Economics 361
Generalized Least Squares

Jun Ishii *
Department of Economics
Amherst College

Fall 2023

1 Overview

The properties of the Ordinary Least Squares (OLS) model were derived assuming the Gauss-Markov assumptions

- Linearity Condition: $E[Y|X] = X\beta$
- Spherical Errors: $\text{Var}(Y|X) = \sigma^2 I_N$
- Full (Column) Rank: rank(X) = k

The above assumptions are necessary and sufficient to show that $b^{ols} = (X'X)^{-1}X'Y$ is the "best" (minimum variance) linear unbiased estimator of β.

Combining the above with

- Normality Assumption: $Y|X \sim N(X\beta, \sigma^2 I)$

we can further show that $b^{ols} \sim N(\beta, \sigma^2(X'X)^{-1})$. This result was used to derive the distribution of the t and z test statistics.

Consider now the world where all of the above four assumptions hold *except* Spherical Errors.

Note that this implies that the available size N sample is *not* random. A random sample naturally satisfies the Spherical Errors assumption as ...

- ... *independently* distributed implies that $\text{Cov}(Y_i, Y_j|X) = 0$ for $i \neq j$
- ... *identically* distributed implies that $\text{Var}(Y_i|X) = \text{Var}(Y_j|X)$ for $i, j = 1, \ldots, N$

*Office: Converse Hall 315 Phone: (413) 542-2901 E-mail: jishii@amherst.edu

Why might we be faced with a *non-random* sample?

- The data we have available is more likely to make up a random sample if they are obtained from *controlled experiments*. With controlled experiments, the value of X can be fixed, the process by which the outcomes (Y_i) is generated made identical, and each outcome ensured to be unaffected by other outcomes.

- But when the data is obtained from *observing* the "regular" outcomes of society and nature, the data need not be random. The outcomes may be generated from non-identical processes and may influence each other.

The former type of data is known as *experimental data* and the latter *observational data*. Unfortunately, most data used in economics (and the social sciences) fall into the second category.

Even if we do not have a random sample, we usually assume that our sample is such that $\text{Var}(Y|X)$ is well-defined. Notationally, we say that

- $\text{Var}(Y|X) = \Sigma$

where Σ is some symmetric, invertible $(N \times N)$ matrix. Spherical errors is the special case where $\Sigma = \sigma^2 I_N$.

More generally, Σ may have non-zero off-diagonal elements (non-zero covariance) and/or diagonal elements that differ from each other (non-identical variance).

$$\Sigma \quad = \quad \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1N} \\ \vdots & \ddots & \vdots \\ \Sigma_{N1} & \cdots & \Sigma_{NN} \end{pmatrix}$$

Σ_{ii} represents the i^{th} diagonal element and the $\text{Var}(Y_i|X)$. Σ_{ij} represents the i^{th} row j^{th} column (or, by symmetry, i^{th} column j^{th} row) element and the $\text{Cov}(Y_i, Y_j|X)$.

The violation of Spherical Errors caused by non-identical diagonal elements – non-identical conditional variances – is referred to as **heteroskedasticity** (as opposed to **homoskedasticity**). The violation caused by non-zero off-diagonal elements – non-zero conditional covariances – is referred to as **autocorrelation**.

As we still have the linearity and full (column) rank conditions, $E[b^{ols}|X] = \beta$. But $\text{Var}(b^{ols}|X)$ may no longer be $\sigma^2(X'X)^{-1}$. It is generally

$$\begin{aligned} \text{Var}(b^{ols}|X) \quad &= \quad \text{Var}(\underbrace{(X'X)^{-1}X'Y}_{AY}|X) \quad = \quad \underbrace{(X'X)^{-1}X'}_{A}\text{Var}(Y|X)\underbrace{X(X'X)^{-1}}_{A'} \\ &= \quad A\Sigma A' \quad = \quad (X'X)^{-1}X'\Sigma X(X'X)^{-1} \end{aligned}$$

The above equals $\sigma^2(X'X)^{-1}$ only if $\Sigma = \sigma^2 I_N$

This implies that b^{ols}, without spherical errors but with Normality Assumption, is distributed

$$b^{ols}|X \quad \sim \quad N(\beta,\ (X'X)^{-1}X'\Sigma X(X'X)^{-1})$$

2 "Known" Σ

Suppose we know Σ. Rather, we know Ω where $\Sigma = \sigma^2\Omega$.[1] Note that this is precisely the case when we have Spherical Errors. Under Spherical Errors, we "know" that $\Omega = I_N$. Under this situation, we have two main approaches to estimating β.

2.1 Use OLS

The first approach is to use the OLS model, as before. As shown earlier, if we still maintain the other two Gauss-Markov assumptions and the Normality Assumption,

$$b^{ols}|X \quad \sim \quad N(\beta,\ (X'X)^{-1}X'\Sigma X(X'X)^{-1})$$

This implies that, from an operational perspective, the only difference is that we use $(X'X)^{-1}X'\Sigma X(X'X)^{-1}$ instead of $\sigma^2(X'X)^{-1}$ as our conditional variance/covariance matrix for b^{ols}.

In the case where we know Ω but not $\Sigma = \sigma^2\Omega$

$$(X'X)^{-1}X'\Sigma X(X'X)^{-1} \quad = \quad (X'X)^{-1}X'\underbrace{\sigma^2\Omega}_{\Sigma}X(X'X)^{-1} \quad = \quad \underbrace{\sigma^2}_{\text{unknown}}\ \underbrace{(X'X)^{-1}X'\Omega X(X'X)^{-1}}_{\text{known}}$$

This is analogous to the case where σ^2 is unknown for the OLS model with Spherical Errors. And the "solution" is the same: use s^2 as your estimate of σ^2

$$s^2 \quad = \quad \frac{1}{N-k}\sum_{i=1}^{N}(\underbrace{Y_i - X_i b^{ols}}_{e_i})^2 \quad = \quad \frac{1}{N-k}\sum_{i=1}^{N} e_i^2$$

And therefore, your estimate of $\mathrm{Var}(b^{ols}|X)$ is

$$\hat{\mathrm{Var}}(b^{ols}) \quad = \quad s^2\ (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

With respect to Hypothesis Testing

- If σ^2 and Ω are both known, use the z-statistic which is distributed $N(0,1)$
- If only Ω is known, use the t-statistic which is distributed t_{N-k}

As we will see, this is not the ideal approach. As the Spherical Errors assumption is violated, the Gauss Markov Theorem no longer holds. OLS provides a linear unbiased estimator of β but there may be a different linear unbiased estimator that achieves lower variance. This, in turns, suggests that hypothesis tests based on this alternative estimator may achieve better power properties.

[1] Note that we can always decompose $\Sigma = \sigma^2\Omega$ for some non-zero σ^2 as $\Omega \equiv \frac{1}{\sigma^2}\Sigma$

2.2 Use GLS

The key problem is that the sample $\{Y, X\}$ is no longer random. If the sample were random, Spherical Errors would hold and the Gauss-Markov Theorem would apply. So, consider the following problem: "How can we transform Y and X such that the transformed sample, $\{\tilde{Y}, \tilde{X}\}$ is random?"

There are two ways in which $\{Y, X\}$ may not be random:

- **Heteroskedasticity**: Diagonal elements of Ω (and thus of Σ) differ – implying that the sample is not identically distributed
- **Autocorrelation**: Off-diagonal elements of Ω (and thus of Σ) are not zero – implying that the sample is not independently distributed

So we need to transform $\{Y, X\}$ such that $\text{Var}(\tilde{Y}_i|\tilde{X})$ are the same for $i = 1, \ldots, N$ and that the $\text{Cov}(\tilde{Y}_i, \tilde{Y}_j|\tilde{X})$ are zero for $i \neq j$ and $i, j = 1, \ldots, N$

Generally, we can express the transformations as

$$\tilde{Y} = HY \qquad \tilde{X} = HX$$

where H is a $(N \times N)$ matrix of known constants

This suggests that we want to choose H such that

$$\begin{aligned} \text{Var}(\tilde{Y}|\tilde{X}) &= \text{Var}(HY|HX) = \text{Var}(HY|X) \\ &= H \underbrace{\text{Var}(Y|X)}_{\Sigma=\sigma^2\Omega} H' = \sigma^2 H\Omega H' = \sigma^2 I_N \end{aligned}$$

So we want $H\Omega H' = I_N$

Let us take a step back and think about Σ. Σ is a symmetric $(N \times N)$ matrix that is also invertible (non-singular). This implies that we can do "LDU" factorization of Σ. But more importantly, this implies that $\Omega = \frac{1}{\sigma^2}\Sigma$ is also a symmetric $(N \times N)$ matrix that is invertible (non-singular). So "LDU" factorization applies to Ω. And, most importantly for our immediate purpose, this implies that "LDU" factorization applies to Ω^{-1}:

$$\Omega^{-1} = LDU = LDL'$$

Ω^{-1} can be expressed as the product of three $(N \times N)$ matrices

- L: $(N \times N)$ matrix whose diagonal elements $= 1$ and whose "top triangle" off-diagonal elements are zero
- D: $(N \times N)$ diagonal matrix (off-diagonal elements $= 0$)
- U: $(N \times N)$ matrix whose diagonal elements $= 1$ and whose "bottom triangle" off-diagonal elements are zero

Moreover, for a symmetric matrix, $U = L'$.

Furthermore, let $D^{\frac{1}{2}}$ be the $(N \times N)$ diagonal matrix whose diagonal elements are the *square root* of the diagonal elements of D. So $D^{\frac{1}{2}} = (D^{\frac{1}{2}})'$ and $D^{\frac{1}{2}} D^{\frac{1}{2}} = D^{\frac{1}{2}} (D^{\frac{1}{2}})' = D$ and

$$\Omega^{-1} \quad = \quad \underbrace{LD^{\frac{1}{2}}}_{(\Omega^{-\frac{1}{2}})'} \underbrace{(D^{\frac{1}{2}})'L'}_{\Omega^{-\frac{1}{2}}}$$

Consider $H = \Omega^{-\frac{1}{2}} = (D^{\frac{1}{2}})'L'$. We can show that

$$\begin{aligned} H\Omega H' \quad = \quad \Omega^{-\frac{1}{2}}\ \Omega\ (\Omega^{-\frac{1}{2}})' \quad &= \quad (D^{\frac{1}{2}})'L'\ (LDL')^{-1}\ LD^{\frac{1}{2}} \\ &= \quad (D^{\frac{1}{2}})'L'\ ((D^{\frac{1}{2}})'L')^{-1} (LD^{\frac{1}{2}})^{-1}\ LD^{\frac{1}{2}} \quad = \quad I_N \end{aligned}$$

So we want to transform the sample using $H = \Omega^{-\frac{1}{2}}$. Note that $H'H = \Omega^{-1}$.

If the only violation of the Spherical Errors is heteroskedasticity (no autocorrelation), then

- L is I_N
- D is the $(N \times N)$ diagonal matrix whose diagonal elements are $\frac{1}{\Omega_{ii}}$ with $\Omega_{ii} = \frac{1}{\sigma^2}\text{Var}(Y_i|X)$
  Note: if Σ is fully known, then this is equivalent to $\Omega = \Sigma$ and $\sigma^2 = 1$
- $\Omega^{-\frac{1}{2}}$ is a diagonal matrix whose diagonal elements are $\frac{1}{\sqrt{\Omega_{ii}}}$

This is the "**Weighted Least Squares**" (WLS) model introduced earlier.

Discussion of the actual "calculation" of $H = \Omega^{-\frac{1}{2}}$ is left for a proper linear algebra course. For the purposes of this course, it is sufficient that H exists and that H can be solved for simple cases, such as the case of "pure" heteroskedasticity (no autocorrelation).

So, applying OLS to the transformed data

$$b^{gls} \quad = \quad (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} \quad = \quad (X'H'HX)^{-1}X'H'HY \quad = \quad (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

And the conditional distribution of b^{gls} is

$$b^{gls}|X \quad \sim \quad N(\beta,\ \underbrace{(X'\Sigma^{-1}X)^{-1}}_{\sigma^2(\tilde{X}'\tilde{X})^{-1}})$$

What is the **intuition** underlying GLS?

Consider the very origin of the least squares model. The OLS model is the model that selects as its estimator the one that minimizes the **sum of squared residuals** (SSR)

$$b^{ols} \quad \equiv \quad \text{argmin}_b \sum_{i=1}^{N} \underbrace{(Y_i - X_i b)^2}_{e_i^2}$$

Gauss proposed the least squares method as solution to the line-fitting problem: choose the line that generates the smallest sum of squared residuals (squared so that over- and under-fitting are symmetrically treated) where the residual (in a regression context) is $e_i = Y_i - X_i b^{ols}$.

So a one unit squared residual in observation i is treated the same as a one unit squared residual in observation j. What matters is the sum. This is fine as long as

1. Each observation contains the same amount of information (identically distributed)

2. Each observation contains information not contained in other observations (independently distributed)

If [1] does not hold, then we should try to minimize the squared residual from a more informative observation more than the squared residual from a less informative observation. If [2] does not hold, then reducing the squared residual from two related observations (with overlapping information) is less desirable than reducing the squared residual from two unrelated observations.

The above "correction" is precisely what GLS does. Note that GLS is OLS applied to the transformed data. So we can re-define b^{gls} as

$$\begin{aligned} b^{gls} &\equiv \text{argmin}_b \sum_{i=1}^{N} \underbrace{\left(\tilde{Y}_i - \tilde{X}_i b\right)^2}_{\tilde{e}_i^2} \\ &\equiv \text{argmin}_b \underbrace{(\tilde{Y} - \tilde{X}b)'(\tilde{Y} - \tilde{X}b)}_{\tilde{e}'\tilde{e}} \qquad \text{in matrix notation} \end{aligned}$$

So the GLS model minimizes the sum of squared residuals associated with the transformed data. But note that

$$\tilde{e} = \tilde{Y} - \tilde{X}b = HY - HXb = H(Y - Xb) = He$$

and therefore

$$\tilde{e}'\tilde{e} = e'H'He = e'\Omega^{-1}e$$

So the sum of squared residuals for GLS is the sum of squared residuals for OLS but with each observation **weighted by** Ω^{-1}. Observations with higher variance and much (positive/negative) covariance with other observations will be weighted less. Observations with lower variance and less (positive/negative) covariance will be weighted more.

Generalized Least Squares is essentially Gauss' line fitting solution, except with observations weighted by their information content.

3 "Unknown" Σ

Suppose even Ω is unknown. We do not have any information about Σ other than that it is a symmetric, invertible $(N \times N)$ matrix. What options do we have?

As long as the Linearity and Full Rank assumptions hold, b^{ols} is still a linear unbiased estimator of β. Not knowing Ω does not alter this result. However, we can no longer derive the exact conditional distribution of b^{ols}, even with the Normality assumption. This is because the conditional variance of b^{ols} depends on the unknown Ω. This, in turn, suggests that hypothesis testing is problematic when Ω is unknown.

Analogous to the earlier situation when σ^2 was unknown, an estimate of Σ must be obtained in order to conduct hypothesis testing, when Σ is unknown.

3.1 Estimating "Σ"

The literature that discusses approaches to estimating Σ is outside the scope of this course. In practice, Σ is rarely directly estimated. What is estimated, instead, is the "asymptotic variances" (such as that of b^{ols}) that is based on Σ. Here, we discuss – very loosely and heuristically – the "intuition" underlying efforts to estimate Σ and/or relevant variances based on Σ.

The key insight lies with the relationship between the regression "error" and the regression residual

- Regression Error: $\epsilon_i \equiv Y_i - X_i\beta$
- Regression Residual: $e_i \equiv Y_i - X_i b$

The regression error is a concept associated with the **population** and the regression residual with the **sample**. Very loosely speaking, e_i can be thought of as the sample analog to ϵ_i, with β replaced by its sample analog, the regression coefficient b.

Note the following relationship between statements about Y and about ϵ

$$\begin{aligned} E[Y|X] = X\beta \quad &\Longleftrightarrow \quad E[\epsilon|X] = 0 \\ \mathrm{Var}[Y|X] = \Sigma \quad &\Longleftrightarrow \quad \mathrm{Var}[\epsilon|X] = \Sigma \end{aligned}$$

Combining the above two statements, we have that $E[\epsilon\epsilon'|X] = \Sigma$. Or, in non-matrix format, $\mathrm{Var}(\epsilon_i|X) = E[\epsilon_i^2|X]$ and $\mathrm{Cov}(\epsilon_i, \epsilon_j|X) = E[\epsilon_i\epsilon_j|X]$

This suggests that

- e_i^2 as the "sample analog" to ϵ_i^2 must contain some information about $E[\epsilon_i^2|X] = \mathrm{Var}(\epsilon_i|X)$
- e_ie_j as the "sample analog" to $\epsilon_i\epsilon_j$ must contain some information about $E[\epsilon_i\epsilon_j|X] = \mathrm{Cov}(\epsilon_i, \epsilon_j|X)$

The key requirement is that $E[e_i|X] = 0$, analogously to $E[\epsilon_i|X] = 0$. This, essentially, requires b to be an unbiased estimate of β.[2] In some situations, this requirement is "weakened" to b being a

[2] $E[e_i|X] = E[Y_i - X_ib|X] = X\beta - X_iE[b|X]$. Hence, need $E[b|X] = \beta$ in order for $E[e_i|X] = 0$

consistent estimate of $\beta : b \xrightarrow{p} \beta$.[3]

This insight can be used to develop *consistent* estimates of the *asymptotic* variance of b^{ols}. Currently, there is no estimate of the *finite sample* variance of b^{ols} that can be obtained analytically.[4]

Recall from the "Estimation" Handout (Section 3.3 "OLS Model under Asymptotic Normality") that if all three Gauss-Markov assumptions are satisfied, the sample is random, and some limit conditions are satisfied for matrices $X'X$ and $X'Y$

$$b^{ols} \quad \overset{a}{\sim} \quad N(\beta, \frac{\sigma^2}{N} Q_{XX}^{-1})$$

In other words, the conditional distribution of b^{ols} as $N \to \infty$ is "said to be" multivariate Normal with mean β and variance $\frac{\sigma^2}{N} Q_{XX}^{-1}$.

A similar result can be derived when Spherical Errors is replaced with $\text{Var}(Y|X) = \Sigma$

$$b^{ols} \quad \overset{a}{\sim} \quad N(\beta, \frac{1}{N} Q_{XX}^{-1} Q_{X'\Sigma X} Q_{XX}^{-1})$$

where $Q_{X'\Sigma X} = \lim_{N\to\infty} \frac{X'\Sigma X}{N}$

Q_{XX} can be estimated with $\frac{1}{N}(X'X)$ and $Q_{X'\Sigma X}$ can be estimated using the regression residuals.

Estimates of this asymptotic variance (rather, their "square root") are referred to as **"robust" standard errors** or, more properly, **"sandwich estimates"** of the asymptotic standard errors.[5] These estimates are jointly credited to two statisticians, F. Eicker and P.J. Huber, and an econometrician, H. White.[6]

Most modern econometrics software will calculate the "general" version of these estimated variances. Two famous "specialized" versions of these "robust" standard errors are

- the Eicker-Huber-White "Heteroskedasticity-Robust (or Consistent)" Standard Errors
- the Newey-West "Autocorrelation-Robust (or Consistent)" Standard Errors

Consider the Eicker-Huber-White version

$$\frac{1}{N} \underbrace{\left(\frac{1}{N}(X'X)\right)^{-1}}_{\hat{Q}_{XX}^{-1}} \underbrace{\frac{1}{N}\sum_{i=1}^{N} e_i^2 x_i x_i'}_{\hat{Q}_{X'\Sigma X}} \underbrace{\left(\frac{1}{N}(X'X)\right)^{-1}}_{\hat{Q}_{XX}^{-1}} \xrightarrow{p} \frac{1}{N} Q_{XX}^{-1} Q_{X'\Sigma X} Q_{XX}^{-1}$$

where x_i' is a $(1 \times k)$ vector representing the i^{th} row of X and $e_i = Y_i - X_i b^{ols}$.

The above version is used when Σ is believed to have non-identical diagonal elements (heteroskedasticity) but zero off-diagonal elements (no autocorrelation).

[3] Technically, unbiasedness does not imply consistency nor does consistency imply unbiasedness

[4] There are some possible simulation-based methods. This topic is outside the scope of the course

[5] The sandwich refers to the variance taking the matrix form of ABA' – B sandwiched between A

[6] White formally developed these estimates in the early 1980s using late 1960s results by Eicker and Huber.

3.2 Using the Estimated "Σ"

In the earlier section, we discussed how the asymptotic (conditional) variance of b^{ols} can be obtained using a "robust" or "sandwich" approach:

$$\text{Estimated Asymptotic Var}(b^{ols}|X) \;=\; \frac{1}{N}\,\hat{Q}_{XX}^{-1}\,\hat{Q}_{X'\Sigma X}\,\hat{Q}_{XX}^{-1}$$

Note that

$$\frac{1}{N}\,\hat{Q}_{XX}^{-1}\,\hat{Q}_{X'\Sigma X}\,\hat{Q}_{XX}^{-1} \xrightarrow{p} \frac{1}{N}\,Q_{XX}^{-1}\,Q_{X'\Sigma X}\,Q_{XX}^{-1}$$

So as $N \to \infty$, Estimated Asymptotic Var$(b^{ols}|X)$ effectively collapses to the actual Asymptotic Var$(b^{ols}|X)$. This result depends on the Law of Large Numbers (LLN).

This estimate can be used to conduct *asymptotic* hypothesis tests: hypothesis tests involving the distribution of the test statistic as the sample size approaches infinity $(N \to \infty)$

For linear hypothesis tests $H_0 : R\beta - r = 0$

$$z \;=\; \frac{Rb^{ols} - r}{\sqrt{R\,\frac{1}{N}\,\hat{Q}_{XX}^{-1}\,\hat{Q}_{X'\Sigma X}\,\hat{Q}_{XX}^{-1}\,R'}} \;\overset{a}{\sim}\; N(R\beta - r, 1)$$

The result above depends on the Central Limit Theorem (CLT).

The idea here is that while the *finite sample* distribution of z cannot be derived, we can derive the distribution for the *asymptotic sample.* This asymptotic distribution of z is used to **approximate** the finite sample distribution of z. This approximation is "justified" when N is a "large" number.

Frankly, the "asymptotic" hypothesis test is something of a desperate move. Ideally, we would conduct the hypothesis test based on the *exact*, finite sample distribution of the test statistic. However, in some cases, like when Ω is unknown, this desired exact, finite sample distribution is unknown. We can only, at best, derive the asymptotic distribution.[7]

Asymptotic hypothesis testing is not the only use of the estimated "Σ." It may be possible to estimate the GLS estimate itself

$$\begin{aligned} b^{FGLS} \;&=\; \hat{Q}_{X'\Sigma^{-1}X}\hat{Q}_{X'\Sigma^{-1}Y} \\ \text{where}\;\; & \hat{Q}_{X'\Sigma^{-1}X} \xrightarrow{p} Q_{X'\Sigma^{-1}X} \qquad \hat{Q}_{X'\Sigma^{-1}Y} \xrightarrow{p} Q_{X'\Sigma^{-1}Y} \\ & Q_{X'\Sigma^{-1}X} \equiv \lim_{N\to\infty} \frac{X'\Sigma^{-1}X}{N} \qquad Q_{X'\Sigma^{-1}Y} \equiv \lim_{N\to\infty} \frac{X'\Sigma^{-1}Y}{N} \end{aligned}$$

This asymptotically valid estimate of the GLS estimator is referred to as the **Feasible GLS** (FGLS) estimator. b^{fgls} has the following *asymptotic* distribution

$$b^{fgls} \;\overset{a}{\sim}\; N(\beta,\; \tfrac{1}{N}Q_{X\Sigma^{-1}X}^{-1})$$

[7] There has been some work studying when N is "sufficiently large enough" to merit the asymptotic approximation. But the literature is still developing.

Unlike the case for GLS, it is not clear that b^{fgls} is a superior estimator to b^{ols}. b^{fgls} is an approximation to b^{gls} for finite samples. Without knowing how close an approximation b^{fgls} is to b^{gls}, the relative merits of b^{fgls} and b^{ols} in a finite sample is not known.

In summary, if Σ is unknown, the options are

1. Use OLS and some robust standard error

   "Default" approach

2. Use FGLS and corresponding asymptotic distribution

   Possibly better approach, if N is suitably large and/or $\hat{\Sigma}$ accurate

In both approaches, "asymptotic" hypothesis tests would need to be used to test any related Null Hypotheses. For OLS, the test statistic would be derived using the "robust" standard errors. For FGLS, the test statistic would be derived using the asymptotic distribution of FGLS.

3.3 Testing Σ

There is an uncountable number of ways in which $\Sigma = \sigma^2 I_N$ may be violated. That has not, however, prevented researchers from developing hypothesis tests of $H_0 : \Sigma = \sigma^2 I_N$. The difficulty lies with defining the allowable alternative hypothesis.

There are currently no test with $H_a : \Sigma \neq \sigma^2 I_N$ with acceptable known power properties. There are, however, tests involving more restricted composite alternatives that have better known power properties.

For example, consider $H_a : \Sigma = \sigma^2 D_Z$ where D_Z is a $(N \times N)$ diagonal matrix consisting of diagonal elements $D_{ii} = Z_i \alpha$. Z_i is a $(1 \times k_z)$ vector of variables whose values are known and α a $(k_z \times 1)$ vector of unknown parameters (constant across the observations). Asymptotic hypothesis tests of $H_0 : \Sigma = \sigma^2 I_N$ versus $H_a : \sigma^2 D_Z$ have been developed, independently, by T.S. Breusch and A. R. Pagan, and by L. Godfrey.[8]

These hypothesis tests usually center on the residuals obtained from applying OLS to the data. The test statistics usually have an asymptotic distribution of χ^2 with some degree of freedom (depends on the actual test statistic).

The details of testing Σ will be left for later in this course (or another course).

[8] "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 1979, Vol. 47, pp.1287-1294 & "Testing for Multiplicative Heteroscedasticity," *Journal of Econometrics*, 1978, Vol. 8, pp.227-236