to\infty} \frac{\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}}{T}$$

is finite and nonsingular. Then the transformed equation

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon} \tag{8-1}$$

satisfies the full ideal condition.

Proof.

Since $\mathbf{P}$ is nonsingular and nonstochastic, $\mathbf{PX}$ is nonstochastic and of full rank if $\mathbf{X}$ is (Condition 2 and 5). Also, for the consistency of OLS estimators

$$\lim_{T\to\infty} \frac{(\mathbf{PX})'(\mathbf{PX})}{T} = \lim_{T\to\infty} \frac{\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}}{T}$$

is finite and nonsingular by assumption. Therefore the transformed regressors matrix satisfies the required conditions, and we need consider only the transformed disturbance $\mathbf{P}\boldsymbol{\varepsilon}$.

Clearly, $E(\mathbf{P}\boldsymbol{\varepsilon}) = 0$ (Condition 3). Also

$$\begin{aligned} E(\mathbf{P}\boldsymbol{\varepsilon})(\mathbf{P}\boldsymbol{\varepsilon})' &= \sigma^2\mathbf{P}\mathbf{\Sigma}\mathbf{P}' \\ &= \sigma^2(\mathbf{\Lambda}^{-1/2}\mathbf{C}')(\mathbf{C}\mathbf{\Lambda}\mathbf{C}')(\mathbf{C}\mathbf{\Lambda}^{-1/2}) \\ &= \sigma^2\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}\mathbf{\Lambda}^{-1/2} \\ &= \sigma^2\mathbf{I} \qquad (Condition\ 4). \end{aligned}$$

Finally, the normality (Condition 6) of $\mathbf{P}\boldsymbol{\varepsilon}$ follows immediately from the normality of $\boldsymbol{\varepsilon}$. ■

Theorem.

Suppose that the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ satisfies the ideal conditions except that $\boldsymbol{\Sigma}$ is not the identity matrix. Then the BLUE of $\boldsymbol{\beta}$ is just

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}.$$

Proof.

Since the transformed equation (8-1) satisfies the full ideal conditions, the BLUE of $\boldsymbol{\beta}$ is just

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= [(\mathbf{PX})'(\mathbf{PX})]^{-1}(\mathbf{PX})'(\mathbf{PY}) \\ &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}.\end{aligned}$$ ■

Indeed, since $\tilde{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$ in the transformed equation, and since the transformed equation (8-1) satisfies the ideal conditions, $\tilde{\boldsymbol{\beta}}$ has all the usual desirable properties–it is unbiased, BLUE, efficient, consistent, and asymptotically efficient. $\tilde{\boldsymbol{\beta}}$ is the OLS of the transformed equation, but it is a generalized least square (GLS) estimator of the original regression model which take the OLS as a subcases when $\boldsymbol{\Sigma} = \mathbf{I}$.

The generalized least squares estimator $\tilde{\boldsymbol{\beta}}$ can also be obtained by minimizing the *GLS criterion function*

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}((\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

which is just the sum of squared residuals from the transformed regression. The criterion function can be thought of as a generalization of the SSE function in which the squares and cross products of the disturbances are weighted by the inverse of the matrix $\boldsymbol{\Sigma}$.

Much attention has been devoted in the literature to the search of conditions for which the ordinary least squares estimator $\hat{\boldsymbol{\beta}}$ is equivalent to the GLS estimator $\tilde{\boldsymbol{\beta}}$, and thus it is BLUE. Anderson was the first who faced this problem, stating (1948, p. 48) and proving (1971, pp. 19 and 560) that equality between $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$, holds if and only if the matrix $\mathbf{X}$ contains k eigenvectors of $\boldsymbol{\Sigma}$ which are normalized to unit length.[2]

[2] Because $\boldsymbol{\Sigma}\mathbf{X} = \mathbf{X}\boldsymbol{\Lambda}$, therefore $\boldsymbol{\Sigma}^{-1} = \mathbf{X}\boldsymbol{\Lambda}^{-1}\mathbf{X}'$. Substitute this equation to $\tilde{\boldsymbol{\beta}}$ we obtain this result.

Theorem.

The variance-covariance of the GLS estimator $\tilde{\boldsymbol{\beta}}$ is $\sigma^2(\mathbf{X'\Sigma^{-1}X})^{-1}$.

Proof.

Viewing $\tilde{\boldsymbol{\beta}}$ as the OLS estimator in the transformed equation, it is clearly has covariance matrix

$$Var(\tilde{\boldsymbol{\beta}}) = \sigma^2[(\mathbf{PX})'(\mathbf{PX})]^{-1} = \sigma^2(\mathbf{X'\Sigma^{-1}X})^{-1}. \qquad \blacksquare$$

Theorem.

An unbiased, consistent, efficient, and asymptotically efficient estimator of σ^2 is

$$\tilde{s}^2 = \frac{\tilde{\mathbf{e}}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{e}}}{T-k},$$

where $\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$.

Proof.

Since the transformed equation satisfies the ideal conditions, the desired estimator of σ^2 is

$$\begin{aligned}
\frac{1}{T-k}(\mathbf{Py} - \mathbf{PX}\tilde{\boldsymbol{\beta}})'(\mathbf{Py} - \mathbf{PX}\tilde{\boldsymbol{\beta}}) &= \frac{1}{T-k}[\mathbf{P}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})]'[\mathbf{P}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})] \\
&= \frac{1}{T-k}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{P}'\mathbf{P}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \\
&= \frac{1}{T-k}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \\
&= \frac{1}{T-k}\tilde{\mathbf{e}}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{e}}. \qquad \blacksquare
\end{aligned}$$

2.1.2 Hypothesis Tests

Finally, for testing hypothesis we can apply the full set of results in Chapter 6 to the transformed equation (8-1). For the testing the m restrictions $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, the appropriate (one of) statistics is

$$\frac{(\mathbf{R}\tilde{\boldsymbol{\beta}}-\mathbf{q})'[\tilde{s}^2\mathbf{R}(\mathbf{PX})'(\mathbf{PX})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\tilde{\boldsymbol{\beta}}-\mathbf{q})}{m}$$

$$= \frac{(\mathbf{R}\tilde{\boldsymbol{\beta}}-\mathbf{q})'[\tilde{s}^2\mathbf{R}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\tilde{\boldsymbol{\beta}}-\mathbf{q})}{m} \sim F_{m,T-k}.$$

Exercise 1.
Derive the other three test statistics (in Chapter 6) of the $F-Ratio$ test statistics to test the hypothesis $\mathbf{R}\boldsymbol{\beta}=\mathbf{q}$ when $\boldsymbol{\Sigma}\neq\mathbf{I}$. ■

2.2 Maximum Likelihood Estimators

Assume that $\boldsymbol{\varepsilon}\sim N(0,\sigma^2\boldsymbol{\Sigma})$, if $\boldsymbol{X}$ are not stochastic, then by results from "functions of random variables" (n $\Rightarrow$ n transformation) we have $\mathbf{y}\sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{\Sigma})$. That is, the log-likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta};\mathbf{y}) &= \ln f(\boldsymbol{\theta};\mathbf{y}) \\ &= -\frac{T}{2}\ln(2\pi)-\frac{1}{2}\ln|\sigma^2\boldsymbol{\Sigma}|-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\sigma^2\boldsymbol{\Sigma})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{T}{2}\ln(2\pi)-\frac{T}{2}\ln\sigma^2-\frac{1}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}), \end{aligned}$$

where $\boldsymbol{\theta}=(\beta_1,\beta_2,...,\beta_k,\sigma^2)'$ since by assumption $\boldsymbol{\Sigma}$ is known.

The necessary conditions for maximizing L are

$$\frac{\partial L}{\partial\boldsymbol{\beta}}=\frac{1}{\sigma^2}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})=0$$

and

$$\frac{\partial L}{\partial\sigma^2}=-\frac{T}{2\sigma^2}+\frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})=0$$

The solution are

$$\tilde{\boldsymbol{\beta}}_{ML} = (\mathbf{X}'\boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{y},$$

and

$$\tilde{\sigma}^2_{ML} = \frac{1}{T}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{ML})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{ML}),$$

which implies that with normally distributed disturbances, generalized least squares are also MLE. As is the classical regression model, the MLE of σ^2 is biased. An unbiased estimator is

$$\tilde{s}^2 = \frac{1}{T-k}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{ML})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{ML}).$$

3 Estimation When Σ is Unknown

If $\mathbf{\Sigma}$ contains unknown parameters that must be estimated, then GLS is not feasible. But with an unrestricted $\mathbf{\Sigma}$, then beside $\boldsymbol{\beta}$ there are $T(T+1)/2$ additional parameters in $\sigma^2\mathbf{\Sigma}$. This number is far too many to estimate with T observations. Obviously, some structures must be imposed on the model if we are to proceed.

3.1 Feasible Generalized Least Squares

The typical problem involves a small set of parameters $\boldsymbol{\theta}$ such that $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$. For example, we may assume autocorrelated disturbances in the beginning of this chapter as

$$\sigma^2\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & . & . & . & \rho_{T-1} \\ \rho_1 & 1 & . & . & . & \rho_{T-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \rho_{T-1} & \rho_{T-2} & . & . & . & 1 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho^1 & . & . & . & \rho^{T-1} \\ \rho^1 & 1 & . & . & . & \rho^{T-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \rho^{T-1} & \rho^{T-2} & . & . & . & 1 \end{bmatrix},$$

then $\mathbf{\Sigma}$ has only one additional unknown parameter, ρ. A model of heteroscedasticity that also has only one new parameter, α, is

$$\sigma_i^2 = \sigma^2 X_{2i}^{\alpha},$$

where α is an unknown parameter.

Definition.

If $\mathbf{\Sigma}$ depends on a finite number of parameters, $\theta_1, \theta_2, ..., \theta_p$, and if $\hat{\mathbf{\Sigma}}$ depends on consistent estimators, $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p$, the $\hat{\mathbf{\Sigma}}$ is called a consistent estimator of $\mathbf{\Sigma}$. ■

Definition.

Let $\hat{\mathbf{\Sigma}}$ be a consistent estimator of $\mathbf{\Sigma}$. Then the feasible generalized least squares estimator (FGLS) of $\boldsymbol{\beta}$ is

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{y}.$$ ■

Conditions that imply that $\breve{\boldsymbol{\beta}}$ is asymptotically equivalent to $\tilde{\boldsymbol{\beta}}$ are

$$plim \left[\left(\frac{1}{T}\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}\right) - \left(\frac{1}{T}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\right)\right] = 0$$

and

$$plim \left[\left(\frac{1}{\sqrt{T}}\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\boldsymbol{\varepsilon}\right) - \left(\frac{1}{\sqrt{T}}\mathbf{X}'\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}\right)\right] = 0.$$

Theorem.

An asymptotically efficient FGLS **does not** require that we have an efficient estimator of $\boldsymbol{\theta}$; only a consistent one is required to achieve full efficiency for the FGLS estimator.■

4 Consequences for OLS Estimation

4.1 Properties of the Least Squares Estimators

We now consider the statistical properties of the *OLS* estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (If you insist to use it !) when the variance-covariance matrix of disturbance in the linear model is now assumed to be $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Sigma}$, which had violated the ideal conditions.

4.1.1 Finite Sample Properties

To reiterate, the OLS estimator is

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}.\end{aligned}$$

(a). Unbiasedness ?

The OLS slope estimator remains unbiased,

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta},$$

but the variance estimator s^2, is not, since

$$\begin{aligned}E(\mathbf{e}'\mathbf{e}) &= E(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}) \\ &= trace\ E(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\ &= \sigma^2\ trace\ \mathbf{M}\boldsymbol{\Sigma} \\ &\neq \sigma^2(T-k).\end{aligned}$$

(b). Efficiency ?

In the classical linear regression model, assume that $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Sigma}$, the covariance matrix of OLS estimator $\hat{\boldsymbol{\beta}}$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. (Instead of $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.)

To see this,

$$\begin{aligned}Var(\hat{\boldsymbol{\beta}}_{OLS}) = E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' &= E(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \qquad (8\text{-}2)\end{aligned}$$

Note that the covariance matrix of $\hat{\boldsymbol{\beta}}$ is no longer equal to $\sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$. It may be either "larger" or "smaller", in the sense that

$$(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} - (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$$

can be either positive semidefinite, negative semidefinite, or neither. However, any inference based on $s^2(\mathbf{X}'\mathbf{X})^{-1}$ is likely to be misleading. That is, the familiar inference procedures based on the t and F test statistics will *no longer be appropriate.*

(c). Exact distribution:
In the classical linear regression model, assume that $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Sigma}$. Because $\hat{\boldsymbol{\beta}}$ is a linear function of $\boldsymbol{\varepsilon}$, therefore if $\boldsymbol{\varepsilon}$ is normally distributed, then

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta},\ \sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}). \tag{8-3}$$

4.1.2 Asymptotic Properties of OLS

(a). Consistency ?:

(i.) $\boxed{\text{Theorem}}$. (Consistency of OLS)
In the classical linear regression model, assume that $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Sigma}$, Furthermore, if $\lim_{T\to\infty}(\mathbf{X}'\mathbf{X}/T)$ and $\lim_{T\to\infty}(\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}/T)$ is finite, then the OLS estimator $\hat{\boldsymbol{\beta}}$ is consistent.

Proof.

$$plim\ \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \lim_{T\to\infty}\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1} plim \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}.$$

But $\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}$ has zero mean and covariance matrix

$$\frac{\sigma^2}{T}\frac{\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}}{T}.$$

If $\lim_{T\to\infty}(\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}/T)$ is finite, then $\frac{\sigma^2}{T}\frac{\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}}{T} = \mathbf{0}$. Hence $\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}$ has zero mean and its covariance matrix vanishes asymptotically, which implies $plim\ \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T} = \mathbf{0}$. Combining with the assumption that $\lim_{T\to\infty}(\mathbf{X}'\mathbf{X}/T)$ is finite, therefore we have the result that $plim\ \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$. ∎

(ii.) Since $E(s^2) \neq \sigma^2$, it is hard to see that it is a consistent estimator of σ^2 from convergence in mean square error.

(b). Asymptotic Normality:

Since by assumption, $\mathbf{X'X}$ is $O(T)$, therefore $(\mathbf{X'X})^{-\mathbf{1}} \to \mathbf{0}$. The exact distribution of $\hat{\boldsymbol{\beta}}$, i.e., $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'\Sigma X}(\mathbf{X'X})^{-\mathbf{1}})$ will degenerate to a point in large sample. To express the limiting distribution of $\hat{\boldsymbol{\beta}}$, we need the following theorem.

$\boxed{\mathfrak{Theorem}}$. (Limiting Distribution of $\hat{\boldsymbol{\beta}}$)

Denote $\lim_{T\to\infty}(\mathbf{X'\Sigma X}/T) = \mathbf{Q}_*$. The asymptotic distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ can be expressed as

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2\underline{\mathbf{Q}}^{-1}\mathbf{Q}_*\underline{\mathbf{Q}}^{-1}) \equiv N(\mathbf{0}, \mathbf{V}), \tag{8-4}$$

where $\underline{\mathbf{Q}} = \lim_{T\to\infty}(\frac{\mathbf{X'X}}{T})$.

$\underline{\mathfrak{Proof}}$.

For any sample size T, the distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is $N\left(\mathbf{0},\, \sigma^2\left(\frac{\mathbf{X'X}}{T}\right)^{-1}\left(\frac{\mathbf{X'\Sigma X}}{T}\right)\left(\frac{\mathbf{X'X}}{T}\right)^{-1}\right)$. The above limiting results are therefore trivial. ■

This asymptotic normality result can be used to test hypotheses about $\boldsymbol{\beta}$. To construct standard tests that are asymptotically invariant to nuisance parameters, an estimate of $\mathbf{V}$ is required.

4.2 Robust Estimation of Asymptotic Covariance Matrices of OLS Estimator

All the testing procedures we have made use of standard errors or estimated covariance matrices for performing statistical inference. When we have assumed that $E(\boldsymbol{\varepsilon\varepsilon}') = \sigma^2\boldsymbol{\Sigma}$, the variance of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}_{OLS}) &= \sigma^2(\mathbf{X'X})^{-\mathbf{1}}(\mathbf{X'\Sigma X})(\mathbf{X'X})^{-\mathbf{1}} \\ &= (\mathbf{X'X})^{-\mathbf{1}}(\mathbf{X'\Omega X})(\mathbf{X'X})^{-\mathbf{1}}, \end{aligned} \tag{8-5}$$

as is seen in Eq. (8-2), where $\mathbf{\Omega} = \sigma^2\mathbf{\Sigma}$. This form of covariance matrix is often called a *sandwich covariance matrix*, for the obvious reason that the matrix $\mathbf{X}'\mathbf{\Omega}\mathbf{X}$ is sandwiched between the two instances of the matrix $(\mathbf{X}'\mathbf{X})^{-1}$. The covariance of an inefficient estimator very often takes this sandwich form.

For the purpose of asymptotic theory, we wish to consider the covariance matrix of, not $\hat{\boldsymbol{\beta}}_{OLS}$, but rather $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. The asymptotic covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is

$$\left(\lim \frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\lim \frac{1}{T}\mathbf{X}'\mathbf{\Omega}\mathbf{X}\right)\left(\lim \frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}, \tag{8-6}$$

as is seen is Eq. (8-4). To estimate the factor $\left[\lim\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\right]$, we can simply use the matrix $\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}$ itself. What is not so trivial is to estimate the middle factor $\left[\lim\left(\frac{1}{T}\mathbf{X}'\mathbf{\Omega}\mathbf{X}\right)\right]$.

It might seem that to estimate $\left[\lim\left(\frac{1}{T}\mathbf{X}'\mathbf{\Omega}\mathbf{X}\right)\right]$ using $\left[\left(\frac{1}{T}\mathbf{X}'\hat{\mathbf{\Omega}}\mathbf{X}\right)\right]$. However as estimator of $\mathbf{\Omega}$, $\hat{\mathbf{\Omega}}$, contains $T(T+1)$ unknown parameters which make this method hopeless with only T observations. But fortunately what is required is an estimator $\hat{\mathbf{Q}}$ of the $T(T+1)$ unknown elements in the matrix[3]

$$plim\ \hat{\mathbf{Q}} = plim\ \left(\frac{1}{T}\mathbf{X}'\mathbf{\Omega}\mathbf{X}\right) = plim\ \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\sigma_{ij}\mathbf{x}_i\mathbf{x}_j'.$$

The OLS estimator $\hat{\boldsymbol{\beta}}$ is consistent estimator of $\boldsymbol{\beta}$, which implies that the OLS residuals e_t are "pointwise" consistent estimators of their population counterparts ε_t. The general approach, then, will be to use $\mathbf{X}$ and $\mathbf{e}$ to devise the estimator $\hat{\mathbf{Q}}$.

In the heteroscedastic case, White (1980) showed that under certain conditions, the estimator

$$\hat{\mathbf{Q}}_0 = \frac{1}{T}\sum_{t=1}^{T} e_t^2\mathbf{x}_t\mathbf{x}_t'$$

has

$$plim\ \hat{\mathbf{Q}}_0 = plim\ \left(\frac{1}{T}\mathbf{X}'\mathbf{\Omega}\mathbf{X}\right) = plim\ \frac{1}{T}\sum_{i=1}^{T}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'.$$

---

[3] To see this, $\mathbf{X}'\mathbf{\Omega}\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_T\end{bmatrix}\begin{bmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2T}\\ \vdots & \vdots & \vdots & \vdots\\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{TT}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1'\\ \mathbf{x}_2'\\ \vdots\\ \mathbf{x}_T'\end{bmatrix} = \sum_{i=1}^{T}\sum_{j=1}^{T}\sigma_{ij}\mathbf{x}_i\mathbf{x}_j'$.

The extension of White's result to the more general case of autocorrelation is much more difficult. The natural counterpart for estimating

$$\frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\sigma_{ij}\mathbf{x}_i\mathbf{x}_j' \tag{8-7}$$

is

$$\hat{\mathbf{Q}}_1 = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}e_ie_j\mathbf{x}_i\mathbf{x}_j'.$$

Unlike the heteroscedasticity case, the matrix in Eq. (8-7) is $1/T$ times a sum of T^2 terms, so it is difficult to conclude yet it will converge to anything at all. To obtain convergence, it is necessary to assume that the terms involving unequal subscripts in (8-7) diminish importance as T grows. A sufficient condition is that the terms with subscript pairs $|i-j|$ grow smaller as the distance between them grows large. In practical terms, observation pairs are progressively less correlated as their separation in time grows.[4] Thus we achieve convergence of $\frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\sigma_{ij}\mathbf{x}_i\mathbf{x}_j'$ by assuming that the rows of $\mathbf{X}$ are well behaved and that the correlations diminish with increasing separation in time.

The practical problem is that $\hat{\mathbf{Q}}_1$ need not to be positive definite. Newey and West (1987) have devised an estimator that overcomes this difficulty:

$$\begin{aligned}
\hat{\mathbf{Q}} &= \hat{\mathbf{Q}}_0 + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w_le_te_{t-l}(\mathbf{x}_t\mathbf{x}_{t-l}' + \mathbf{x}_{t-l}\mathbf{x}_t')\\
&= \frac{1}{T}\sum_{t=1}^{T}e_t^2\mathbf{x}_t\mathbf{x}_t' + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w_le_te_{t-l}(\mathbf{x}_t\mathbf{x}_{t-l}' + \mathbf{x}_{t-l}\mathbf{x}_t')\\
&= \frac{1}{T}\left(\sum_{t=1}^{T}e_t^2\mathbf{x}_t\mathbf{x}_t' + \sum_{l=1}^{L}\sum_{t=l+1}^{T}w_le_te_{t-l}(\mathbf{x}_t\mathbf{x}_{t-l}' + \mathbf{x}_{t-l}\mathbf{x}_t')\right), \qquad (8\text{-}8)
\end{aligned}$$

where $w_l = 1 - \frac{l}{L+1}$.

Consequently we use the matrix

$$\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\hat{\mathbf{Q}}\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1} \tag{8-9}$$

to estimate expression (8-4), i.e. $\mathrm{Var}(\sqrt{T}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}))$

$$\hat{\mathbf{V}} \equiv \left(\lim\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\lim\frac{1}{T}\mathbf{X}'\mathbf{\Omega}\mathbf{X}\right)\left(\lim\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1},$$

[4] In time series analysis term, it is ergodicity. If the autocovariance of a stationary process Y_t satisfy $\sum_{j=0}^{\infty}|\gamma_j| < \infty$, then Y_t is ergodic.

consistently.

Of course, in practice, we ignore the factor of T^{-1} and use the matrix

$$\begin{aligned}
&\widehat{Var(\hat{\boldsymbol{\beta}})}_{HAC} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{t=1}^{T} e_t^2\mathbf{x}_t\mathbf{x}_t' + \sum_{l=1}^{L}\sum_{t=l+1}^{T} w_l e_t e_{t-l}(\mathbf{x}_t\mathbf{x}_{t-l}' + \mathbf{x}_{t-l}\mathbf{x}_t')\right)(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned} \tag{8-10}$$

directly to estimate the covariance of $\hat{\boldsymbol{\beta}}$. The matrix $\widehat{Var(\hat{\boldsymbol{\beta}})}_{HAC}$ is called *Heteroscedasticity and Autocorrelation Consistent Covariance* (HAC) estimator. By taking square roots of the diagonal elements of (8-10), we can obtain standard errors that are asymptotically valid in the presence of heteroscedasticity and autocorrelation of unknown form.

The null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, where $\mathbf{R}_{m\times k}$ has full rank. Under the null hypothesis, it follows from (8-4) that

$$\sqrt{T}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{R}\underline{\mathbf{Q}}^{-1}\mathbf{Q}_*\underline{\mathbf{Q}}^{-1}\mathbf{R}') = N(\mathbf{0}, \mathbf{RVR}'),$$

and therefore from Theorem at p.62 of Ch. 2 that

$$(\mathbf{R}\hat{\mathbf{V}}\mathbf{R}')^{-1/2}\sqrt{T}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_m),$$

The Wald test statistic is

$$T(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'(\mathbf{R}\hat{\mathbf{V}}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \xrightarrow{d} \chi_m^2. \tag{8-11}$$

If the true form of serial correlation/heteroskedasticity are known, then generalized least squares (GLS) provides efficient estimates and standard inference can be conducted on the GLS transformed model. But, in practice the form of serial correlation/heteroskedasticity is often unknown, and this has led to the development of techniques that permit valid asymptotic inference without having to specify a model of the serial correlation or heteroscedasticity. The most common approach is to estimate the variance-covariance matrix of the OLS estimate semi-parametrically using spectral methods (heteroscedasticity and autocorrelation consistent (HAC) estimators) and construct standard tests using the asymptotic normality of the OLS estimate.

There is a final problem to be solved. It must be determined in advance how large L is to be. Unfortunately, there is little theoretical guidance. The use of a HAC estimator involves the specification of a kernel (w_j) and a truncation lag or bandwidth. The bandwidth choice determines the fraction of the available covariance information that goes into the calculation of the long run variances. Kiefer et al. (2000) showed that even if a data-dependent method is used to choose the truncation lag (bandwidth), arbitrary choices of the truncation lag are inevitable. Furthermore, HAC has a poor finite sample performance, (see, for example, Kiefer et al., 2000; Kiefer and Vogelsang, 2005). Kiefer et al. (2000) proposed an alternative method of constructing robust test statistics; in this method, estimates of the variance covariance matrix are not explicitly required to construct the test. This approach requires a nonsingular data-dependent stochastic transformation to the OLS estimates. Therefore, arbitrary choices of the truncation lags in HAC can be avoided, and the test based on KVB approach is asymptotically invariant to serial correlation/heteroskedasticity nuisance parameters.

$\boxed{\text{Example}}$.
See the paper by Peter Boswijk, Philip Hans Franses and Dick van Dijk (2010),"Cointegration in a historical perspective", *Journal of Econometrics*, **158**, pp.156-159.

4.3 KVB estimator

Kiefer, Vogelsang, and Bunzel (2000) propose an alternative method of constructing robust test statistics. They apply a nonsingular data dependent stochastic transformation to the OLS estimates. The asymptotic distribution of the transformed estimates does not depend on nuisance parameters. Then, test statistics that are asymptotically invariant to nuisance parameters (asymptotic pivotal statistics) are constructed. The resulting test statistics have nonstandard asymptotic distributions that only depend on the number of restrictions being tested, and critical values are easy to simulate using standard techniques.

Define $\hat{\mathbf{s}}_t = \sum_{j=1}^{t} \mathbf{x}_j e_j$, and consider $\hat{\mathbf{C}} = T^{-2}\sum_{t=1}^{T} \mathbf{s}_t\mathbf{s}_t'$, KVB (2000) show that

$$\hat{\mathbf{C}} = T^{-2}\sum_{t=1}^{T} \mathbf{s}_t\mathbf{s}_t' \xrightarrow{d} \mathbf{\Lambda}\mathbf{P}_k\mathbf{\Lambda}' \equiv \mathbf{\Lambda}\mathbf{Z}_k\mathbf{Z}_k'\mathbf{\Lambda}' \equiv \mathbf{C}^{1/2}\mathbf{C}'^{1/2}, \tag{8-12}$$

where $\mathbf{\Lambda}\mathbf{\Lambda}' = \sigma^2\mathbf{Q}_*$ and

$$\mathbf{P}_k = \int_0^1 \left[(\mathbf{W}_k(r) - r\mathbf{W}_k(1))\,(\mathbf{W}_k(r) - r\mathbf{W}_k(1))'\right] dr = \mathbf{Z}_k\mathbf{Z}_k',$$

here, $\mathbf{W}_k(r)$ denote a k-vector of independent standard Wiener processes.[5] While the kerenl HAC estimator $\hat{\mathbf{Q}}$ in (8-8) has a nonstochastic limit, $\hat{\mathbf{C}}$ in (8-12) has a random limit depending on $\mathbf{\Lambda}$ and a functional of the Brownian bridge $\mathbf{P}_k$.

Define $\hat{\mathbf{M}} = \left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\hat{\mathbf{C}}^{1/2}$, where $\hat{\mathbf{C}}^{1/2}\hat{\mathbf{C}}'^{1/2} = \hat{\mathbf{C}}$. It follows from (8-4) that

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \underline{\mathbf{Q}}^{-1}\mathbf{Q}_* \underline{\mathbf{Q}}^{-1}) \equiv N(\mathbf{0}, \underline{\mathbf{Q}}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\underline{\mathbf{Q}}^{-1}), \tag{8-13}$$

therefore

$$\begin{aligned}
\hat{\mathbf{M}}^{-1}\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \quad & \xrightarrow{d} N(\mathbf{0}, \mathbf{M}^{-1}\underline{\mathbf{Q}}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\underline{\mathbf{Q}}^{-1}\mathbf{M}'^{-1}) \\
& \equiv N(\mathbf{0}, (\mathbf{C}^{-1/2}\underline{\mathbf{Q}})\underline{\mathbf{Q}}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\underline{\mathbf{Q}}^{-1}(\underline{\mathbf{Q}}\mathbf{C}'^{-1/2})) \\
& = N(\mathbf{0}, \mathbf{C}^{-1/2}\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{C}'^{-1/2}) \\
& = N(\mathbf{0}, \mathbf{Z}_k^{-1}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Lambda}'^{-1}\mathbf{Z}_k'^{-1}) \\
& = \mathbf{Z}_k^{-1}N(\mathbf{0},\ \mathbf{I}).
\end{aligned}$$

This transformation results in a limiting distribution that does not depend on the nuisance parameters.

4.3.1 Tests of Several Linear Restrictions on $\boldsymbol{\beta}$: the F-ratio Test Statistic

Suppose we are interested in testing a general linear hypothesis of the form $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$ against $H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{q}$, where $\mathbf{R}$ is a known $m \times k$ matrix. KVB (2000) suggest using the $\hat{\mathbf{B}} = (\frac{1}{T}\mathbf{X}'\mathbf{X})^{-1}\hat{\mathbf{C}}(\frac{1}{T}\mathbf{X}'\mathbf{X})^{-1}$ to replace $\hat{\mathbf{V}}$ as in (8-9). The Wald test statistic they propose is

$$W^* = T(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'(\mathbf{R}\hat{\mathbf{B}}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}). \tag{8-14}$$

The asymptotic distribution of F^* is stated in the following theorem.

$\boxed{\mathfrak{Theorem}}$.
Under the null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$,

$$W^* \xrightarrow{d} \mathbf{W}_m(1)'\mathbf{P}_m^{-1}\mathbf{W}_m(1). \tag{8-15}$$

[5] See Chapter 21 for a detailed introduction in Wiener Process.

Proof.

Under the null hypothesis, it follows from (8-13) and (8-14) that

$$\begin{aligned} W^* &= T(\mathbf{R}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}))'\left[\mathbf{R}\hat{\mathbf{B}}\mathbf{R}'\right]^{-1}(\mathbf{R}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})) \\ &= (\mathbf{R}T^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}))'\left[\mathbf{R}\hat{\mathbf{B}}\mathbf{R}'\right]^{-1}(\mathbf{R}T^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})) \\ &\xrightarrow{d} (\mathbf{R}\underline{\mathbf{Q}}^{-1}\boldsymbol{\Lambda}\mathbf{W}_k(1))'\left[\mathbf{R}\underline{\mathbf{Q}}^{-1}\boldsymbol{\Lambda}\mathbf{P}_k\boldsymbol{\Lambda}'\underline{\mathbf{Q}}^{-1}\mathbf{R}'\right]^{-1}(\mathbf{R}\underline{\mathbf{Q}}^{-1}\boldsymbol{\Lambda}\mathbf{W}_k(1)). \end{aligned}$$

Denote

$$\mathbf{R}\underline{\mathbf{Q}}^{-1}\boldsymbol{\Lambda}\mathbf{W}_k(1) \equiv \boldsymbol{\Lambda}^*\mathbf{W}_m(1), \tag{8-16}$$

where $\boldsymbol{\Lambda}^*$ is the $(m \times m)$ matrix square root of $\mathbf{R}\underline{\mathbf{Q}}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\underline{\mathbf{Q}}^{-1}\mathbf{R}'$ and hence it is easy to see that[6]

$$\mathbf{R}\underline{\mathbf{Q}}^{-1}\boldsymbol{\Lambda}\mathbf{P}_k\boldsymbol{\Lambda}'\underline{\mathbf{Q}}^{-1}\mathbf{R}' \equiv \boldsymbol{\Lambda}^*\mathbf{P}_m\boldsymbol{\Lambda}'^* \tag{8-17}$$

Substitute (8-16) and (8-17) into (8-15), we complete the proof. ■

Compared with (8-10), W^* does not have a limiting χ^2 distribution, yet it is asymptotically pivotal because the null limit in (8-14) does not depend on the matrix of nuisance parameters, $\boldsymbol{\Lambda}$. Although the asymptotic distribution of W^*, it can be easily simulated.

Construction of the W^* statistic amounts to replacing the HAC estimator, $\hat{\mathbf{Q}}$ with $\hat{\mathbf{C}}$ and using the scaling matrix $\hat{\mathbf{B}}$ in place of the usual scaling matrix $\hat{\mathbf{V}}$. The scaling matrix $\hat{\mathbf{B}}$ converges to a random matrix rather than the fixed variance-covariance matrix.

[6]See the proof at p.713 of KVB (2000).

5 An Example of GLS: SURE model

The seemingly unrelated regression model or SURE model, is a multivariate linear regression model investigated by Zellner (1962). Consider the following two classical linear regression models:

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1,\ E(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1') = \sigma_{11}\mathbf{I}_T,$$

and

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2,\ E(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1') = \sigma_{22}\mathbf{I}_T.$$

where

$$\mathbf{y}_i \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ . \\ . \\ . \\ Y_{iT} \end{bmatrix}_{T\times 1}, \text{ and } \mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \\ . \\ . \\ . \\ \mathbf{x}'_{iT} \end{bmatrix}_{T\times k_i}.$$

The reason for the label "seemingly" unrelated regression should now be clear. Though initially it may appear that the first equation is *not in any way related to the second equation*, in fact there may be random disturbances which are pertinent to both. The common effect of the random disturbances is reflected in the covariance of the two equation's disturbance term. If the disturbances of the above two equations are assumed to be contemporaneously correlated in that

$$E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}, \quad i,j = 1,2; \quad t = 1,2,...,T,$$

then the variance covariance in the combined equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8-18}$$

is

$$\begin{aligned} E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = E\begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}\begin{bmatrix} \boldsymbol{\varepsilon}'_1 & \boldsymbol{\varepsilon}'_2 \end{bmatrix} = E\begin{bmatrix} \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}'_1 & \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}'_2 \\ \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}'_1 & \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}'_2 \end{bmatrix} &= \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \sigma_{12}\mathbf{I}_T \\ \sigma_{21}\mathbf{I}_T & \sigma_{22}\mathbf{I}_T \end{bmatrix} \\ &= \boldsymbol{\Sigma}\otimes\mathbf{I}_T, \end{aligned} \tag{8-19}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

and

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

Clearly, this combined equation does not satisfy the classical assumption since its disturbance variance-covariance, $\mathbf{\Sigma} \otimes \mathbf{I}_T$ in (8-20) is heteroscedastic and autocorrelated. To estimate $\boldsymbol{\beta}$ efficiently, GLS or FGLS is called for.

Let us now generalize the seemingly unrelated regression model to m equations rather than just two and define the standard conditions for the seemingly unrelated regression. These conditions are sufficient to ensure that the seemingly unrelated regression model meets the requirements of generalized least squares estimation. Consider the m regression equations

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \\
\mathbf{y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \\
&. \\
&. \\
&. \\
\mathbf{y}_m &= \mathbf{X}_m\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m,
\end{aligned}$$

where $\mathbf{y}_i$ is of dimension $T \times 1$ and $\mathbf{X}_i$ is $T \times k_i$. We also write them as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, ..., m. \tag{8-20}$$

These m equations can be further written in the combined form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8-21}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ . \\ . \\ . \\ \mathbf{y}_m \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{X}_M \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ . \\ . \\ . \\ \boldsymbol{\beta}_m \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ . \\ . \\ . \\ \boldsymbol{\varepsilon}_m \end{bmatrix}.$$

We assume the SURE model satisfies the following assumptions,

Assumptions

Assume that the seemingly unrelated regression system (8-22) satisfies the conditions:
(a). $E(\boldsymbol{\varepsilon}) = \mathbf{0}$,
(b). $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}_T$ and $\boldsymbol{\Sigma} = [\sigma_{ij}], i, j = 1, 2, ..., m$.
(c). The matrix $\mathbf{X}$ is nonstochastic and $\lim_{T\to\infty} \frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{T}$ is finite and nonsingular. ■

These assumptions are called the standard conditions for seemingly unrelated regression.

5.1 The GLS Estimator

For the present, let us examine the estimation of SURE in the instance where $\boldsymbol{\Omega}$ is assumed known. The most efficient estimator of slope parameters are the GLS.

Theorem.

The BLUE of $\boldsymbol{\beta}$ is just

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}, \qquad (8\text{-}22)$$

with covariance matrix $(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}$.

Proof.

The SURE estimator $\tilde{\boldsymbol{\beta}}$ satisfies the conditions required in generalized least squares estimation. Therefore these results follow directly from the development in Section 8.3.■

This estimator is obviously different from ordinary least squares. The GLS estimator $\tilde{\boldsymbol{\beta}}$ is more efficient, in general, than the OLS estimator. To see this, using partitioned matrix multiplication and the Kronecker product property $(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})$, the GLS's estimator of SURE model can be written as

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} \\
&= \left(\begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \mathbf{0} & . & \mathbf{0} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \mathbf{0} & . & . & \mathbf{0} & \mathbf{X}_m' \end{bmatrix} \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & . & . & . & \sigma^{1m}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & . & . & . & \sigma^{2m}\mathbf{I}_T \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma^{m1}\mathbf{I}_T & . & . & . & . & \sigma^{mm}\mathbf{I}_T \end{bmatrix} \cdot \right. \\
&\quad \left. \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} & . & \mathbf{0} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \mathbf{0} & . & . & \mathbf{0} & \mathbf{X}_m \end{bmatrix} \right)^{-1} \times \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' & \mathbf{0} & . & \mathbf{0} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \mathbf{0} & . & . & \mathbf{0} & \mathbf{X}_m' \end{bmatrix} \cdot \\
&\quad \begin{bmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & . & . & . & \sigma^{1m}\mathbf{I}_T \\ \sigma^{21}\mathbf{I}_T & \sigma^{22}\mathbf{I}_T & . & . & . & \sigma^{2m}\mathbf{I}_T \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma^{m1}\mathbf{I}_T & . & . & . & . & \sigma^{mm}\mathbf{I}_T \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ . \\ . \\ . \\ . \\ \mathbf{y}_m \end{bmatrix}
\end{aligned}
$$

$$
= \begin{bmatrix} \sigma^{11}\mathbf{X}_1'\mathbf{X}_1 & . & . & . & . & \sigma^{1m}\mathbf{X}_1'\mathbf{X}_m \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma^{m1}\mathbf{X}_m'\mathbf{X}_1 & . & . & . & . & \sigma^{mm}\mathbf{X}_m'\mathbf{X}_m \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^m \sigma^{1j}\mathbf{X}_1'\mathbf{y}_j \\ . \\ . \\ . \\ . \\ . \\ \sum_{j=1}^m \sigma^{mj}\mathbf{X}_m'\mathbf{y}_j \end{bmatrix}, \qquad (8\text{-}23)
$$

where σ^{ij} represents the $(i,j)-th$ elements of $\boldsymbol{\Sigma}^{-1}$.

The OLS estimators of SURE model (8-22) is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2 & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{X}_m'\mathbf{X}_m \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{y}_2 \\ . \\ . \\ . \\ \mathbf{X}_m'\mathbf{y}_m \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1 \\ (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}_2 \\ . \\ . \\ . \\ (\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'\mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ . \\ . \\ . \\ \hat{\boldsymbol{\beta}}_m \end{bmatrix}, \qquad (8\text{-}24)$$

where $\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i, i = 1, 2, ..., m$. While The GLS estimator (8-24) links equations by their disturbance, the OLS estimator ignores this linkage since (8-25) represents the individual OLS estimator for the $i - th$ equation from (8-21).

There are, however, two cases in which the GLS and OLS estimators are identical.

(a). If the equations are actually unrelated, that is, if $\sigma_{ij} = 0$ for $i \neq j$. Then there is obviously no payoff to GLS. Indeed, GLS is OLS.

(b). If the equations have identical explanatory variables, that is, if $\mathbf{X}_i = \mathbf{X}_j$, then OLS and GLS are identical.[7] □

These two cases are shown clearly in the following theorems.

Theorem.

If $\sigma_{jk} = 0$ for $j \neq k$ then $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$, and OLS is fully efficient.

Proof.

Set $\sigma_{ij} = 0$ to (8-24), we obtain the results identical to (8-25). ■

[7] A vector autoregressive model (VAR) in *Time Series Analysis* is an example of this case. See Chapter 18.

Therefore under the situation that $\sigma_{ij} = 0$, the equations of the SURE system are "truly" unrelated when the disturbances of the various equations are uncorrelated and nothing is lost by using an estimator which ignores the possibility of contemporaneously correlated disturbance terms.

The other case where the GLS estimator $\tilde{\boldsymbol{\beta}}$ and OLS estimator $\hat{\boldsymbol{\beta}}$ are numerically equivalent and equally efficient when the regressor $\mathbf{X}_i, i = 1, 2, ..., m$ are numerically identical. This result is stated formally in the following theorem.

$\boxed{\text{Theorem}}$.

Consider the set of SURE equations that $\mathbf{X}_i = \mathbf{X}^*$, $\forall i = 1, 2, ..., m$,

$$
\begin{aligned}
\mathbf{y}_1 &= \mathbf{X}^*\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \\
\mathbf{y}_2 &= \mathbf{X}^*\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 \\
&. \\
&. \\
&. \\
\mathbf{y}_m &= \mathbf{X}^*\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m.
\end{aligned}
$$

In this case the OLS estimator is fully efficient in that $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$.

Proof.

In this case, $\mathbf{X}$ in (8-22) can be expressed as

$$
\mathbf{X} = \begin{bmatrix}
\mathbf{X}_1 & \mathbf{0} & . & . & . & \mathbf{0} \\
\mathbf{0} & \mathbf{X}_2 & \mathbf{0} & . & . & \mathbf{0} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
\mathbf{0} & . & . & . & \mathbf{0} & \mathbf{X}_m
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{X}^* & \mathbf{0} & . & . & . & \mathbf{0} \\
\mathbf{0} & \mathbf{X}^* & \mathbf{0} & . & . & \mathbf{0} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
\mathbf{0} & . & . & . & \mathbf{0} & \mathbf{X}^*
\end{bmatrix}
= \mathbf{I}_m \otimes \mathbf{X}^*,
$$

and therefore

$$
\begin{aligned}
\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X} &= (\mathbf{I}_m \otimes \mathbf{X}^*)'(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1}(\mathbf{I}_m \otimes \mathbf{X}^*) \\
&= (\mathbf{I}_M \otimes \mathbf{X}'^*)(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)(\mathbf{I}_m \otimes \mathbf{X}^*) \\
&= (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'^*)(\mathbf{I}_m \otimes \mathbf{X}^*) \\
&= \boldsymbol{\Sigma}^{-1} \otimes (\mathbf{X}'^*\mathbf{X}^*),
\end{aligned}
$$

and

$$\begin{aligned}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} &= (\mathbf{I}_m \otimes \mathbf{X}^*)'(\mathbf{\Sigma} \otimes \mathbf{I}_T)^{-1}\mathbf{y} \\ &= (\mathbf{I}_m \otimes \mathbf{X}'^*)(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{y} \\ &= (\mathbf{\Sigma}^{-1} \otimes \mathbf{X}'^*)\mathbf{y}.\end{aligned}$$

The GLS estimator is therefore

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} \\ &= \left[\mathbf{\Sigma}^{-1} \otimes (\mathbf{X}'^*\mathbf{X}^*)\right]^{-1}\left[(\mathbf{\Sigma}^{-1} \otimes \mathbf{X}'^*)\mathbf{y}\right] \\ &= \left[\mathbf{\Sigma} \otimes (\mathbf{X}'^*\mathbf{X}^*)^{-1}\right]\left[\mathbf{\Sigma}^{-1} \otimes \mathbf{X}'^*\right] \cdot \mathbf{y} \\ &= (\mathbf{I}_m \otimes (\mathbf{X}'^*\mathbf{X}^*)^{-1}\mathbf{X}'^*) \cdot \mathbf{y}\end{aligned}$$

$$= \begin{bmatrix} (\mathbf{X}'^*\mathbf{X}^*)^{-1}\mathbf{X}'^* & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}'^*\mathbf{X}^*)^{-1}\mathbf{X}'^* & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & (\mathbf{X}'^*\mathbf{X}^*)^{-1}\mathbf{X}'^* \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ . \\ . \\ . \\ \mathbf{y}_m \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{X}'^*\mathbf{X}^*)^{-1}\mathbf{X}'^*\mathbf{y_1} \\ (\mathbf{X}'^*\mathbf{X}^*)^{-1}\mathbf{X}'^*\mathbf{y_2} \\ . \\ . \\ . \\ (\mathbf{X}'^*\mathbf{X}^*)^{-1}\mathbf{X}'^*\mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ . \\ . \\ . \\ \hat{\boldsymbol{\beta}}_m \end{bmatrix} = \hat{\boldsymbol{\beta}}. \qquad \blacksquare$$

Note that when the numerical values of the m design matrices are identical, i.e., $\mathbf{X}_1 = \mathbf{X}_2 = ... = \mathbf{X}_m$, this theorem holds regardless of the degree of contemporaneous correlation among the disturbance terms. This result is particular important in the estimation of the Vector Autoregressive Model (VAR) where each individual equation is just the case here to have the same numerical value of regressors.

5.2 The FGLS Estimators

The preceding discussion assumes that $\mathbf{\Sigma}$ is known, which, as usual, is rarely the case. FGLS estimators have been devised. The OLS squared residuals of individual equations residuals may be used to estimated consistently the elements of $\mathbf{\Sigma}$ with

$$\hat{\sigma}_{ij} = \frac{\mathbf{e}_i'\mathbf{e}_j}{T},$$

where

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i,$$

is the residual from OLS estimation of (8-21). A degree of freedom correction in the divisor is suggested:

$$\tilde{\sigma}_{ij} = \frac{\mathbf{e}_i'\mathbf{e}_j}{[(T-k_i)(T-k_j)]}.$$

5.3 An Alternative Formulation of the SURE Model

An alternative way of developing the SURE estimator–which does not involve Kronecker products – is to write the m equations together as[8]

$$\ddot{\mathbf{y}}_t = \ddot{\mathbf{X}}_t\boldsymbol{\beta} + \ddot{\boldsymbol{\varepsilon}}_t, \quad t = 1, 2, ..., T, \tag{8-25}$$

where

$$\ddot{\mathbf{y}}_t = \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ . \\ . \\ . \\ Y_{mt} \end{bmatrix}, \ \ddot{\mathbf{X}}_t = \begin{bmatrix} \mathbf{x}_{1t}' & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2t}' & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{x}_{mt}' \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ . \\ . \\ . \\ \boldsymbol{\beta}_m \end{bmatrix},$$

[8]This is equivalent to equation (8-21).

and

$$\ddot{\boldsymbol{\varepsilon}}_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ . \\ . \\ . \\ \varepsilon_{mt} \end{bmatrix}.$$

Here,

$$E(\ddot{\boldsymbol{\varepsilon}}_t\ddot{\boldsymbol{\varepsilon}}_t') = E\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ . \\ . \\ . \\ \varepsilon_{mt} \end{bmatrix}\begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} & . & . & . & \varepsilon_{mt} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & . & . & . & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & . & . & . & \sigma_{2M} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_{m1} & . & . & . & . & \sigma_{mm} \end{bmatrix} = \boldsymbol{\Sigma}.$$

If the T equations in (8-26) are stacked in the usual way, we have[9]

$$\ddot{\mathbf{y}} = \ddot{\mathbf{X}}\boldsymbol{\beta} + \ddot{\boldsymbol{\varepsilon}}, \tag{8-26}$$

where

$$\ddot{\mathbf{y}} = \begin{bmatrix} \ddot{\mathbf{y}}_1 \\ \ddot{\mathbf{y}}_2 \\ . \\ . \\ . \\ \ddot{\mathbf{y}}_T \end{bmatrix}, \quad \ddot{\mathbf{X}} = \begin{bmatrix} \ddot{\mathbf{X}}_1 \\ . \\ . \\ . \\ \ddot{\mathbf{X}}_T \end{bmatrix}, \quad and \ \ddot{\boldsymbol{\varepsilon}} = \begin{bmatrix} \ddot{\boldsymbol{\varepsilon}}_1 \\ \ddot{\boldsymbol{\varepsilon}}_2 \\ . \\ . \\ . \\ \ddot{\boldsymbol{\varepsilon}}_T \end{bmatrix}.$$

The covariance matrix of the disturbances in the stacked equation is

$$\begin{aligned} E(\ddot{\boldsymbol{\varepsilon}}\ddot{\boldsymbol{\varepsilon}}') &= E\begin{bmatrix} \ddot{\boldsymbol{\varepsilon}}_1 \\ \ddot{\boldsymbol{\varepsilon}}_2 \\ . \\ . \\ . \\ \ddot{\boldsymbol{\varepsilon}}_T \end{bmatrix}\begin{bmatrix} \ddot{\boldsymbol{\varepsilon}}_1' & \ddot{\boldsymbol{\varepsilon}}_2' & . & . & . & \ddot{\boldsymbol{\varepsilon}}_T' \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & . & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & . & \boldsymbol{\Sigma} \end{bmatrix} = \mathbf{I}_T \otimes \boldsymbol{\Sigma} \equiv \boldsymbol{\Lambda}. \end{aligned}$$

[9]This is equivalent to equation (8-22).

The GLS of $\boldsymbol{\beta}$ in this form (8-27) is

$$\tilde{\tilde{\boldsymbol{\beta}}} = (\ddot{\mathbf{X}}'\boldsymbol{\Lambda}^{-1}\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\boldsymbol{\Lambda}^{-1}\ddot{\mathbf{y}}$$

$$= \left(\begin{bmatrix} \ddot{\mathbf{X}}_1' & . & . & . & . & \ddot{\mathbf{X}}_T' \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} & . & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & . & \boldsymbol{\Sigma}^{-1} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{X}}_1 \\ . \\ . \\ . \\ . \\ . \\ \ddot{\mathbf{X}}_T \end{bmatrix} \right)^{-1} .$$

$$\left(\begin{bmatrix} \ddot{\mathbf{X}}_1' & . & . & . & . & \ddot{\mathbf{X}}_T' \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} & . & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & . & \boldsymbol{\Sigma}^{-1} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{y}}_1 \\ . \\ . \\ . \\ . \\ . \\ \ddot{\mathbf{y}}_T \end{bmatrix} \right)$$

$$= \left(\sum_{t=1}^{T} \ddot{\mathbf{X}}_t'\boldsymbol{\Sigma}^{-1}\ddot{\mathbf{X}}_t \right)^{-1} \left(\sum_{t=1}^{T} \ddot{\mathbf{X}}_t'\boldsymbol{\Sigma}^{-1}\ddot{\mathbf{y}}_t \right). \tag{8-27}$$

It is easy to show that $\tilde{\tilde{\boldsymbol{\beta}}}$ in (8-28) is equal to $\tilde{\boldsymbol{\beta}}$ in (8-24). To show this result, it is easy to see that the first bracket in (8-28) can be written as

$$\sum_{t=1}^{T} \ddot{\mathbf{X}}_t'\boldsymbol{\Sigma}^{-1}\ddot{\mathbf{X}}_t$$

$$= \sum_{t=1}^{T} \left(\begin{bmatrix} \mathbf{x}_{1t} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2t} & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{x}_{mt} \end{bmatrix} \begin{bmatrix} \sigma^{11} & \sigma^{12} & . & . & . & \sigma^{1m} \\ \sigma^{21} & \sigma^{22} & . & . & . & \sigma^{2m} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma^{m1} & . & . & . & . & \sigma^{mm} \end{bmatrix} \right.$$

$$
\left.\begin{bmatrix} \mathbf{x}'_{1t} & \mathbf{0} & . & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{x}'_{2t} & \mathbf{0} & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & . & . & . & \mathbf{0} & \mathbf{x}'_{mt} \end{bmatrix}\right)
$$

$$
= \begin{bmatrix} \sigma^{11}\sum_{t=1}^{T}\mathbf{x}_{1t}\mathbf{x}'_{1t} & . & . & . & . & \sigma^{1m}\sum_{t=1}^{T}\mathbf{x}_{1t}\mathbf{x}'_{mt} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma^{m1}\sum_{t=1}^{T}\mathbf{x}_{mt}\mathbf{x}'_{1t} & . & . & . & . & \sigma^{mm}\sum_{t=1}^{T}\mathbf{x}_{mt}\mathbf{x}'_{mt} \end{bmatrix}
$$

$$
= \begin{bmatrix} \sigma^{11}\mathbf{X}'_1\mathbf{X}_1 & . & . & . & . & \sigma^{1m}\mathbf{X}'_1\mathbf{X}_m \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma^{m1}\mathbf{X}'_m\mathbf{X}_1 & . & . & . & . & \sigma^{mm}\mathbf{X}'_m\mathbf{X}_m \end{bmatrix}
$$

$$
= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}) \text{ i.e. the first bracket in (8-24),}
$$

using the fact that

$$
\mathbf{X}'_i\mathbf{X}_j = \begin{bmatrix} \mathbf{x}_{i1} & . & . & . & . & \mathbf{x}_{iT} \end{bmatrix} \begin{bmatrix} \mathbf{x}'_{j1} \\ . \\ . \\ . \\ . \\ \mathbf{x}'_{jT} \end{bmatrix} = \sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}'_{jt}.
$$

A Monkey at the Window of Student's Dorm, NSYSU.

End of this Chapter