Topic 6: Non-Spherical Disturbances

Our basic linear regression model is

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad ; \qquad \boldsymbol{\varepsilon} \sim N[\boldsymbol{0}\,, \sigma^2 I_n]$$

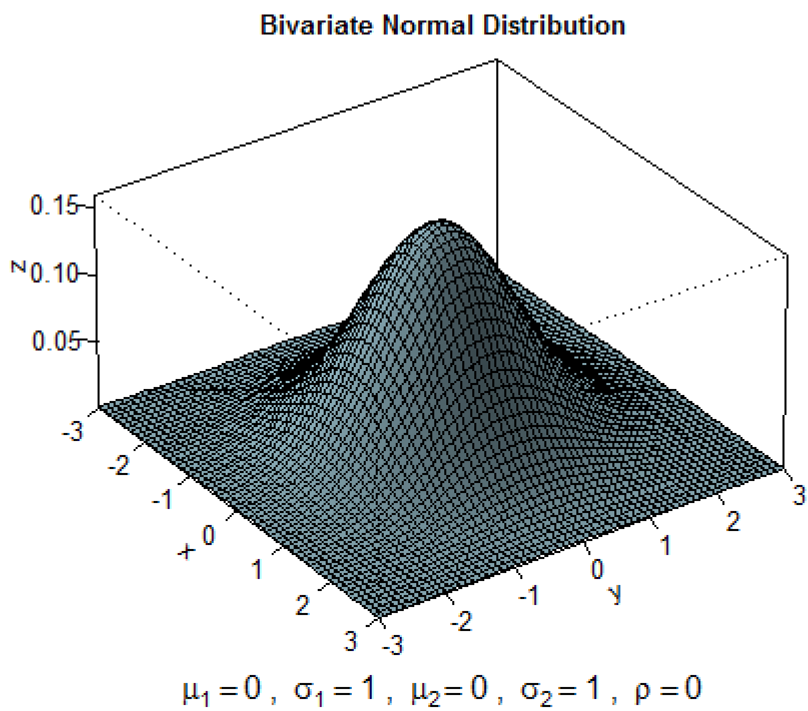
Now we'll generalize the specification of the error term in the model:

$$E[\boldsymbol{\varepsilon}] = \boldsymbol{0} \quad ; \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \Sigma = \sigma^2\Omega \quad ; \qquad \text{(\& Normal)}$$

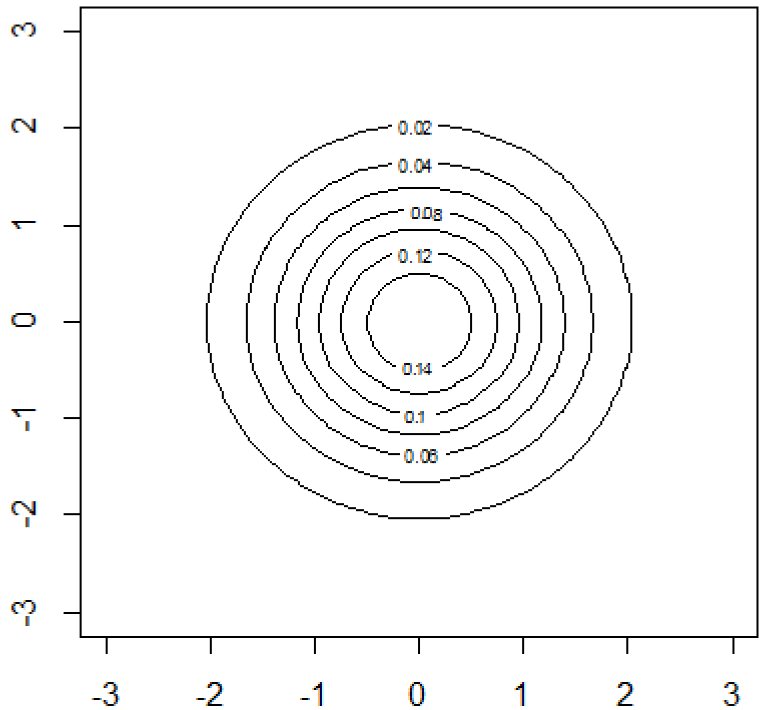
This allows for the possibility of one or both of

- Heteroskedasticity
- Autocorrelation (Cross-section; Time-series; Panel data)

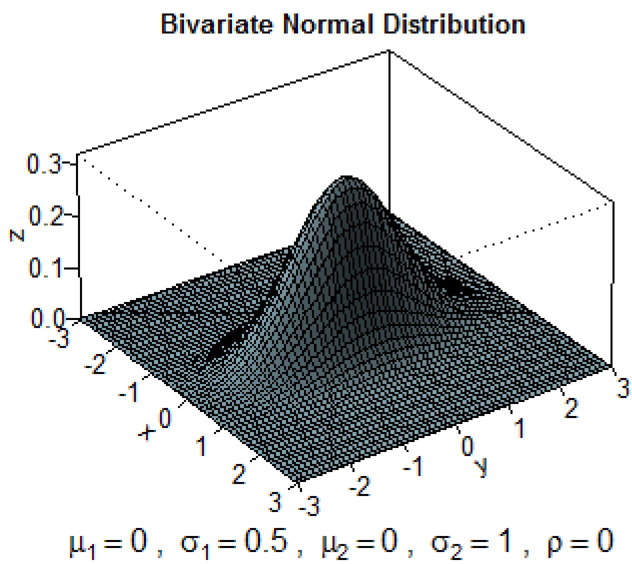
Spherical Disturbances – Homoskedasticity and Non-Autocorrelation

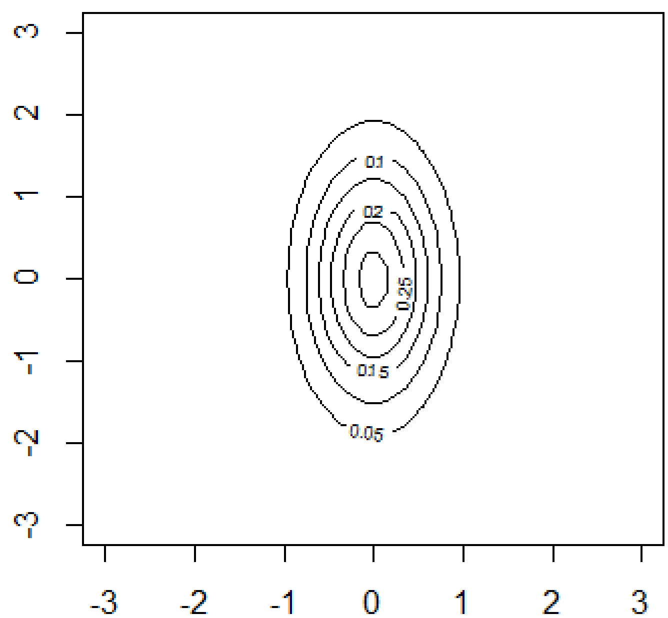


In the above, consider $x = \varepsilon_i$ and $y = \varepsilon_j$. The joint probability density function, $p(\varepsilon_i, \varepsilon_j)$, is in the direction of the z axis. Below is a contour of the above perspective. If we consider the joint distribution of *three* error terms instead of *two*, the circles below would become spheres, hence the terminology "spherical disturbances."

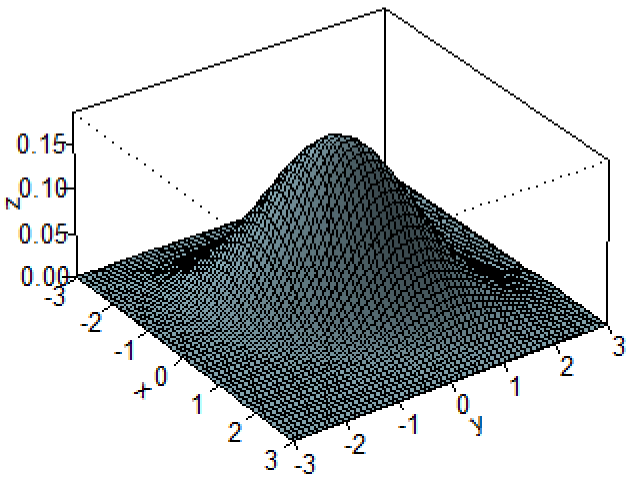


Non-Spherical Disturbances – Heteroskedasticity and Non-Autocorrelation

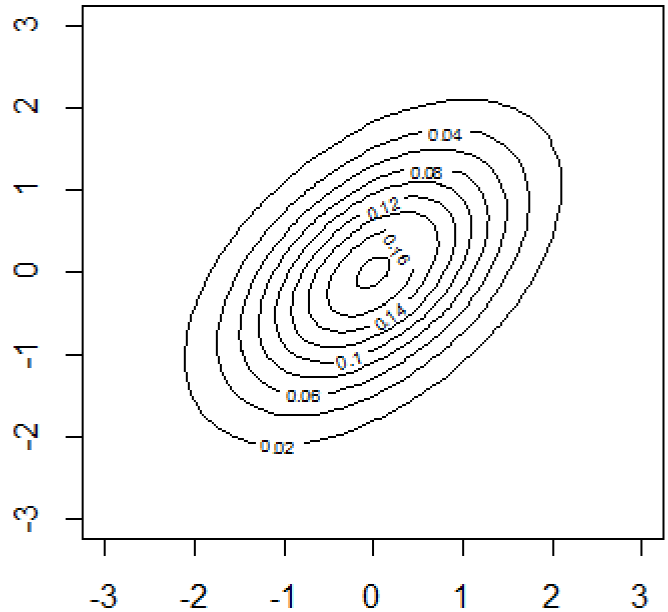




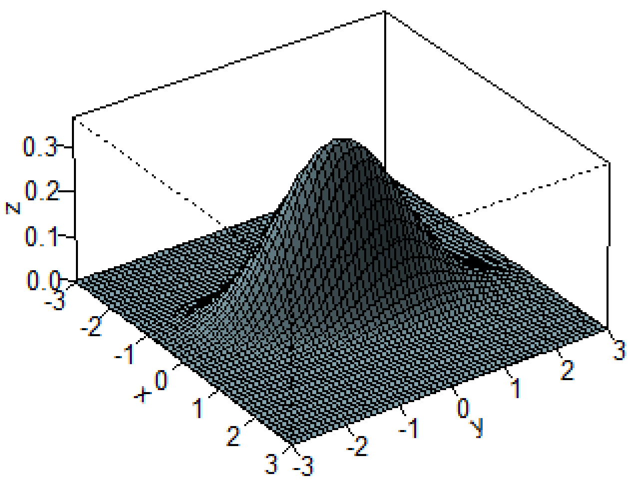
Non-Spherical Disturbances – Homoskedasticity and Autocorrelation



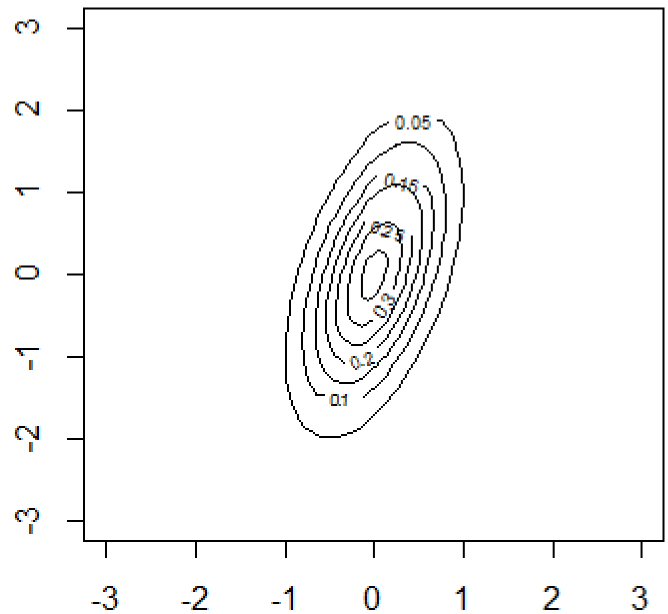
$\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 1$, $\rho = 0.5$



Non-Spherical Disturbances – Heteroskedasticity and Autocorrelation



$\mu_1 = 0$, $\sigma_1 = 0.5$, $\mu_2 = 0$, $\sigma_2 = 1$, $\rho = 0.5$



- How does the more general situation of non-spherical disturbances affect our (Ordinary) Least Squares estimator?
- In particular, let's first look at the sampling distribution of $\boldsymbol{b}$:

$$\boldsymbol{b} = (X'X)^{-1}X'\boldsymbol{y} = (X'X)^{-1}X'(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$= \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon} .$$

So,

$$E(\boldsymbol{b}) = \boldsymbol{\beta} + (X'X)^{-1}X'E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta} \ .$$

The more general form of the covariance matrix for the error term does not alter the fact that the OLS estimator is ***unbiased***.

- Next, consider the covariance matrix of our OLS estimator in this more general situation:

$$V(\boldsymbol{b}) = V[\boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon}] = V[(X'X)^{-1}X'\boldsymbol{\varepsilon}]$$

$$= [(X'X)^{-1}X'V(\boldsymbol{\varepsilon})X(X'X)^{-1}]$$

$$= [(X'X)^{-1}X'\sigma^2\Omega X(X'X)^{-1}]$$

$$\neq [\sigma^2(X'X)^{-1}] \ .$$

- So, under our full set of modified assumptions about the error term:

$$\boldsymbol{b} \sim N[\boldsymbol{\beta}\,, V^*]$$

where

$$V^* = \sigma^2[(X'X)^{-1}X'\Omega X(X'X)^{-1}] \, .$$

- So, the usual computer output will be misleading, *numerically*, as it will be based on using the wrong formula, namely $s^2(X'X)^{-1}$.
- The standard errors, t-statistics, *etc*. will all be incorrect.
- As well as being *unbiased*, the OLS point estimator of $\boldsymbol{\beta}$ will still be ***weakly consistent***.
- The I.V. estimator of $\boldsymbol{\beta}$ will still be ***weakly consistent***.

- The NLLS estimator of the model's parameters will still be ***weakly consistent***.
- **However**, the usual estimator for the covariance matrix of ***b***, namely $s^2(X'X)^{-1}$, will be an *inconsistent estimator* of the true covariance matrix of ***b***!
- This has serious implications for inferences based on confidence intervals, tests of significance, *etc*.
- So, we need to know how to deal with these issues.
- This will lead us to some *generalized estimators*.
- First, let's deal with the most pressing issue – the inconsistency of the estimator for the covariance matrix of ***b***.

White's Heteroskedasticity-Consistent Covariance Matrix Estimator

- If we knew $\sigma^2\Omega$, then the "estimator" of the covariance matrix for ***b*** would just be:

$$V^* = [(X'X)^{-1}X'\sigma^2\Omega X(X'X)^{-1}]$$

$$= \frac{1}{n}\left[\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\sigma^2\Omega X\right)\left(\frac{1}{n}X'X\right)^{-1}\right]$$

$$= \frac{1}{n}\left[\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\Sigma X\right)\left(\frac{1}{n}X'X\right)^{-1}\right]$$

- If Σ is *unknown*, then we need to find a consistent estimator of $\left(\frac{1}{n}X'\Sigma X\right)$.
- (Why not an estimator of just Σ ?)
- Note that at this stage of the discussion, the form of the Σ matrix is quite arbitrary.
- Let $Q^* = \left(\frac{1}{n}X'\Sigma X\right)$ $(k \times k)$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\boldsymbol{x}_i\boldsymbol{x}_j'$$

$(k \times 1)$ $(1 \times k)$

- In the case of *heteroskedastic errors*, things simplify, because $\sigma_{ij} = 0$, for $i \neq j$.

Then, we have

$$Q^* = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$$

- White (1980) showed that if we define

$$S_0 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$$

  Then , $plim(S_0) = Q^*$.

- This means that we can estimate the model by OLS; get the associated residual vector, ***e*** ; and then a consistent estimator of V^*, the covariance matrix of ***b***, will be:

$$\hat{V}^* = \frac{1}{n}\left[\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'\right)\left(\frac{1}{n}X'X\right)^{-1}\right]$$

  or,

$$\hat{V}^* = n[(X'X)^{-1}S_0(X'X)^{-1}] .$$

- $\hat{V}^*$ is a **consistent estimator** of V^*, *regardless of the* (*unknown*) *form of the heteroskedasticity*.
- This includes no heteroskedasticity (*i.e.*, homoscedastic errors).
- Newey & West produced a corresponding consistent estimator of V^* for when the errors possibly exhibit autocorrelation (of some unknown form).
- Note that the White and the Newey-West estimators modify just the estimated covariance matrix of ***b*** – not ***b*** itself.
- As a result, the *t*-statistics, *F*-statistic, *etc*., will be modified, but only in a manner that is appropriate *asymptotically*.
- So, if we have heteroskedasticity (or autocorrelation), whether we modify the covariance estimator or not, the usual test statistics will be unreliable **in finite samples**.
- A good practical solution is to use White's (or Newey-West's) adjustment, and then use the Wald test, rather than the *F*-test for exact linear restrictions.
- This Wald test will incorporate the consistent estimator of the covariance matrix of ***b***, and so it will still be valid, *asymptotically*.

- Now let's turn to the estimation of $\boldsymbol{\beta}$, taking account of the fact that the error term has a non-scalar covariance matrix.
- Using this information should enable us to improve the *efficiency* of the LS estimator of the coefficient vector.

Generalized Least Squares (Alexander Aitken, 1935)

- In the present context, (Ordinary) LS ignores some important information, and we'd anticipate that this will result in a loss of efficiency when estimating $\boldsymbol{\beta}$.
- Let's see how to obtain the fully efficient (linear unbiased) estimator.
- Recall that $V(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \Sigma = \sigma^2\Omega$.
- Generally, Ω will be *unknown*. However, to begin with, let's consider the case where it is actually *known*.
- Clearly, Ω must be *symmetric*, as it is a covariance matrix.
- Suppose that Ω is also *positive-definite*.
- Then, Ω^{-1} is also positive-definite, and so there exists a *non-singular* matrix, P, such that $\Omega^{-1} = P'P$.
- In fact, $P' = C\Lambda^{-1/2}$, where the columns of C are the characteristic vectors of Ω, and $\Lambda^{1/2} = diag.(\sqrt{\lambda_i})$. Here, the $\{\lambda_i\}$ are the characteristic roots of Ω.
- Our model is:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad ; \quad \boldsymbol{\varepsilon} \sim [0\,, \sigma^2\Omega]$$

- Pre-multiply the equation by P:

$$P\boldsymbol{y} = PX\boldsymbol{\beta} + P\boldsymbol{\varepsilon}$$

  or,

$$\boldsymbol{y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \quad ; \quad \text{say}$$

- Now, Ω is non-random, so P is also non-random.
- So, $E[\boldsymbol{\varepsilon}^*] = E[P\boldsymbol{\varepsilon}] = P\,E[\boldsymbol{\varepsilon}] = \mathbf{0}$.

- And

$$\begin{aligned} V[\boldsymbol{\varepsilon}^*] &= V[P\boldsymbol{\varepsilon}] \\ &= PV(\boldsymbol{\varepsilon})P' \\ &= P(\sigma^2\Omega)P' = \sigma^2 P\Omega P' \end{aligned}$$

- Note that $P\Omega P' = P(\Omega^{-1})^{-1}P'$

$$= P(P'P)^{-1}P'$$

$$= PP^{-1}(P')^{-1}P' = I$$

- (Because P is both square and non-singular.)
- So, $E[\boldsymbol{\varepsilon}^*] = \mathbf{0}$ and $V[\varepsilon^*] = \sigma^2 I$.
- The transformed model, $\boldsymbol{y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, has an error-term that satisfies the *usual assumptions*. **In particular, it has a scalar covariance matrix**.

- So, if we apply (Ordinary) Least Squares to the model, $\boldsymbol{y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, we'll get the BLU estimator of $\boldsymbol{\beta}$, by the Gauss-Markhov Theorem.

- We call this the **Generalized Least Squares Estimator** of $\boldsymbol{\beta}$.

- The formula for this estimator is readily determined:

$$\hat{\boldsymbol{\beta}} = [X^{*\prime}X^*]^{-1}X^{*\prime}\boldsymbol{y}^*$$
$$= [(PX)'(PX)]^{-1}(PX)'(P\boldsymbol{y})$$
$$= [X'P'PX]^{-1}X'P'P\boldsymbol{y}$$

$$= [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\boldsymbol{y}$$

- Note that we can also write the GLS estimator as:

$$\hat{\boldsymbol{\beta}} = [X'(\sigma^2\Omega)^{-1}X]^{-1}X'(\sigma^2\Omega)^{-1}\boldsymbol{y}$$

$$= [X'\Sigma^{-1}X]^{-1}X'\Sigma^{-1}\boldsymbol{y} = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\boldsymbol{y}$$

- Clearly, because $E[\boldsymbol{\varepsilon}^*] = \mathbf{0}$ as long as the regressors are non-random, the GLS estimator, $\hat{\boldsymbol{\beta}}$ is **unbiased**.
- Moreover, the covariance matrix of the GLS estimator is:

$$V(\hat{\boldsymbol{\beta}}) = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}V(\boldsymbol{y})\{[X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\}'$$

$$= [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\sigma^2\Omega\Omega^{-1}X[X'\Omega^{-1}X]^{-1}$$

$$= \sigma^2[X'\Omega^{-1}X]^{-1} .$$

- If the errors are Normally distributed, then the full sampling distribution of the GLS estimator of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta}\,, \sigma^2[X'\Omega^{-1}X]^{-1},]$$

- The GLS estimator is just the OLS estimator, applied to the transformed model, and the latter model satisfies all of the usual conditions.
- So, the *Gauss-Markhov Theorem* applies to the GLS estimator.
- The GLS estimator is BLU for this more general model (with a non-scalar error covariance matrix).
- Note: OLS must be *inefficient* in the present context.
- Have a more general form of the GMT – the OLS version is a special case.
- Moreover, all of the results that we established with regard to testing for linear restrictions and incorporating them into our estimation, also apply if we make some obvious modifications.
- $\hat{\boldsymbol{\beta}} =$ GLS estimator

  $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y}^* - X^*\hat{\boldsymbol{\beta}}$

  $\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/(n-k)$
- Then, to test $H_0: R\boldsymbol{\beta} = \boldsymbol{q}$ *vs.* $H_A: R\boldsymbol{\beta} \neq \boldsymbol{q}$ we would use the test statistic,

  $$F = \left(R\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)'[R(X^{*\prime}X^*)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)/J\hat{\sigma}^2$$
- If H_0 is true, then is distributed as $F_{J,n-k}$.
- We can also construct the **Restricted GLS estimator**, in the same way that we obtained the restricted OLS estimator of $\boldsymbol{\beta}$.

- Check for yourself that this restricted estimator is

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_r &= \hat{\boldsymbol{\beta}} - (X^{*\prime}X^*)^{-1}R'[R(X^{*\prime}X^*)^{-1}R']^{-1}(R\hat{\boldsymbol{\beta}} - \boldsymbol{q}) \\
&= \hat{\boldsymbol{\beta}} - (X'\Omega^{-1}X)^{-1}R'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\hat{\boldsymbol{\beta}} - \boldsymbol{q}) \\
&= \hat{\boldsymbol{\beta}} - (X'\Sigma^{-1}X)^{-1}R'[R(X'\Sigma^{-1}X)^{-1}R']^{-1}(R\hat{\boldsymbol{\beta}} - \boldsymbol{q})
\end{aligned}
$$

- Then, if the residuals from this restricted GLS estimation are defined as $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{r}} = \boldsymbol{y} - X\hat{\boldsymbol{\beta}}_r$, we can also write the F-test statistic as:

$$F = [\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}'\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{r}} - \hat{\boldsymbol{\varepsilon}}'\,\hat{\boldsymbol{\varepsilon}}]\,/(J\hat{\boldsymbol{\varepsilon}}'\,\hat{\boldsymbol{\varepsilon}}/(n-k))$$

- Recalling our formula for the GLS estimator, we see that it depends on the (usually unknown) covariance matrix of the error term:

$$\hat{\boldsymbol{\beta}} = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\boldsymbol{y}\ .$$

“Feasible” GLS Estimator

- In order to be able to implement the GLS estimator, in practice, we’re usually going to have to provide a *suitable estimator* of Ω (or Σ).
- Presumably we’ll want to obtain an estimator that is *at least consistent*, and place this into the formula for the GLS estimator, yielding:

$$\tilde{\boldsymbol{\beta}} = \left[X'\hat{\Omega}^{-1}X\right]^{-1}X'\hat{\Omega}^{-1}\boldsymbol{y}$$

- Problem: The Ω matrix is $(n \times n)$, and it has $n(n+1)/2$ *distinct* elements. However, we have only n observations on the data. This precludes obtaining a consistent estimator.
- We need to constrain the elements of Ω in some way.
- In practice, this won’t be a big problem, because usually there will be lots of “structure” on the form of Ω .
- Typically, we’ll have $\Omega = \Omega(\boldsymbol{\theta})$, where the vector, $\boldsymbol{\theta}$ has low dimension.

Example: **Heteroskedasticity**

Suppose that $var.(\varepsilon_i) \propto (\theta_1 + \theta_2 z_i) = \sigma^2(\theta_1 + \theta_2 z_i)$

Then,

$$\Omega = \begin{pmatrix} \theta_1 + \theta_2 z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_1 + \theta_2 z_n \end{pmatrix}$$

There are just two parameters that have to be estimated, in order to obtain $\widehat{\Omega}$.

Example: **Autocorrelation**

Suppose that the errors follow a *first-order autoregressive process*:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad ; \quad u_t \sim N[0\,,\sigma_u^2] \qquad \text{(i.i.d.)}$$

Then (for reasons we'll see later),

$$\Omega = \frac{\sigma_u^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^{n-2} \\ \vdots & \rho & \ddots & \vdots \\ \rho^{n-1} & \dots & \dots & 1 \end{bmatrix} = \Omega(\rho).$$

- So, typically, we'll just have to estimate a very small number of parameters in order to get an estimator of Ω .
- As long as we use a *consistent estimator* for these parameters – the elements of θ, this will give us a consistent estimator of Ω and of Ω^{-1}, by Slutsky's Theorem.
- This in turn, will ensure that our Feasible GLS estimator of β is also *weakly consistent*:

$$plim(\tilde{\beta}) = plim\left\{\left[X'\widehat{\Omega}^{-1}X\right]^{-1}X'\widehat{\Omega}^{-1}\boldsymbol{y}\ \right\}$$

$$= plim\{[X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\boldsymbol{y}\ \}$$

$$= plim(\hat{\beta}) = \beta\ .$$

- Also, if $\widehat{\Omega}$ is consistent for Ω then $\tilde{\beta}$ will be *asymptotically efficient*.

- In general, we can say little about the *finite-sample* properties of our feasible GLS estimator.
- Usually it will be *biased*, and the nature of the bias will depend on the form of Ω and our choice of $\hat{\Omega}$.
- In order to apply either the GLS estimator, or the feasible GLS estimator, we need to know the form of Ω .
- Typically, this is achieved by postulating various forms, and testing to see if these are supported by the data.

Appendix – R-Code for perspective plots and contours

(see http://quantcorner.wordpress.com/2012/09/21/bivariate-normal-distribution-with-r/)

```
# Édouard Tallent @ TaGoMa.Tech
# September 2012
# This code plots simulated bivariate normal distributions

# Some variable definitions
mu1 <- 0   # expected value of x
mu2 <- 0   # expected value of y
sig1 <- 0.5     # variance of x
sig2 <- 1  # variance of y
rho <- 0.5 # corr(x, y)

# Some additional variables for x-axis and y-axis
xm <- -3
xp <- 3
ym <- -3
yp <- 3

x <- seq(xm, xp, length= as.integer((xp + abs(xm)) * 10))   # vector
series x
y <- seq(ym, yp, length= as.integer((yp + abs(ym)) * 10))   # vector
series y

# Core function
bivariate <- function(x,y){
	term1 <- 1 / (2 * pi * sig1 * sig2 * sqrt(1 - rho^2))
	term2 <- (x - mu1)^2 / sig1^2
	term3 <- -(2 * rho * (x - mu1)*(y - mu2))/(sig1 * sig2)
	term4 <- (y - mu2)^2 / sig2^2
	z <- term2 + term3 + term4
	term5 <- term1 * exp((-z / (2 *(1 - rho^2))))
	return (term5)
}
```

```
# Computes the density values
z <- outer(x,y,bivariate)

# Plot
persp(x, y, z, main = "Bivariate Normal Distribution",
           sub = bquote(bold(mu[1])==.(mu1)~",  "~sigma[1]==.(sig1)~",
           "~mu[2]==.(mu2)~", "~sigma[2]==.(sig2)~", "~rho==.(rho)),
           col="lightblue", theta = 55, phi = 30, r = 40, d = 0.1, expand
           = 0.5,ltheta = 90, lphi = 180, shade = 0.4, ticktype =
           "detailed", nticks=5)

#In order to see the contours, use:
#contour(x,y,z)
```