# Lecture 7-b
# Departures from OLS Assumptions

Brooks (4th edition): Chapter 5

 1

## Review - CLM: Departures from (A3)

• The CLM assumes

(**A3**) Var[$\boldsymbol{\varepsilon}$|$\mathbf{X}$] = $\sigma^2 \mathbf{I}_T$

Now, we will assume:

(**A3'**) Var[$\boldsymbol{\varepsilon}$|$\mathbf{X}$] = $\boldsymbol{\Sigma}$ (also written as $\sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega} \neq \mathbf{I}_T$)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

• Two Leading Cases:

– **Pure heteroscedasticity**: We model only the diagonal elements.

– **Pure autocorrelation**: We model only the off-diagonal elements.

11

## Review - CLM: Heteroscedasticity

- **Pure heteroscedasticity**:

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

$$\Rightarrow \text{Var}[\varepsilon_i \,|\, \mathbf{X}] = \sigma_i^2$$

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

- 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**:

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

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

- 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 - Testing for Heteroscedasticity

• Recall that **b** is unbiased in the presence of heteroscedasticity. We test for heteroscedasticity for efficiency and to do proper inference.

• We want to test: $H_0$: $E(\varepsilon_i^2) = \sigma^2$ for all $i$.
$H_1$: $E(\varepsilon_i^2) = \sigma_i^2 \neq \sigma^2$ 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 $\sigma_i^2$: a particular variable, say $\boldsymbol{x}_j$, a regime (before & after some event), or past volatility, $\sigma_{t-j}^2$.

• 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 $H_0$: $\sigma_i^2 = \sigma^2$
$H_1$: $\sigma_i^2 = f(\boldsymbol{x}_j)$ $\boldsymbol{x}_j$: 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 ($d \approx$ 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

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

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 $\sigma_i^2$, the $z_i$'s, under $H_1$.

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

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

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

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

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

- **Step 3**. Calculate

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

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:

$\boldsymbol{y} = \mathbf{X}\,\beta + \varepsilon.$ 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: $\text{LM} = T\,R_{e2}^2 \xrightarrow{d} \chi_m^2.$

8

## Review: Finding Auto/cross-correlation

- We look for autocorrelation in the error structure. Usually, we model autocorrelation using two models: autoregressive (AR) and moving averages (MA).

- In an AR model, the errors, $\varepsilon_t$, show a correlation over time. 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:

$H_0$: $\rho_1 = \ldots = \rho_p = 0$
$H_1$: at least one $\rho_i \neq 0$, for $i = 1, 2, \ldots, p$

Under $H_0$, Breusch & Godfrey use OLS residuals, $e_i$, to construct an LM test (**BG test**), similar to the BP test. 9

## Review: LM Test for Autocorrelation

- Under the null hypothesis of no AR($p$) we have

$H_0$: $\rho_1 = \ldots = \rho_p = 0$.
$H_1$: at least one $\rho_i \neq 0$, for $i = 1, 2, \ldots, p$

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

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

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

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

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

    $\text{LM} = (T - p) * R_e^2 \xrightarrow{d} \chi_p^2$. 10

## Review: LM Test for Autocorrelation

**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
```

11

## Review: LM Test for Autocorrelation

**Example (continuation):**

```
> r2_e1
[1] 0.0303721
> (T-p_lag)
[1] 557
lm_t <- (T-p_lag ) * r2_e1                      # LM-test with p lags
> lm_t
[1] 16.91726
df <- ncol(e_lag)                               # degrees of freedom for the LM Test
> 1-pchisq(lm_t,df)
[1] 0.1560063
```

LM-AR(**12**) Test: **16.91726** ⇒ cannot reject $H_0$ at 5% level (*p-value* > .05)

- If we run the test with p = **4 lags**, we get

LM-AR(4) Test: **2.9747** (*p-value* = 0.56) ⇒ cannot reject $H_0$ at 5% level

12

## Review: LM Test for Autocorrelation

**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 = 16.259, df  = 12, p-value = 0.1797
```

(minor difference with the previous test, likely due to multiplication by $T$. 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: **16.83253** ⇒ very similar. Not entirely correct, but it works well.

13

## Review: LM Test for Autocorrelation

- 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: for daily data, 5 or 20 lags; for weekly, 4 or 12 lags; for monthly data, 12 lags; for quarterly data, 4 lags.

14

## 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 $T \rightarrow \infty$, $d \approx 2(1 - \rho_1)$.

• $\rho_1$ is estimated by the sample correlation $r$.

• Under $H_0$, $\rho_1$=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 **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*.

15

## 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] 2.042728       ⇒ 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] 2.048281
```

• Similar finding for Disney returns:

```
> DW
        [,1]
[1,] 2.1609          ⇒ 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 = 2.0427, p-value = 0.7087
```

16

## 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
alternative hypothesis: true autocorrelation is greater than 0
```

⇒ not significant at 5% level.

17

## 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 $H_0$: $\rho_1 = ... = \rho_p = 0$ 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.

Then, under $H_0$:

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

18

## Testing for Autocorrelation: Portmanteau tests

• Ljung-Box (1978) test (LB test).
A variation of the Box-Pierce test. It has a small sample correction.

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

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

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 **X**. Thus, it is more powerful..

19

## 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**:

```
RSS <- sum(e_ibm^2)
r_sum <- 0
lb_sum <- 0
p_lag <- 12
a <- 1
while (a <= p_lag) {
        za <- as.numeric(t(e_ibm[(p_lag+1):T]) %*% e_ibm[a:(T-p_lag+a-1)])
  r_sum <- r_sum + (za/RSS)^2                    #sum cor(e[(p_lag+1):T], e[a:(T-p_lag+a-1)])^2
  lb_sum <- lb_sum + (za/RSS)^2/(T-a)
a <- a + 1
}
Q <- T*r_sum
LB <- T*(T-2)*lb_sum
> Q
[1] 16.39559          (p-value = 0.1737815)          ⇒ cannot reject H0 at 5% level.
> LB
[1] 16.46854          (p-value = 0.1707059)          ⇒ cannot reject H0 at 5% level.
```

20

## Testing for Autocorrelation: Portmanteu tests

**Example (continuation):** The *Box.test* function computes Q & LB:

- Q test

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

        Box-Pierce test

data:  e_ibm
X-squared = 16.304, df = 12, p-value = 0.1777
```

- LB test

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

        Box-Ljung test

data:  e_ibmX-squared = 16.61, df = 12, p-value = 0.1649
```

Note: There is a minor difference between the previous code and the code in Box.test. They are based on how the correlations of e are computed (centered around the mean, or assumed zero mean).

21

## Testing for Autocorrelation: Portmanteu tests

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

- For **DIS** (e_dis), we get:

> Q
[1] **25.563** (*p-value* = **0.01237**) ⇒ reject $H_0$ at 5% level.
> LB
[1] **25.879** (*p-value* = **0.01117**) ⇒ reject $H_0$ at 5% level.

- For **GE** (e_ge), we get:

> Q
[1] **27.087** (*p-value* = **0.007507**) ⇒ reject $H_0$ at 5% level.
> LB
[1] **27.523** (*p-value* = **0.006493**) ⇒ reject $H_0$ at 5% level.

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

22

## 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:

```
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.

```
> 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.

23

## Testing for Autocorrelation: Portmanteu 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, see Diebold (1986), Robinson (1991), Lobato et al. (2001). The "robustified" Portmanteau statistic uses $\tilde{r}_j$ instead of $r_j$:

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

Thus, for Q we have:

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

24

## 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 $\rho$ and AIC for bigger $\rho$).

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

25

## Testing for Autocorrelation: Portmanteau 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
```

26

## 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 $\sigma_i^2$.

- We use a Portmanteu test on the squared residuals to check for a particular kind of heteroscedasticity: the variance, $\sigma_i^2$, 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).

27

## 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
```

28

## 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. 29

## Generalized Regression Model (GRM)

- Now, we go back to the CLM Assumptions:

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

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

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

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

-a ($T$x$T$) 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[$\boldsymbol{\varepsilon}$|**X**] = $\sigma^2 \mathbf{I}_T$ $\Rightarrow$ Var[**b**|**X**] = $\sigma^2$ (**X′X**)$^{-1}$.
- Now, we have (**A3'**) Var[$\boldsymbol{\varepsilon}$|**X**] = $\boldsymbol{\Sigma}$
- Recall **b** = $\boldsymbol{\beta}$ + (**X′X**)$^{-1}$ **X′** $\boldsymbol{\varepsilon}$.
- The true variance of **b** under (**A3'**) should be:

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

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

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

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

## GR Model: True Variance for b

**Example (continuation):**

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

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

$$\text{Var}[\mathbf{b}|\mathbf{X}] = \frac{\sigma^2}{\sum_{i=1}^{T}(x_i-\bar{x})^2} \neq \text{Var}_T[\mathbf{b}|\mathbf{X}]$$

- Under (**A3'**), the usual OLS estimator of Var[**b**|**X**] –i.e., $s^2$ (**X′X**)$^{-1}$– is *biased*. If we want to use OLS for inferences (say, with *t-test* or *F-test*), we need to estimate Var$_T$[**b**|**X**].
- That is, we need to estimate the unknown $\boldsymbol{\Sigma}$. But, it has $T*(T+1)/2$ parameters. Too many parameters to estimate with only $T$ observations!

## GR Model: Robust Covariance Matrix

- We will not be estimating $\mathbf{\Sigma}$. Impossible with $T$ data points.
- We will estimate $\mathbf{X}'\mathbf{\Sigma}\mathbf{X} = \sum_{i=1}^{T}\sum_{j=1}^{T} \sigma_{ij}\, \boldsymbol{x}_i\, \boldsymbol{x}_j{}'$, 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 $\text{Var}_T[\mathbf{b}|\mathbf{X}]$:

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

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

## Covariance Matrix: The White Estimator

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

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

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

where

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

- Q: How do we estimate $\sigma_i^2$?

## Covariance Matrix: The White Estimator

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

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

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

with $\mathbf{S_0} = (1/T) \sum_{i=1}^{T} e_i^2\ \boldsymbol{x}_i\ \boldsymbol{x}_i'$

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

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

which we estimate using OLS residuals, $e_i$:

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

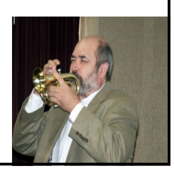
## Covariance Matrix: The White Estimator

• White (1980) shows that a consistent estimator of $\text{Var}_T[\mathbf{b}\,|\,\mathbf{X}]$ is obtained if $e_i^2$ is used as an estimator of $\sigma_i^2$. 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 the *White estimator* (also known as *Eiker-Huber-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"):

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

(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 $\mathbf{x}_i$ or $\mathbf{x}_i^2$, 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 ($i_{MX}$) with a linear model including US interest rates, changes in exchange rates (MXN/USD), Mexican inflation and Mexican GDP growth, using quarterly data 1978:II – 2020:II (T=166):

$$i_{MX,t} = \beta_0 + \beta_1\, i_{US,t} + \beta_2\, e_t + \beta_3\, mx\_I_t + \beta_4\, mx\_y_t + \varepsilon_t$$

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

us_i <- FMX_da$US_int                    # US short-term interest rates (iUS)
mx_CPI <- FMX_da$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 (iMX)
mx_GDP <- FMX_da$MX_GDP                  # Mexican GDP
S_mx <- FMX_da$MXN_USD                   # St = 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 St.
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[**ε**|**X**] = **Σ**

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

Now, we need to estimate

$$\mathbf{Q}^* = (1/T)\,\mathbf{X'\Sigma X} = (1/T)\sum_{i=1}^{T}\sum_{j=1}^{T}\sigma_{ij}\,\boldsymbol{x}_i\,\boldsymbol{x}_j'$$
$$= (1/T)\sum_{i=1}^{T}\{\sigma_{i1}\,\boldsymbol{x}_i\,\boldsymbol{x}_1 + \sigma_{i2}\,\boldsymbol{x}_i\,\boldsymbol{x}_2 + \cdots + \sigma_{iT}\,\boldsymbol{x}_i\,\boldsymbol{x}_T\}$$

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

## Newey-West Estimator

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

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

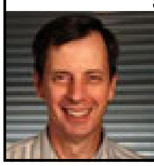
$\Rightarrow$ natural estimator of $\mathbf{Q}^*$:

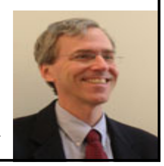
$$\mathbf{S}_T = (1/T)\ \textstyle\sum_{i=1}^{T} \sum_{j=1}^{T}\ e_i\ e_j\ \boldsymbol{x}_i\ \boldsymbol{x}_j'$$

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

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

- 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 $\mathbf{Q}^*$:

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

**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:

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

We add the sum of the autocovariances of $w_i$ $(= x_i\ e_i)$ to the White estimator of $\text{Var}_T[\mathbf{b}\,|\,\mathbf{X}]$. If the autocovariances of $w_i$ $(= x_i\ e_i)$ 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:

$\mathbf{S_0} = \frac{1}{T}\sum_{t=1}^{T} e_t^2\, \boldsymbol{x_t}\, \boldsymbol{x_t}'$ – the White estimator.

(2) Add the Autocorrelation Component

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

where

$k(\frac{j}{L(T)}) = \frac{L+1-|j|}{L+1}$ –decaying weights (*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,

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

$k(2) = 29/31 = 0.9354839$ $\quad k(\frac{j}{L(T)}) = \frac{L+1-|j|}{L+1}$

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

## Newey-West Estimator

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

Then,

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

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

- Asymptotic inferences can be based on OLS **b**, with *t-tests* and *Wald tests* using N(0,1) and $\chi^2$ 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:

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

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

Technical Detail: Above, it is mentioned that the asymptotics need that as $L$ & $T \to \infty$, and $L/T \to 0$, to get $\mathbf{S}_T \to \mathbf{Q}^*$. 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 $\mathbf{Q}^*$ 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: $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correctly specified.

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

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

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

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

• When we know (**A3'**), we transform $\boldsymbol{y}$ & $\mathbf{X}$, 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, $\boldsymbol{\Omega}$:

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

## Generalized Least Squares (GLS)

Note: $\boldsymbol{\Omega}$ can be decomposed as

$$\boldsymbol{\Omega} = \mathbf{T}'\,\mathbf{T} \quad \text{(think of } \mathbf{T} \text{ as } \boldsymbol{\Omega}^{1/2}) \qquad \Rightarrow \mathbf{T}'^{-1}\,\boldsymbol{\Omega}\,\mathbf{T}^{-1} = \mathbf{I}$$

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

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

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

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

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

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

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

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

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

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

## Generalized Least Squares (GLS)

Remarks:

– The transformed model is homoscedastic:

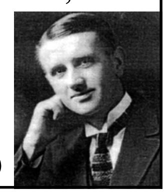
$\text{Var}[\boldsymbol{\varepsilon}^*|\mathbf{X}^*] = \text{E}[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}|\mathbf{X}^*] = \mathbf{P}\text{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}^*]\mathbf{P}' = \sigma^2\,\mathbf{P}\boldsymbol{\Omega}\mathbf{P}' = \sigma^2\,\mathbf{I}_T$

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

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

– 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)

- The GLS estimator is:

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

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

- Check unbiasedness:

$$\mathbf{b}_{GLS} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{\Omega}^{-1}\,\boldsymbol{y} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{\Omega}^{-1}\,(\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$
$$= \boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{\Omega}^{-1}\,\boldsymbol{\varepsilon}$$
$$\text{E}[\mathbf{b}_{GLS}\,|\mathbf{X}] = \boldsymbol{\beta}$$

- Efficient Variance

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

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

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