Lecture 7-b
Departures from OLS Assumptions

Brooks (4th edition): Chapter 5

 1

Review - CLM: Departures from (A3)

- The CLM assumes

$$(\mathbf{A3}) \quad \text{Var}[\boldsymbol{\varepsilon}\,|\,\boldsymbol{X}] = \sigma^2 \mathbf{I}_T = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

- Now, we assume:

$$(\mathbf{A3'}) \quad \text{Var}[\boldsymbol{\varepsilon}\,|\,\boldsymbol{X}] = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

- Two Special Cases:
  – **Pure heteroscedasticity**: We model only the diagonal elements.
  – **Pure autocorrelation**: We model only the off-diagonal elements. 11

Review – CLM: Heteroscedasticity

• **Pure heteroscedasticity:**

$$E[\varepsilon_i \, \varepsilon_j \,|\, \boldsymbol{X}] = \sigma_{ij} = \sigma_i^2 \quad \text{if } i = j$$
$$= 0 \quad \text{if } i \neq j$$
$$\Rightarrow \text{Var}[\varepsilon_i \,|\, \boldsymbol{X}] = \sigma_i^2$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix}$$

• Common structure in:

- Time series: The variance of the errors changing over time or subject to different regimes (say, bear and bull regimes).

- Cross sections: Firms in different industries have different variances.

3

Review – CLM: Cross/auto-correlation

• **Pure cross/auto-correlation:**

$$E[\varepsilon_i \, \varepsilon_j \,|\, \boldsymbol{X}] = \sigma_{ij} \quad \text{if } i \neq j$$
$$= \sigma^2 \quad \text{if } i = j$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma^2 \end{bmatrix}$$

• Common structure in:

- Cross sections: Errors of two firms in the same industry can be correlated, since they are subject to common (industry) shocks.

- Time series: Returns show clustering of errors ("news") over time, since it takes time to absorb shocks.

4

Review – CLM: (A3') Implications

- OLS **b** is still **unbiased** and **consistent**. (Proofs do not rely on (**A3**).)

- OLS **b** still follows an **asymptotic normal distribution**.

- But, OLS **b** is **no longer** BLUE. There are more efficient estimators; estimators that take into account the heteroscedasticity in the data.

Note: We used (**A3**) to derive our test statistics. A revision is needed!

5

Review - Testing for Heteroscedasticity

- We test for heteroscedasticity for efficiency & proper inference.

- We want to test: H_0: E(ε_i^2) = σ^2 for all i.
  H_1: E(ε_i^2) = $\sigma_i^2 \neq \sigma^2$ for at least some i.

- The structure of H_1 drives the form (& power) of the test. It depend on what we consider the drivers of σ_i^2: a particular variable, say $\boldsymbol{x}_j$, a regime (before & after some event), or past volatility, σ_{t-j}^2.

- We went over three tests of heteroscedasticity:
  - **Goldfeld & Quandt** (**GQ**) -in general, H_1 involves regimes
  - **Breusch & Pagan** (**BP**) -we have a particular H_1 in mind
  - **White** -general departure of H_0

21

Review - Heteroscedasticity Test: GQ Test

- GQ tests H_0: $E(\varepsilon_i^2) = \sigma^2$
  H_1: $\sigma_i^2 = h(\boldsymbol{x}_j)$ $\boldsymbol{x}_j$: variable/regime dummy.

- Steps for the **GQ test**:

– **Step 1**. Arrange the data from small to large values of the independent variable suspected of causing heteroscedasticity, $\boldsymbol{x}_j$.

– **Step 2**. Run two separate regressions, one for small values of $\boldsymbol{x}_j$ and one for large values of $\boldsymbol{x}_j$, omitting ***d*** middle observations (***d*** ≈ 20%). Get the RSS for each regression: RSS_1 for small values of $\boldsymbol{x}_j$ and RSS_2 for large $\boldsymbol{x}_j$'s.

– **Step 3**. Calculate the F ratio

$$GQ = \frac{RSS_2}{RSS_1}, \sim F_{df,df}, \quad \text{with } df = \frac{(T-d) - 2*(k+1)}{2} \qquad (\textbf{A5} \text{ holds})$$

22

Review - Heteroscedasticity Test: BP Test

- The derivation of the BP test is complicated. The implementation of the **studentized BP** test is simple, based on the squared OLS residuals, e_i^2, & the specific set of drivers of σ_i^2, the z_i's, under H_1.

- Steps for the **studentized Breusch-Pagan LM test**

- **Step 1**. Run OLS on DGP:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}. . \qquad \text{–Keep } e_i$$

- **Step 2**. (Auxiliary Regression). Run the regression of e_i^2/σ_R^2 on the m explanatory variables, $\boldsymbol{z}$. In our example,

$$e_i^2 = \alpha_0 + \alpha_1 z_{1,i} + ... + \alpha_m z_{m,i} + v_i \qquad \text{–Keep } R^2 \ (R_{e2}^2)$$

- **Step 3**. Calculate

$$LM = T\ R_{e2}^2 \xrightarrow{d} \chi_m^2.$$

25

Review - Heteroscedasticity Test: White Test

- The White test derivation is complicated, but, easy to compute.

- Steps for the **White LM test**:

– **Step 1**. (Same as BP's Step 1). Run OLS on DGP:

$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$. Keep residuals, e_i.

– **Step 2**. (Auxiliary Regression). Regress e_i^2 on all the explanatory variables (x_j), their squares (x_j^2), & all their cross products $(x_j * x_i)$.

For example, with $k = 2$ explanatory variables, the test is based on:

$$e_i^2 = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i}^2 + \beta_4 x_{2,i}^2 + \beta_5 x_{1,i} x_{2,i} + v_i$$

Let m be the number of regressors in auxiliary regression (in the above example, $m = 5$). Keep R^2, say R_{e2}^2.

– **Step 3**. Compute the statistic: LM = $T\, R_{e2}^2 \xrightarrow{d} \chi_m^2$.

9

Finding Auto/cross-correlation

- We test for autocorrelation for efficiency & proper inference. Usually, we consider an AR model for the errors, ε_t. For example, AR(p):

$$\varepsilon_t = \rho_1\, \varepsilon_{t-1} + \rho_2\, \varepsilon_{t-2} + \cdots + \rho_p\, \varepsilon_{t-p} + u_t \qquad - u_t \sim D(0, \sigma^2)$$

- Breusch & Godfrey (1978) use this AR(p) structure as the base of H_1 & the structure of the LM test, which is joint test:

H_0 (No autocorrelation): $\rho_1 = \ldots = \rho_p = 0$.

H_1: At least one $\rho_i \neq 0$. $\qquad i = 1, 2, \ldots, p$

Under H_0, Breusch & Godfrey use OLS residuals, e_i, to construct an LM test (**BG test**), similar to the BP test.

10

Review: (BG) LM Test for Autocorrelation

- Steps for the **Breusch–Godfrey (1978) LM test**:
  - **Step 1**. (Same as BP's Step 1). Run OLS on DGP:

    $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$. - Keep residuals, e_t.

  - **Step 2**. (Auxiliary Regression). Run the regression of e_t on all the explanatory variables, $\boldsymbol{X}$, and p lags of residuals, e_t:

    $e_t = \boldsymbol{x}_t{}'\gamma + \alpha_1 e_{t-1} + ... + \alpha_p e_{t-p} + v_t$ - Keep R^2 (R_e^2)

  - **Step 3**. Keep R_e^2. Then, calculate:

    $\text{LM} = (T - p) * R_e^2 \xrightarrow{d} \chi_p^2$.

Note: In general, in **Step 2**, if we do not include $\boldsymbol{x}_t$, the LM test is not that different.

11

Testing for Autocorrelation: LM Test

Example: LM-AR Test for the 3 factor F-F model for **IBM returns** (p = **12 lags**):

```
e_ibm <- fit_ibm_ff3$residuals              # OLS residuals
p_lag <- 12                                 # Select # of lags for test (set p)
e_lag <- matrix(0,T-p_lag,p_lag)            # Matrix to collect lagged residuals
a <- 1
while (a<=p_lag) {                          # loop creates matrix (e_lag) with lagged e
  za <- e_ibm[a:(T-p_lag+a-1)]
   e_lag[,a] <- za
a <- a+1
}

Mkt_RF_p <- Mkt_RF[(p_lag+1):T]             # Adjust for new sample size: T – p_lag
SMB_p <- SMB[(p_lag+1):T]
HML_p <- HML[(p_lag+1):T]
fit_ibm_ar <- lm(e_ibm[(p_lag+1):T] ~ e_lag + Mkt_RF_p + SMB_p + HML_p) # Aux R
r2_e1 <- summary(fit_ibm_ar)$r.squared      # get R² from Auxiliary Regression
```

12

Testing for Autocorrelation: LM Test

Example (continuation):
The package *lmtest,* performs this test, *bgtest,* (and many others, used in this class, encompassing, jtest, waldtest, etc).

```
library(lmtest)
> bgtest(ibm_x ~ Mkt_RF + SMB + HML, order=12)

	Breusch-Godfrey test for serial correlation of order up to	12

data:  lr_ibm ~ Mkt_RF + SMB + HML
LM test = 13.206, df = 12, p-value = 0.3543
```

(minor difference with the previous test, due to starting values of lags (here, all set to 0). Results do not change much.)

Note: If you do not include in the Auxiliary Regression the original regressors (Mkt_RF, SMB, HML) the test do not change much. You get LM-AR(**12**) Test: **13.731** $\Rightarrow$ very similar. Not entirely correct, but it works well.

13

Testing for Autocorrelation: LM Test

Example (continuation):
Autocorrelation is very common. If I run the test for Disney, or CAT, instead, we get significant test results.

- For **DIS**:

```
lr_dis <- log(x_dis[-1]/x_dis[-T])
dis_x <- lr_dis – RF

> bgtest(fit_dis_ff3, order=4)
	Breusch-Godfrey test for serial correlation of order up to 4
data:  fit_dis_ff3
LM test = 9.2059, df = 4, p-value = 0.05615
```

$\Rightarrow$ cannot reject H_0 at 5% level (*p-value* >.05)

```
> bgtest(dis_x ~ Mkt_RF + SMB + HML, order=12)
	Breusch-Godfrey test for serial correlation of order up to  12

data:  dis_x ~ Mkt_RF + SMB + HML
LM test = 28.706, df = 12, p-value = 0.004356
```

$\Rightarrow$ reject H_0 at 5% level (*p-value* < .05)

14

Testing for Autocorrelation: LM Test

Example (continuation):
LM tests for autocorrelation (with 4 or 12 lags) for **CAT** :

```
lr_cat <- log(x_ge[-1]/x_ge[-T])
cat_x <- lr_cat – RF
```

• For **CAT**:
> bgtest(fit_cat_ff3, order=**4**)
Breusch-Godfrey test for serial correlation of order up to **4**

data: fit_cat_ff3LM test = **6.253**, df = **4**, p-value = **0.181** ⇒ cannot reject H_0 at 5% level (*p-value* >.05)

> bgtest(fit_cat_ff3, order=**12**)
Breusch-Godfrey test for serial correlation of order up to **12**

data: fit_cat_ff3
LM test = **.20.259**, df = **12**, p-value = **0.0623** ⇒ cannot reject H_0 at 5% level (*p-value* < .05)

Note: In both examples, adding more lags decreases *p*-values.

15

Testing for Autocorrelation: LM Test

• Q: How many lags are needed in the test?
A: Enough to make sure there is no auto-correlation left in the residuals.

There are some popular rule of thumbs:
- Daily data, 5 or 20 lags
- Weekly, 4 or 12 lags
- Monthly data, 12 lags
- Quarterly data, 4 lags

16

Testing for Autocorrelation: Durbin-Watson

• The Durbin-Watson (1950) (DW) test for AR(1) autocorrelation: H_0: $\rho_1 = 0$ against H_1: $\rho_1 \neq 0$. Based on simple correlations of $\boldsymbol{e}$.

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

• It is easy to show that when $T \rightarrow \infty$, $d \approx 2(1 - \rho_1)$.

• ρ_1 is estimated by the sample correlation r.

• Under H_0, ρ_1=0. Then, d should be distributed randomly around 2.

• Small values (close to 0) or Big values (close to 4) of d lead to rejection of H_0. The distribution depends on $\boldsymbol{X}$. Since there are better tests, in practice, the DW is used "**visually**:" Is d close to 2?

The R function *dwtest* from the *lmtest* package produces also a *p-value*.

17

Testing for Autocorrelation: DW Test

Example: DW Test for the 3 factor F-F model for IBM returns

```
RSS <- sum(e_ibm^2)                                    # RSS
DW <- sum((e_ibm[1:(T-1)] - e_ibm[2:T])^2)/RSS         # DW stat
> DW
[1] 2.048635        ⇒ DW statistic ≈ 2  ⇒ No evidence for autocorrelation of order 1.
> 2 * (1 - cor(e_ibm[1:(T-1)], e_ibm[2:T]))            # approximate DW stat
[1] 2.049084
```

• Similar finding for Disney returns:

```
> DW
       [,1]
[1,] 2.1327          ⇒ DW statistic ≈ 2  ⇒  But, DIS suffers from autocorrelation!
```

⇒ This is why DW are not that informative. They only test for AR(1) in residuals.

Note: The package *lmtest* performs this test too, *dwtest*:

```
> dwtest(fit_ibm_ff3)
DW = 2.0486, p-value = 0.7266
```

18

Testing for Autocorrelation: DW Test

Example: DW Test for the residuals of the encompassing model (IFE + PPP) for changes in **USD/GBP**:

```
e_gbp <- fit_gbp$residuals
> dwtest(fit_gbp)

        Durbin-Watson test

data:  fit_gbp
DW = 1.8588, p-value = 0.08037                    ⇒ not significant at 5% level.
alternative hypothesis: true autocorrelation is greater than 0
```

19

Testing for Autocorrelation: Portmanteu tests

- Portmanteu tests are tests with a well-defined H_0, but not specific H_1. We will present two: Box-Pierce Q test and the Ljung-Box test.

- **Box-Pierce (1970) test (Q test).**

It tests H_0 (No autocorrelation): $\rho_1 = ... = \rho_p = 0$, using the sample correlation, $r_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}$, where (using time series notation)

$\hat{\gamma}_j$ = Sample covariance between y_t & y_{t-j} = $\frac{\sum_{t=j+1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y})}{T-j}$

$\hat{\gamma}_0$ = Sample variance = $\frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T-1}$

Then, under H_0:

$$Q = T * \sum_{j=1}^{p} r_j^2 \xrightarrow{d} \chi_p^2.$$

20

Testing for Autocorrelation: Portmanteau tests

• **Ljung-Box (1978) test (LB test).**

A variation of the Box-Pierce test. It has a small sample correction.

$$LB = T * (T+2) * \sum_{j=1}^{p} \frac{r_j^2}{T-j} \xrightarrow{d} \chi_p^2.$$

Technical Note: The asymptotic distribution of both tests is based on the fact that, under the null of independent data, $\sqrt{T}\,\boldsymbol{r} \xrightarrow{d}$ N(0, **I**).

Note: When analyzing residuals, e_t, of a regression we compute r_j as:

$$r_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0} = \frac{\sum_{t=j+1}^{T} e_t\, e_{t-j}}{\sum_{t=1}^{T} e_t^2}$$

• The LB statistic is widely used. But, the BG (1978) LM tests conditions on $\boldsymbol{X}$. Thus, it is more powerful..

21

Testing for Autocorrelation: Portmanteau tests

Example: Q and LB tests with p =12 lags for the residuals in the 3-factor FF model for **IBM excess returns**, using the *Box.test* function:

• **Q test**

```
> Box.test(e_ibm, lag = 12, type="Box-Pierce")

        Box-Pierce test

data:  e_ibm
X-squared = 13.017, df = 12, p-value = 03678
```

• **LB test**

```
> Box.test(e_ibm, lag = 12, type="Ljung-Box")

        Box-Ljung test

data:  e_ibmX-squared = 13.24, df = 12, p-value = 0.3519
```

22

Testing for Autocorrelation: Portmanteu tests

Example (continuation): Same tests (p=12 lags) & same model:

• For **DIS** (e_dis), we get:

• **Q**

[1] **25.22863** (*p-value* = **0.01378**) ⇒ reject H_0 at 5% level.

• **LB**

[1] **25.539** (*p-value* = **0.01246**) ⇒ reject H_0 at 5% level.

• For **CAT** (e_cat), we get:

• **Q**

[1] **23.071** (*p-value* = **0.02713**) ⇒ reject H_0 at 5% level.

• **LB**

[1] **23.409** (*p-value* = **0.0244**) ⇒ reject H_0 at 5% level.

• Autocorrelation in financial asset returns is a usual finding in monthly, weekly and daily data.

23

Testing for Autocorrelation: Portmanteu tests

Example: Same Q and LB tests (**p = 12 lags**) for the **USD/GBP** residuals in the encompassing (PPP + IFE) model:

• **Q**

```
e_gbp <- fit_gbp$residuals
> Box.test(e_gbp, lag = 12, type="Box-Pierce")

        Box-Pierce test

data:  e_gbp
X-squared = 19.587, df = 12, p-value = 0.0753
```

⇒ cannot reject H_0 at 5% level, but close.

• **LB**

```
> Box.test(e_gbp, lag = 12, type="Ljung-Box")

        Box-Ljung test

data:  e_gbp
X-squared = 20.032, df = 12, p-value = 0.06649
```

⇒ cannot reject H_0 at 5% level.

24

Testing for Autocorrelation: Portmanteau tests

• Q & LB tests are widely use, but they have two main limitations:

(1) The test was developed under the independence assumption.

If y_t shows dependence, such as heteroscedasticity, the asymptotic variance of $\sqrt{T}\,\boldsymbol{r}$ is no longer **I**, but a non-diagonal matrix.

There are several proposals to "**robustify**" both Q & LB tests. The "robustified" Portmanteau statistic uses $\tilde{r}_j$ instead of r_j ($\tilde{r}_j$ has an extra term in the denominator):

$$\tilde{r}_j = \frac{\hat{\gamma}_j^2}{\tau_j} = \frac{\sum_{t=j+1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y})}{\sum_{t=j+1}^{T}(y_t - \bar{y})^2\ (y_{t-j} - \bar{y})^2}$$

Thus, for Q we have:

$$\text{Q}^* = T\ \sum_{j=1}^{p} \tilde{r}_j^2 \xrightarrow{d} \chi_p^2.$$

25

Testing for Autocorrelation: Portmanteau tests

(2) The selection of the number of autocorrelations p is arbitrary.

The traditional approach is to try different p values, say 3, 6 & 12. Another popular approach is to let the data "select" p, for example, using AIC or BIC, an approach sometimes referred as "**automatic selection**."

Escanciano and Lobato (2009) propose combining BIC's and AIC's penalties to select p in Q* (BIC for small ρ and AIC for bigger ρ).

• It is possible to reach very different conclusion from Q and Q*.

26

Testing for Autocorrelation: Portmanteu tests

Example: Q* tests with automatic selection of **p** for the residuals in the 3-factor FF model for **IBM & DIS excess returns.** We use Auto.Q funcition in R package *vrtest*.

```
- For IBM (e_ibm), we get:
> library(vrtest)
> Auto.Q(e_ibm, 12)            #Maximum potential lag = 12
> $Stat
[1] 0.2781782

$Pvalue
[1] 0.5978978

- For DIS (e_dis), we get:
> Auto.Q(e_dis, 12)
$Stat
[1] 2.649553

$Pvalue
[1] 0.103579                   ⇒ Reversal for DIS
```

27

Testing for Autocorrelation: Heteroscedasticity

• Time-varying volatility is very common in financial time series. We can use the Q & LB tests for autocorrelation to check for autocorrelation in squared errors, e_i^2, which based on White's idea, we use to estimate σ_i^2.

• We use a Portmanteu test on the squared residuals to check for a particular kind of heteroscedasticity: the variance, σ_i^2, is driven by lagged squared errors.

$$H_0: \sigma_i^2 = \sigma^2$$

$$H_1: \sigma_i^2 = f(\varepsilon_{i-1}^2, \varepsilon_{i-2}^2, ..., \varepsilon_{i-p}^2)$$

• Of course, an LM-BP test can also be used, using lagged squared residuals as the drivers of heteroscedasticity (more on this topic in Lecture 10).

28

Testing for Autocorrelation: Heteroscedasticity

Example: Q and LB tests with ***p*** = **12 lags** for the squared residuals in the 3-factor FF model for **IBM returns**:

```
> e_ibm2 <- e_ibm^2
```

• Q test

```
> Box.test(e_ibm2, lag = 12, type="Box-Pierce")

        Box-Pierce test

data:  e_ibm2
X-squared = 37.741, df = 12, p-value = 0.0001693
```

• LB test

```
> Box.test(e_ibm2, lag = 12, type="Ljung-Box")

        Box-Ljung test

data:  e_ibm2
X-squared = 38.435, df = 12, p-value = 0.0001304
```

29

Testing for Autocorrelation: Heteroscedasticity

Example (continuation): Q and LB tests with ***p*** = **12 lags** for the squared residuals in the 3-factor FF model for **DIS** & **GE** returns:

• For **DIS** (dis_x), we get

```
> Box.test(e_dis2, lag = 12, type="Ljung-Box")

        Box-Ljung test

data:  e_dis2
X-squared = 73.798, df = 12, p-value = 6.195e-11
```

• For **GE** (ge_x), we get

```
> Box.test(e_ge2, lag = 12, type="Ljung-Box")

        Box-Ljung test

data:  e_ge2
X-squared = 115.9, df = 12, p-value < 2.2e-16
```

• Strong evidence for time-varying heteroscedasticity in the residuals. 30

Generalized Regression Model (GRM)

- Now, we go back to the CLM Assumptions:

(**A1**) DGP: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) or (**A2'**)

(**A3'**) Var[$\boldsymbol{\varepsilon}$|$\boldsymbol{X}$] = $\boldsymbol{\Sigma}$ (sometimes written Var[$\boldsymbol{\varepsilon}$|$\boldsymbol{X}$] = $\sigma^2\boldsymbol{\Omega}$)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$ -a (TxT) symmetric matrix

(**A4**) or (**A4'**)

- This is the **generalized regression model** (**GRM**).

- OLS **b** is still unbiased (& consistent). Can we still use OLS?

GR Model: True Variance for b

- From (**A3**) Var[$\boldsymbol{\varepsilon}$|$\boldsymbol{X}$] = σ^2 $\mathbf{I}_T$ ⇒ Var[**b**|$\boldsymbol{X}$] = σ^2 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

- Now, we have (**A3'**) Var[$\boldsymbol{\varepsilon}$|$\boldsymbol{X}$] = $\boldsymbol{\Sigma}$

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

- The true variance of **b** under (**A3'**) should be:

$$\begin{aligned} \text{Var}_T[\mathbf{b}|\boldsymbol{X}] &= \text{E}[(\mathbf{b}-\beta)(\mathbf{b}-\beta)'|\boldsymbol{X}] \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\text{E}[\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}|\boldsymbol{X}]\,(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\boldsymbol{\Sigma}\,\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1} \end{aligned}$$

Example: We compute the true variance for the simplest case, a regression with only one explanatory variable and heteroscedastic $\boldsymbol{\varepsilon}$:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \qquad \varepsilon_i \sim \text{D}(0, \sigma_i^2)$$

$$\Rightarrow \quad \text{Var}_T[\mathbf{b}|\boldsymbol{X}] = \left(\frac{1}{\sum_i^T (x_i-\bar{x})^2}\right)^2 \sum_{i=1}^T \sigma_i^2 (x_i-\bar{x})^2.$$

GR Model: True Variance for b

Example (continuation):

$$\Rightarrow \qquad \text{Var}_T[\mathbf{b} \mid \mathbf{X}] = \left(\frac{1}{\sum_{i=1}^{T}(x_i-\bar{x})^2}\right)^2 \sum_{i=1}^{T} \sigma_i^2 (x_i - \bar{x})^2.$$

If we compute the OLS variance, we see how both estimators differ:

$$\text{Var}[\mathbf{b} \mid \boldsymbol{X}] = \frac{\sigma^2}{\sum_i^T (x_i-\bar{x})^2} \neq \text{Var}_T[\mathbf{b} \mid \boldsymbol{X}].$$

• Under **(A3')**, the OLS estimator of $\text{Var}_T[\mathbf{b} \mid \boldsymbol{X}]$. –i.e., $s^2\ (\boldsymbol{X}'\boldsymbol{X})^{-1}$– is **biased**. If we want to use OLS for inferences (say, with *t-test* or *F-test*), we need to estimate $\text{Var}_T[\mathbf{b} \mid \boldsymbol{X}]$.

• That is, we need to estimate the unknown $\boldsymbol{\Sigma}$. But, $\boldsymbol{\Sigma}$ has Tx$(T+1)/2$ parameters. Too many to estimate with only T observations!

GR Model: Robust Covariance Matrix

• We will not be estimating $\boldsymbol{\Sigma}$. Impossible with T data points.

• We will estimate $\boldsymbol{X}'\ \boldsymbol{\Sigma}\ \boldsymbol{X} = \sum_{i=1}^{T} \sum_{j=1}^{T} \sigma_{ij}\ \boldsymbol{x}_i\ \boldsymbol{x}_j{}'$, a $(k$x$k)$ matrix. That is, we are estimating $[k * (k+1)]/2$ elements.

• This distinction is very important in modern applied econometrics:
  – **The White estimator**
  – **The Newey-West estimator**

• Both estimators produce a **consistent** estimator of $\text{Var}_T[\mathbf{b} \mid \boldsymbol{X}]$:

$$\text{Var}_T[\mathbf{b} \mid \boldsymbol{X}] = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\ \boldsymbol{\Sigma}\ \boldsymbol{X}\ (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Since $\mathbf{b}$ consistently estimates $\boldsymbol{\beta}$, the OLS residuals, $\boldsymbol{e}$, are also consistent estimators of $\boldsymbol{\varepsilon}$. We use $\boldsymbol{e}$ to consistently estimate $\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}$.

Covariance Matrix: The White Estimator

• The White estimator simplifies the estimation since it only assumes heteroscedasticity. Then, $\boldsymbol{\Sigma}$ is a diagonal matrix, with elements σ_i^2.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix} \qquad \text{-a } (T\text{x}T) \text{ matrix}$$

Thus, we need to estimate: $\boldsymbol{Q}^* = (1/T)\ \boldsymbol{X}'\ \boldsymbol{\Sigma}\ \boldsymbol{X}$ -a $(k\text{x}k)$ matrix

where

$$\boldsymbol{X}'\ \boldsymbol{\Sigma}\ \boldsymbol{X} = \begin{bmatrix} \sum_{i=1}^{T} \boldsymbol{x}_{1i}^2\, \sigma_i^2 & \cdots & \sum_{i=1}^{T} \boldsymbol{x}_{1i}\boldsymbol{x}_{ki}\sigma_i^2 \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{T} \boldsymbol{x}_{ki}\boldsymbol{x}_{1i}\, \sigma_i^2 & \cdots & \sum_{i=1}^{T} \boldsymbol{x}_{ki}^2\sigma_i^2 \end{bmatrix} = \sum_{i=1}^{T} \sigma_i^2\, \boldsymbol{x}_i\, \boldsymbol{x}_i{}'$$

• Q: How do we estimate σ_i^2?

Covariance Matrix: The White Estimator

• We need to estimate: $\boldsymbol{Q}^* = (\frac{1}{T})\ \boldsymbol{X}'\ \boldsymbol{\Sigma}\ \boldsymbol{X} = (\frac{1}{T})\ \sum_{i=1}^{T} \sigma_i^2 \boldsymbol{x}_i \boldsymbol{x}_i{}'$

• The OLS residuals, $\boldsymbol{e}$, are consistent estimators of $\boldsymbol{\varepsilon}$. This suggests using e_i^2 to estimate σ_i^2. That is,

we estimate $\boldsymbol{Q}^* = (\frac{1}{T})\ \sum_{i=1}^{T} \sigma_i^2 \boldsymbol{x}_i \boldsymbol{x}_i{}'$

with $\boldsymbol{S}_0 = (\frac{1}{T})\ \sum_{i=1}^{T} e_i^2\ \boldsymbol{x}_i\ \boldsymbol{x}_i{}'$

Example: Back to the simplest case, a regression with one explanatory variable and heteroscedastic error term, we have:

$$\text{Var}_T[\mathbf{b}\,|\,\boldsymbol{X}] = \left(\frac{1}{\sum_i^T (x_i - \bar{x})^2}\right)^2 \sum_{i=1}^{T} \sigma_i^2 (x_i - \bar{x})^2$$

which we estimate using OLS residuals, e_i:

$$\text{Est Var}_T[\mathbf{b}\,|\,\boldsymbol{X}] = \left(\frac{1}{\sum_i^T (x_i - \bar{x})^2}\right)^2 \sum_{i=1}^{T} e_i^2 (x_i - \bar{x})^2.$$

Covariance Matrix: The White Estimator

• White (1980) shows that a consistent estimator of $\text{Var}_T[\mathbf{b}|\boldsymbol{X}]$ is obtained if e_i^2 is used as an estimator of σ_i^2. Taking the square root, we get a **heteroscedasticity-consistent** (**HC**) standard errors (**HCSE**).

• (**A3'**) was not specified. That is, the White estimator is **robust** to a potential misspecifications of heteroscedasticity in (**A3'**).

• The White estimator allows us to make inferences using the OLS estimator **b** in situations where heteroscedasticity is suspected, but we do not know enough to identify its nature.

Note: The estimator is also called the **sandwich estimator** or just the **White estimator**.

Halbert White (1950-2012, USA)

The White Estimator: Some Remarks

(1) Since there are many refinements of the White estimator, the White estimator is usually referred as HC0 (or just "HC"):

$$\text{HC0} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\,[\boldsymbol{X}'\,Diag[e_i^2]\,\boldsymbol{X}]\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

(2) In large samples, SEs, t-tests and F-tests are asymptotically valid.

(3) The OLS estimator remains inefficient. But inferences are asymptotically correct.

(4) The HC SEs can be larger or smaller than the OLS SEs (in general, HC SEs are larger when positively correlated to $\boldsymbol{x}_i$ or $\boldsymbol{x}_i{}^2$, which tends to be the case). It can make a difference to the tests.

(5) It is used, along the Newey-West estimator, in almost all finance applied work. Included in all the packaged software programs.

The White Estimator: Some Remarks

(6) In R, you can use the library "*sandwich,*" to calculate White SEs. They are easy to program:

```
# White SE in R
White_f <- function(y,X,b) {
T <- length(y)
 k <- length(b)
yhat <- X%*%b                                  # fitted values
e <- y-yhat                                    # residuals
hhat <- t(X)*as.vector(t(e))                   # x_i e_i
G <- matrix(0,k,k)                             # Create empty kxk matrix to place x'e ex
 za <- hhat[,1:k]%*%t(hhat[,1:k])              # X' diag[e_i] X
 G <- G + za                                   # X' diag[e_i] X
 F <- t(X)%*%X                                 # X'X
 V <- solve(F)%*%G%*%solve(F)                  # S_0
 white_se <- sqrt(diag(V))
 ols_se <- sqrt(diag(solve(F)*drop((t(e)%*%e))/(T-k)))
l_se = list(white_se,olse_se)
return(l_se) }
```

The White Estimator: Application 1 – IBM

Example: We estimate t-values using OLS and White SE, for the 3 factor F-F model for IBM returns:

$$(r_{i=IBM,t} - r_f) = \beta_0 + \beta_1 (r_{m,t} - r_f) + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_t$$

```
fit_ibm_ff3 <- lm(ibm_x ~ Mkt_RF + SMB + HML)      # OLS Regression with lm
b_ibm <- fit_ibm_ff3$coefficients                  # Extract OLS coeff's from fit_ibm_ff3
SE_OLS <- sqrt(diag(vcov(fit_ibm_ff3)))            # Extract OLS SE from fit_ibm_ff3
t_OLS <- b_ibm/SE_OLS                              # Calculate  OLS t-values

> b_ibm
 (Intercept)      Mkt_RF         SMB         HML
-0.005191356  0.910379487 -0.221385575 -0.139179020
> SE_OLS
(Intercept)      Mkt_RF         SMB         HML
0.002482305 0.056784474 0.084213761 0.084060299
> t_OLS
(Intercept)      Mkt_RF         SMB         HML
  -2.091345   16.032190   -2.628853   -1.655705
```

The White Estimator: Application 1 – IBM

Example (continuation):

```
> library(sandwich)
White <- vcovHC(fit_ibm_ff3, type = "HC0")
SE_White <- sqrt(diag(White))                                   # White SE HC0
t_White <- b_ibm/SE_White

> SE_White
(Intercept)      Mkt_RF        SMB        HML
0.002505978 0.062481080 0.105645459 0.096087035
> t_White
(Intercept)      Mkt_RF        SMB        HML
  -2.071589   14.570482   -2.095552   -1.448468          ⟹ HML not longer significant at 10% level

White3 <- vcovHC(fit_ibm, type = "HC3")                         # White SE HC3 (refinement)
SE_White3 <- sqrt(diag(White3))# White SE HC0
t_White <- b_i/SE_White3
> SE_White3
(Intercept)      Mkt_RF        SMB        HML
0.002533461 0.063818378 0.108316056 0.098800721
> t_White3
(Intercept)      Mkt_RF        SMB        HML
  -2.049116   14.265162   -2.043885   -1.408684          ⟹ similar results with HC3 refinement
```

The White Estimator: Application 2 – i_{MX}

Example: We estimate Mexican interest rates ($i_{MX,t}$) with a linear model including US interest rates, changes in exchange rates (MXN/USD), $e_{MX,t}$, Mexican inflation, $I_{MX,t}$, and Mexican GDP growth, $y_{MX,t}$, using quarterly data 1978:II – 2020:II (T=166):

$$i_{MX,t} = \beta_0 + \beta_1\, i_{US,t} + \beta_2\, e_{MX,t} + \beta_3\, I_{MX,t} + \beta_4\, y_{MX,t} + \varepsilon_t$$

```
FMX_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/FX_USA_MX.csv", head=TRUE,
sep=",")

us_i <- FMX_da$US_int                        # US short-term interest rates (i_US)
mx_CPI <- FMX_da$MX_CPI                      # Mexican CPI
mx_M1 <- FMX_da$MX_M1                        # Mexican Money Supply (M1)
mx_i <- FMX_da$MX_int                        # Mexican short-term interest rates (i_MX)
mx_GDP <- FMX_da$MX_GDP                      # Mexican GDP
S_mx <- FMX_da$MXN_USD                       # S_t = exchange rates (MXN/USD)
T <- length(mx_CPI)
mx_I <- log(mx_CPI[-1]/mx_CPI[-T])           # Mexican Inflation: Log  changes in CPI
mx_y <- log(mx_GDP[-1]/mx_GDP[-T])           # Mexican growth: Log  changes in GDP
```

The White Estimator: Application 2 – i_{MX}

Example (continuation):

```
mx_mg <- log(mx_M1[-1]/mx_M1[-T])           # Money growth: Log  changes in M1
e_mx <- log(S_mx[-1]/S_mx[-T])              # Log changes in S_t.
us_i_1 <- us_i[-1]/100                      # Adjust sample size.
mx_i_1 <- mx_i[-1]/100
mx_i_0 <- mx_i[-T]/100
fit_i <- lm(mx_i_1 ~ us_i_1 + e_mx + mx_I + mx_y)
b_i <- fit_i$coefficients                   # Extract OLS coeff's from fit_i
> summary(fit_i)

Coefficients:
            Estimate  Std. Error t value Pr(>|t|)
(Intercept) 0.04022   0.01506   2.671  0.00834 **
us_i_1      0.85886   0.31211   2.752  0.00661 **
e_mx        -0.01064  0.02130  -0.499  0.61812
mx_I        3.34581   0.19439  17.212  < 2e-16 ***
mx_y        -0.49851  0.73717  -0.676  0.49985
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The White Estimator: Application 2 – i_{MX}

Example (continuation):

```
White <- vcovHC(fit_i, type = "HC0")        # Extract White Var Matrix from fit_i
SE_White <- sqrt(diag(White))               # White SE HC0
t_White <- b_i/SE_White

> SE_White
(Intercept)     us_i_1      e_mx       mx_I       mx_y
0.009665759 0.480130221 0.026362820 0.523925226 1.217901733
> t_White
(Intercept)     us_i_1      e_mx       mx_I       mx_y
  4.1613603   1.7888018  -0.4035554  6.3860367  -0.4093221
```

⇒ $i_{US,t}$ not longer significant at 5% level.

```
White3 <- vcovHC(fit_i, type = "HC3")       # Using popular refinement HC3
SE_White3 <- sqrt(diag(White3))             # White SE HC3
t_White <- b_i/SE_White3
> t_White3
(Intercept)     us_i_1      e_mx       mx_I       mx_y
  3.6338983   1.5589936  -0.2117600  5.4554986  -0.3519886
```

⇒ $i_{US,t}$ not longer significant at 10% level

Newey-West Estimator

- Newey-West allow for both heteroscedasticy and autocorrelation.

(**A3'**) Var[$\boldsymbol{\varepsilon}$|$\boldsymbol{X}$] = $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$ -a (TxT) symmetric matrix

Now, we need to estimate

$$\boldsymbol{Q}^* = (\tfrac{1}{T})\, \boldsymbol{X}'\, \boldsymbol{\Sigma}\, \boldsymbol{X} = (1/T) \sum_{i=1}^{T} \sum_{j=1}^{T} \sigma_{ij}\, \boldsymbol{x}_i\, \boldsymbol{x}_j'$$

$$= (\tfrac{1}{T}) \sum_{i=1}^{T} \{\sigma_{i1}\, \boldsymbol{x}_i\, \boldsymbol{x}_1' + \sigma_{i2}\, \boldsymbol{x}_i\, \boldsymbol{x}_2' + \cdots + \sigma_{iT}\, \boldsymbol{x}_i\, \boldsymbol{x}_T'\}$$

- Newey and West (1987) follow White (1980) to produce a **HAC** (**Heteroscedasticity and Autocorrelation Consistent**) estimator of $\boldsymbol{Q}^*$, also referred as *long-run variance* (**LRV**): Use $e_i e_j$ to estimate σ_{ij}.

Newey-West Estimator

- Now, we also have autocorrelation. We need to estimate

$$\boldsymbol{Q}^* = (\tfrac{1}{T}) \sum_{i=1}^{T} \sum_{j=1}^{T} \sigma_{ij}\, \boldsymbol{x}_i\, \boldsymbol{x}_j'$$

⇒ natural estimator of $\boldsymbol{Q}^*$:

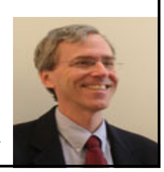
$$\boldsymbol{S}_T = (\tfrac{1}{T}) \sum_{i=1}^{T} \sum_{j=1}^{T} e_i\, e_j\, \boldsymbol{x}_i\, \boldsymbol{x}_j'$$

Or using time series notation, estimator of $\boldsymbol{Q}^*$:

$$\boldsymbol{S}_T = (\tfrac{1}{T}) \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{x}_t e_t\, e_s \boldsymbol{x}_s'$$

- There are some restrictions that need to be imposed:
  - **Q*** needs to be a pd matrix (use a quadratic form)
  - The double sum cannot explode (use decaying weights to cut the sum short, after lag L the weights are zero).

Whitney Newey, USA



Kenneth D. West, USA

Newey-West Estimator

- Using time series notation, estimator of **Q***:

$$S_T = (\frac{1}{T}) \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{x}_t e_t \, e_s \boldsymbol{x}_s'$$

Example: Back to the simplest case, a regression with only one explanatory variable, but now with a heteroscedastic and autocorrelated error term. We estimate the "true" variance of **b** with:

$$\text{Var}_T[\mathbf{b}\,|\,\boldsymbol{X}] = \left(\frac{1}{\sum_i^T (x_i-\bar{x})^2}\right)^2 \{\sum_{i=1}^{T} e_i^2 (x_i-\bar{x})^2 + + \sum_{i=1}^{T} \sum_{j=i+1}^{T} (x_i-\bar{x})\, e_i \; e_j (x_j-\bar{x})\}$$

We add the sum of the autocovariances of w_i (= $\boldsymbol{x}_i e_i$) to the White estimator of $\text{Var}_T[\mathbf{b}\,|\,\boldsymbol{X}]$. If the autocovariances of w_i (are positive, the NW estimator will be bigger than the White estimator. This is a very common case.

Newey-West Estimator

- Two components for the NW HAC estimator:

(1) Start with Heteroscedasticity Component:

$$\boldsymbol{S}_0 = \frac{1}{T} \sum_{t=1}^{T} e_t^2 \, \boldsymbol{x}_t \, \boldsymbol{x}_t' \qquad \text{– the White estimator.}$$

(2) Add the Autocorrelation Component

$$\boldsymbol{S}_T = \boldsymbol{S}_0 + \frac{1}{T} \sum_{l=1}^{L} k(l) \sum_{t=l+1}^{T} (\boldsymbol{x}_{t-l} e_{t-l} e_t \boldsymbol{x}_t' + \boldsymbol{x}_t e_t e_{t-l} \boldsymbol{x}_{t-l}')$$

where

$$k(\frac{j}{L(T)}) = \frac{L+1-|j|}{L+1} \qquad \text{–decaying weights (\textit{Bartlett kernel})}$$

L is the cut-off lag, which is a function of T. (More data, longer L).

The weights are linearly decaying, suppose L = 30. Then,

$k(1) = 30/31 = 0.9677419$

$k(2) = 29/31 = 0.9354839$

$k(3) = 28/31 = 0.9032258$

$$k(\frac{j}{L(T)}) = \frac{L+1-|j|}{L+1}$$

Newey-West Estimator

- $\boldsymbol{S}_T = \boldsymbol{S}_0 + \frac{1}{T} \sum_{l=1}^{L} k(l) \sum_{t=l+1}^{T} (\boldsymbol{x}_{t-l} e_{t-l} e_t \boldsymbol{x}_t' + \boldsymbol{x}_t e_t e_{t-l} \boldsymbol{x}_{t-l}')$

Then,

Est. Var[**b**] = $\frac{1}{T}$ $(\boldsymbol{X}'\boldsymbol{X}/T)^{-1}\boldsymbol{S}_T\, (\boldsymbol{X}'\boldsymbol{X}/T)^{-1}$ –NW's HAC Var.

- Under suitable conditions, as $L, T \rightarrow \infty$, and $L/T \rightarrow 0$, $\boldsymbol{S}_T \rightarrow \boldsymbol{Q}^*$.

- Asymptotic inferences can be based on OLS **b**, with *t-tests* and *Wald tests* using N(0,1) and χ^2 critical values, respectively.

- There are many refinements of the NW estimators. Today, all HAC estimators are usually referred as NW estimators, regardless of the weights (*kernel*) used if they produce a positive (semi-) definite covariance matrix.

Newey-West Estimator

Example: Back to the simplest case, a regression with only one explanatory variable, but with a heteroscedastic and autocorrelated error term. Suppose we set L =12, then:

$$\text{Var}_T[\mathbf{b}\,|\,\boldsymbol{X}] = \left(\frac{1}{\sum_i^T (x_i - \bar{x})^2}\right)^2 \{\sum_{t=1}^{T} e_t^2 (x_t - \bar{x})^2 + $$
$$+ \sum_{l=1}^{L=12} \left\{\frac{13 - |j|}{13}\right\} \sum_{t=i+1}^{T} (x_t - \bar{x})\, e_t\, e_{t-l} (x_{t-l} - \bar{x})\}$$

To compute $\boldsymbol{S}_T$, we only add 12 autocovariances of w_t (= $x_t\, e_t$) to the White estimator, $\boldsymbol{S}_0$.

<u>Technical Detail</u>: Above, it is mentioned that the asymptotics need that as L & $T \rightarrow \infty$, and $L/T \rightarrow 0$, to get $\boldsymbol{S}_T \rightarrow \boldsymbol{Q}^*$. That is, as we gather more data, we need to increase L –i.e., use more lags.

NW Estimator: In all Econometric Packages

- All econometric packages (SAS, SPSS, Eviews, etc.) calculate NW SE. In R, you can use the library "*sandwich*," to calculate NW SEs:

```
> library(sandwich)
> NeweyWest(x, lag = NULL, order.by = NULL, prewhite = TRUE, adjust = FALSE,
diagnostics = FALSE, sandwich = TRUE, ar.method = "ols", data = list(), verbose = FALSE)
```

- Install R package sandwich and then call it.

Example:

```
## fit the 3 factor Fama French Model for IBM returns:
fit_ibm <- lm(ibm_x ~ Mkt_RF + SMB + HML)

## NeweyWest computes the NW SEs. It requires lags=L & suppression of prewhitening
NeweyWest(fit_ibm_ff3, lag = 4, prewhite = FALSE)
```

Note: It is usually found that the NW SEs are downward biased.

NW Estimator: Script in R

- You can also program the NW SEs yourself. In R:

```
NW_f <- function(y, X, b, lag)
{
T <- length(y);
k <- length(b);
yhat <- X%*%b
e <- y - yhat
hhat <- t(X)*as.vector(t(e))
G <- matrix(0,k,k)
a <- 0
w <- numeric(T)
while (a <= lag) {
  Ta <- T - a
  ga <- matrix(0,k,k)
  w[lag+1+a] <- (lag+1-a)/(lag+1)
  za <- hhat[,(a+1):T] %*% t(hhat[,1:Ta])
  ga <- ga + za
  G <- G + w[lag+1+a]*ga
a <- a+1
}
F <- t(X)%*%X
V <- solve(F)%*%G%*%solve(F)
 nw_se <- sqrt(diag(V))
 ols_se <- sqrt(diag(solve(F)*drop((t(e)%*%e))/(T-k)))
l_se = list(nw_se,ols_se)
return(l_se)
}

NW_f(y,X,b,lag=4)
```

NW Estimator: Application 1 – IBM

Example: We estimate the 3 factor F-F model for IBM returns:

```
> t_OLS
(Intercept)      Mkt_RF         SMB         HML
  -2.091345   16.032190   -2.628853   -1.655705
```
⟹ with SE_OLS: SMB significant at 1% level

```
NW <- NeweyWest(fit_ibm_ff3, lag = 4, prewhite = FALSE)        # with 4 lags
SE_NW <- diag(sqrt(abs(NW)))
t_NW <- b_ibm/SE_NW
> SE_NW
(Intercept)      Mkt_RF         SMB         HML
0.002527425 0.069918706 0.114355320 0.104112705
> t_NW
(Intercept)      Mkt_RF         SMB         HML
  -2.054010   13.020543   -1.935945   -1.336811
```
⟹ SMB close to significant at 5% level

- If we add more lags in the NW function (**lag** = **8**)

```
NW <- NeweyWest(fit_ibm_ff3, lag = 8, prewhite = FALSE)
SE_NW <- diag(sqrt(abs(NW)))
t_NW <- b_ibm/SE_NW
> t_NW
(Intercept)      Mkt_RF         SMB         HML
  -2.033648   12.779060   -1.895993   -1.312649
```
⟹ not very different results.

NW Estimator: Application 2 – i_{MX}

Example: Mexican short-term interest rates

```
NW <- NeweyWest(fit_i, lag = 4, prewhite = FALSE)        # with 4 lags
SE_NW <- diag(sqrt(abs(NW)))
t_NW <- b_i/SE_NW
> SE_NW
(Intercept)      us_i_1        e_mx        mx_I        mx_y
 0.01107069  0.55810758  0.01472961  0.51675771  0.93960295
> t_NW
(Intercept)      us_i_1        e_mx        mx_I        mx_y
  3.6332593   1.5388750  -0.7222770   6.4746121  -0.5305582
```
⟹ $i_{US,t}$ not longer significant at 10% level

- If we add more lags in the text (**lag = 8**)

```
NW <- NeweyWest(fit_i, lag = 8, prewhite = FALSE)
SE_NW <- diag(sqrt(abs(NW)))
t_NW <- b_i/SE_NW
> t_NW
(Intercept)      us_i_1        e_mx        mx_I        mx_y
  3.0174983   1.4318654  -0.8279016   6.5897816  -0.5825521
```
⟹ similar results.

NW Estimator: Remarks

• There are many estimators of $\boldsymbol{Q}^*$ based on a specific parametric model for $\boldsymbol{\Sigma}$, using time series models (Lecture 8). Thus, they are not *robust* to misspecification of (**A3'**). This is the appeal of White & NW.

• NW SEs are used almost universally in academia. However:
- NW SEs perform poorly in Monte Carlo simulations:
- NW SEs tend to be **downward biased**.
- The finite-sample performance of tests using NW SE is not well approximated by the asymptotic theory.
- Tests have size distortions.

• Q: What happens if we know the specific form of (**A3'**)?
We can do much better –i.e., more efficient- than using OLS with NW SEs. In this case, we can do Generalized LS (GLS), a method that delivers the most efficient estimators.

Generalized Least Squares (GLS)

• GRM: Assumptions (**A1**), (**A2**), (**A3'**) & (**A4**) hold. That is,

(**A1**) DGP: $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) $\text{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}] = \mathbf{0}$

(**A3'**) $\text{Var}[\boldsymbol{\varepsilon}|\boldsymbol{X}] = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$ ($\boldsymbol{\Omega}$ is symmetric $\Rightarrow \mathbf{T}'\mathbf{T} = \boldsymbol{\Omega}$)

(**A4**) $\boldsymbol{X}$ has full column rank –rank($\boldsymbol{X}$)= k–, where $T \geq k$.

• Suppose we know the form of (**A3'**). We can use this information to gain efficiency.

• We transform $\boldsymbol{y}$ & $\boldsymbol{X}$, in such a way, that we can do again OLS with the transformed data.

To do this transformation, we exploit a property of symmetric matrices, like the variance-covariance matrix, $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} \text{ is symmetric} \quad \Rightarrow \text{exists } \mathbf{T} \ni \quad \mathbf{T}'\mathbf{T} = \boldsymbol{\Omega} \quad \Rightarrow \mathbf{T}'^{-1}\,\boldsymbol{\Omega}\,\mathbf{T}^{-1} = \mathbf{I}$$

Generalized Least Squares (GLS)

- We transform the linear model in (**A1**) using $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$.

$$\mathbf{P} = \boldsymbol{\Omega}^{-1/2} \qquad \Rightarrow \mathbf{P}'\mathbf{P} = \boldsymbol{\Omega}^{-1}$$

$$\mathbf{P}\boldsymbol{y} = \mathbf{P}\boldsymbol{X}\beta + \mathbf{P}\boldsymbol{\varepsilon} \text{ or}$$

$$\boldsymbol{y}^* = \boldsymbol{X}^*\beta + \boldsymbol{\varepsilon}^*.$$

$$\text{E}[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}|\boldsymbol{X}^*] = \text{E}[\mathbf{P}\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'\mathbf{P}'|\boldsymbol{X}^*] = \mathbf{P}\,\text{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\text{X}]\,\mathbf{P}' = \sigma^2\mathbf{P}\,\boldsymbol{\Omega}\,\mathbf{P}'$$

$$= \sigma^2\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}\,\boldsymbol{\Omega}^{-1/2} = \sigma^2\,\mathbf{I}_T \qquad \Rightarrow \text{back to } (\mathbf{A3})$$

- The transformed model is homoscedastic: We have the CLM framework back $\Rightarrow$ we can use OLS!

$$\mathbf{b}_{GLS} = \mathbf{b}^* = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\,\boldsymbol{X}^{*\prime}\,\boldsymbol{y}^*$$

$$= (\boldsymbol{X}'\mathbf{P}'\,\mathbf{P}\boldsymbol{X})^{-1}\,\boldsymbol{X}'\mathbf{P}'\,\mathbf{P}\boldsymbol{y}$$

$$= (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$$

Generalized Least Squares (GLS)

Remarks:

– The transformed model is homoscedastic:

$$\text{Var}[\boldsymbol{\varepsilon}^*|\boldsymbol{X}^*] = \text{E}[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}|\boldsymbol{X}^*] = \mathbf{P}\text{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X}^*]\mathbf{P}' = \sigma^2\,\mathbf{P}\boldsymbol{\Omega}\mathbf{P}' = \sigma^2\,\mathbf{I}_T$$

– We have the CLM framework back: We do OLS with the transformed model, we call this OLS estimator, the GLS estimator:

$$\mathbf{b}_{GLS} = \mathbf{b}^* = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\,\boldsymbol{y}^* = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$$

– Key assumption: $\boldsymbol{\Omega}$ is known, and, thus, $\mathbf{P}$ is also known; otherwise we cannot transformed the model.

- Big Question: Is $\boldsymbol{\Omega}$ known?

Alexander C. Aitken (1895 –1967, NZ)

Generalized Least Squares (GLS) – Summary

- The GLS estimator is:

$$\mathbf{b}_{GLS} = (\boldsymbol{X}'\Omega^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\Omega^{-1}\boldsymbol{y}$$

Note I: $\mathbf{b}_{GLS} \neq \mathbf{b}$. $\mathbf{b}_{GLS}$ is **BLUE** by construction, $\mathbf{b}$ is not.

- Check unbiasedness:

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

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

$$\mathrm{E}[\mathbf{b}_{GLS} \,|\boldsymbol{X}] = \beta$$

- Efficient Variance

$\mathbf{b}_{GLS}$ is BLUE. The "best" variance can be derived from

$$\mathrm{Var}[\mathbf{b}_{GLS}\,|\boldsymbol{X}] = \sigma^2\,(\boldsymbol{X}^{*\prime}\boldsymbol{X}^{*})^{-1} = \sigma^2\,(\boldsymbol{X}'\Omega^{-1}\boldsymbol{X})^{-1}$$

Then, the usual OLS variance for $\mathbf{b}$ is biased and inefficient!

Generalized Least Squares (GLS) - Properties

- Steps for GLS:

Step 1. Find transformation matrix $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$ (in the case of heteroscedasticity, $\mathbf{P}$ is a diagonal matrix).

Step 2. Transform the model: $\boldsymbol{X}^* = \mathbf{P}\mathbf{X}$ & $\boldsymbol{y}^* = \mathbf{P}\boldsymbol{y}$.

Step 3. Do GLS; that is, OLS with the transformed variables.

- Key step to do GLS: **Step 1**, getting the transformation matrix:

$$\mathbf{P} = \boldsymbol{\Omega}^{-1/2}.$$

GLS – Relaxing Assumptions (A2) & (A4)

Technical detail: If we relax the CLM assumptions (**A2**) and (**A4**), as we did in Lecture 7-a, we only have asymptotic properties for GLS:

– Consistency - "*well behaved data*."

– Asymptotic distribution under usual assumptions.
(easy for heteroscedasticity, complicated for autocorrelation.)

– Wald tests and *F*-tests with usual asymptotic χ^2 distributions.

(Weighted) GLS: Pure Heteroscedasticity

• **Step 1.** Find the transformation matrix $\mathbf{P} = \mathbf{\Omega}^{-1/2}$.

$$(\textbf{A3'})\;\; Var[\varepsilon] = \mathbf{\Sigma} = \sigma^2\mathbf{\Omega} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ 0 & 0 & & 0 \\ 0 & 0 & \dots & \omega_T \end{bmatrix}$$

$$\mathbf{\Omega}^{-1/2} = \mathbf{P} = \begin{bmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1/\sqrt{\omega_T} \end{bmatrix}$$

• **Step 2.** Now, transform $\boldsymbol{y}$ & $\mathbf{X}$:

$$\boldsymbol{y}^* = \mathbf{P}\boldsymbol{y} = \begin{bmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1/\sqrt{\omega_T} \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_T/\sqrt{\omega_T} \end{bmatrix}$$

(Weighted) GLS: Pure Heteroscedasticity

- **Step 2 (continuation).** Each observation of $\boldsymbol{y}$, y_i, is divided by $\sqrt{\omega_i}$. Similar transformation occurs with **X**:

$$\boldsymbol{X}^* = \mathbf{P}\mathbf{X} = \begin{bmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1/\sqrt{\omega_T} \end{bmatrix} * \begin{bmatrix} 1 & x_{21} & \cdots & x_{k1} \\ 1 & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2T} & \cdots & x_{kT} \end{bmatrix} =$$

$$= \begin{bmatrix} 1/\sqrt{\omega_1} & x_{21}/\sqrt{\omega_1} & \dots & x_{k1}/\sqrt{\omega_1} \\ 1/\sqrt{\omega_2} & x_{22}/\sqrt{\omega_2} & \dots & x_{k2}/\sqrt{\omega_2} \\ \vdots & \vdots & \dots & \vdots \\ 1/\sqrt{\omega_T} & x_{2T}/\sqrt{\omega_T} & \dots & x_{kT}/\sqrt{\omega_T} \end{bmatrix}$$

- **Step 3.** We do GLS (OLS with the transformed variables):

$$\mathbf{b}_{GLS} = \mathbf{b}^* = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\,\boldsymbol{y}^* = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$$

(Weighted) GLS: Pure Heteroscedasticity

- In the case of heteroscedasticity, GLS is also called *Weighted Least Squares* (WLS): Think of $1/\sqrt{\omega_i}$ as weights. The GLS estimator is:

$$\mathbf{b}_{GLS} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y} = \left(\sum_{i=1}^{T}\frac{1}{\omega_i}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\sum_{i=1}^{T}\frac{1}{\omega_i}\boldsymbol{x}_i y_i$$

Observations with lower (bigger) variances –i.e., lower (bigger) ω_i– are given higher (lower) weights in the sums:

More precise observations, more weight!

- The GLS variance is given by:

$$\hat{\sigma}^2_{GLS} = \frac{\sum_{i=1}^{T}\left(\frac{y_i - \boldsymbol{x}_i'\mathbf{b}_{GLS}}{\omega_i}\right)^2}{T-k}$$

(Weighted) GLS: Pure Heteroscedasticity

Example: Suppose we believe that $(r_{m,t} - r_f)^2$ drives the variance for DIS excess returns. Suppose we assume:
(**A3'**) $\sigma_t^2 = (r_{m,t} - r_f)^2$.

Steps for GLS:

1. Find transformation matrix, **P**, with i^{th} diagonal element: $1/\sqrt{\sigma_i^2}$
2. Transform model: Each y_i and x_i is divided ("weighted") by
   $\sigma_i = \text{sqrt}[(r_{m,i} - r_f)^2]$.
3. Do GLS, that is, OLS with transformed variables.

```
T <- length(dis_x)
Mkt_RF2 <- Mkt_RF^2                                   # (A3')
y_w <- dis_x/sqrt(Mkt_RF2)                            # transformed y = y*
x0 <- matrix(1,T,1)
xx_w <- cbind(x0, Mkt_RF, SMB, HML)/sqrt(Mkt_RF2)     # transformed X = X*
fit_dis_wls <- lm(y_w ~ xx_w - 1)                     # GLS
```

(Weighted) GLS: Pure Heteroscedasticity

Example (continue):

```
> summary(fit_dis_wls)

Call:
lm(formula = y_w ~ xx_w)

Residuals:
    Min      1Q  Median      3Q     Max
-59.399  -0.891   0.316   1.503  77.434

Coefficients:
              Estimate   Std. Error t value Pr(>|t|)
xx_w         -0.006607   0.001586  -4.165 3.59e-05 ***
xx_wMkt_RF    1.588057   0.334771   4.744 2.66e-06 ***   ⇒ OLS b: 1.26056
xx_wSMB      -0.200423   0.067498  -2.969  0.00311 **    ⇒ OLS b: -0.028993
xx_wHML      -0.042032   0.072821  -0.577  0.56404       ⇒ OLS b:  0.174545
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.984 on 566 degrees of freedom
Multiple R-squared:  0.09078,    Adjusted R-squared:  0.08435
F-statistic: 14.13 on 4 and 566 DF,  p-value: 5.366e-11
```