# VI. Problems

**A. Multicollinearity (Chapter 8)** ✓

**B. Random Regressors** ✓

**C. Heteroskedasticity (Chapter 9)**

**D. Autocorrelation (Chapter 10)**

## *B. Heteroskedasticity*

CRMA # 4

**1. Definition**: One of our CRM assumptions is ***violated***. The disturbances do not have constant variances: $Var(u_i) = E[u_i^2] = \sigma_i^2$

Examples: $Var(u_i) = \sigma^2 X_i$ ← as $X \uparrow$ $\sigma_i^2 \uparrow$

or: $Var(u_i) = \sigma^2 / X_i$ ← as $X \uparrow$ $\sigma_i^2 \downarrow$

## 2. Consequences (Eg. Simple Regression)

- OLS estimates the variance as: $s_{b_1}^2 = \dfrac{\hat{\sigma}^2}{\sum x_i^2}$

OLS assumes all CRMA are true.

$E[b_1] = \beta_1$ still unbiased

- CRMA#4 is violated – the correct estimated variance is: $s_{b_1}^2 = \dfrac{\sum_i x_i^2 \hat{\sigma}_i^2}{\left(\sum_i x_i^2\right)^2}$

OLS uses wrong formula for $s_{b_1}^2$ !!

- ***OLS uses the wrong formula*** – OLS results would give you the "***wrong standard errors***"

*

## 3. Diagnosis: How can we tell if we have this problem?

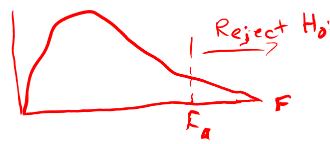
- ✓ Nothing on your printout tells you – you have to look.
- ✓ Plot: Errors (or $e_i^2$) vs. Xs, or Y.
- ✓ Regressions: $\mathbf{e_i^2}$ used as dependent variable.
  - ***Hypothesis:***
  - ***Bruesch-Pagan Test***: regress $\mathbf{e_i^2}$ on Xs and other factors.
  - ***White's Test***: regress $\mathbf{e_i^2}$ on $X_k$s, $X_k \cdot X_l$ and $X_k^2$s.
    - **Statistical Test:** $\chi_{calc}^2 = nR^2 \sim \chi_K^2$
    - df for the chi-square is the number of coefficients estimated in the $\mathbf{e_i^2}$ regression.

3. Diagnosis: How can we tell if we have this problem?

- ✓ Group-wise heteroskedasticity: consider 2 groups
  - ***Hypothesis:*** $H_0: \sigma_1^2 = \sigma_2^2$ ; $H_a: \sigma_1^2 \neq \sigma_2^2$
  - ***Goldfeld-Quandt Test*:** Compare variances for two groups.
    - **Estimate two regressions:** $\Rightarrow \hat{\sigma}_1^2$, $\Rightarrow \hat{\sigma}_2^2$
    - **Statistical Test:**

always put biggest $\hat{\sigma}^2$ on top!!

$$F_{calc} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{(n_1 - K_1 - 1,\, n_2 - K_2 - 1)}$$




## 4. Solutions:

- ***OLS – calculate proper standard errors*** asymptotic covariance; this is no longer BLUE
- ***Generalized Least Squares***: These are the ***BLUE*** estimators. But, you must know $\sigma_i^2$.
- "***Feasible Generalized Least Squares***"
  - Weight the dependent and independent variables.
  - Estimate using the weighted variables.
  - Example: suppose $E[u_i^2] = \sigma_i^2 = X_i \sigma^2$ assume a relationship between $\sigma_i^2$ and $X_i$

    Transform data: $Y_i^* = \frac{Y_i}{\sqrt{X_i}}$ ; $X_i^* = \frac{X_i}{\sqrt{X_i}}$ ; $u_i^* = \frac{u_i}{\sqrt{X_i}}$.

    Estimate:

$$Y_i^* = \beta_0^* + \beta_1 X_i^* + u_i^*$$

```
proc reg data=new2;
  model wage = yrsed exp expsq fe expfe/spec;
    output out=ests residual=e;
run;
```

SAS

save residuals, or errors

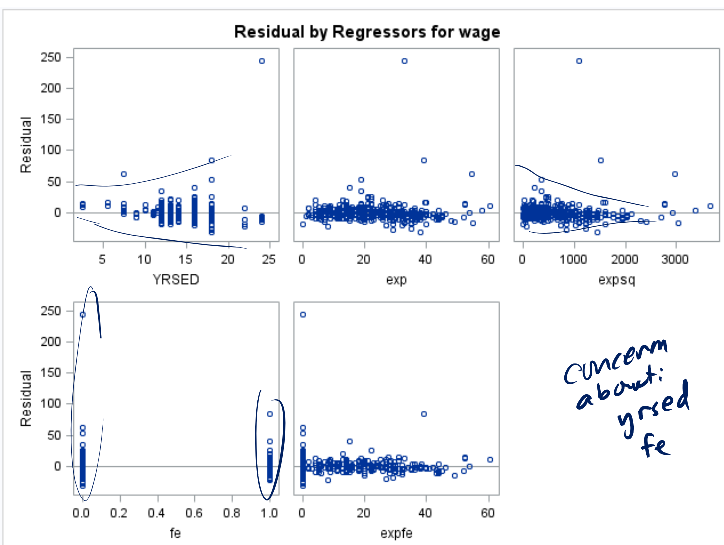
OLS

```
                         Model: MODEL1
                   Dependent Variable: wage
              Number of Observations Used        410

                     Analysis of Variance
                         Sum of           Mean
Source             DF   Squares         Square    F Value   Pr > F
Model               5     20635     4126.91882      16.07   <.000
Error             404    103781      256.88486
Corrected Total   409    124416

          Root MSE              16.02763    R-Square     0.1659
          Dependent Mean        16.11876    Adj R-Sq     0.1555
          Coeff Var             99.43465

                         Parameter Estimates
                    Parameter       Standard
Variable      DF     Estimate          Error    t Value    Pr > |t|
Intercept      1    -16.68316        5.23960      -3.18      0.0016
YRSED          1      2.00215        0.27844       7.19      <.0001
exp            1      0.46415        0.24737       1.88      0.0613
expsq          1     -0.00322        0.00457      -0.70      0.4812
fe             1     -4.69790        3.38288      -1.39      0.1657
expfe          1     -0.06215        0.13623      -0.46      0.6485
```




concern about:
yrsed
fe

# Example: Wage model

check this

- Heteroskedasticity arises due to level of education (Yrsed).
- Yrsed varies from a few years to 24. Many levels.
- What might we assume about the form of heteroskedasticity? $\sigma_i^2 = \frac{\sigma^2}{yrsed}$
- How would we test for this form of heteroskedasticity?

BP or ~~White's test~~

$e_i^2 = a_0 + a_1 yrsed + a_2 exp + \dots +$

WARNING: The average covariance matrix for the SPEC test has been deemed singular which violates an assumption of the test. Use caution when interpreting the results of the test.

**The SAS System**

The REG Procedure
Model: MODEL1
Dependent Variable: wage

Test of First and Second Moment Specification

| DF | Chi-Square | Pr > ChiSq |
|---|---|---|
| 16 | 15.20 | 0.5102 |

$\Rightarrow$ Fail to reject $H_0: \sigma_i^2 = \sigma^2$

```
proc reg data=new2;    ** where wage gt 0;
  model wage = yrsed exp expsq fe expfe/spec ;
    output out=ests residual=e;
run;

data whites; set ests;
 esq = e**2;
 edsq = yrsed**2; edexp = yrsed*exp;  edfe = yrsed*fe;
run;
```

going to "White's test"

```
proc reg data=whites;
  model esq = yrsed edsq exp expsq edexp fe edfe expfe ;
run;
```

auxiliary Regression

Model: MODEL1
Dependent Variable: esq
Number of Observations Used 410

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 8 | 742807074 | 92850884 | 12.69 | <.0001 |
| Error | 401 | 2934619424 | 7318253 | | |
| Corrected Total | 409 | 3677426498 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2705.22696 | R-Square | 0.2020 |
| Dependent Mean | 253.12557 | Adj R-Sq | 0.1861 |
| Coeff Var | 1068.72925 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 19943 | 3288.46649 | 6.06 | <.0001 |
| YRSED | 1 | -2396.48670 | 343.91431 | -6.97 | <.0001 |
| edsq | 1 | 68.48255 | 9.14762 | 7.49 | <.0001 |
| exp | 1 | -462.09615 | 90.53337 | -5.10 | <.0001 |
| expsq | 1 | 1.95919 | 0.87955 | 2.23 | 0.0265 |
| edexp | 1 | 29.47714 | 4.36668 | 6.75 | <.0001 |
| fe | 1 | 2834.09781 | 1576.96005 | 1.80 | 0.0731 |
| edfe | 1 | -173.93544 | 96.96479 | -1.79 | 0.0736 |
| expfe | 1 | -22.52145 | 24.12013 | -0.93 | 0.3510 |

# BP or White's Test

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

$$\chi^2_{calc} = nR^2 \sim \chi^2_K$$

From our *esq* regression: $R^2 = 0.202$ and $n = 410$

$$n \cdot R^2 = 410 \cdot (0.202) = 82.82$$

p-value $< 0.0001$

compare to $\chi^2_8 \cong 15.50$ $\quad \alpha = 0.05$

1. Use OLS – properly compute standard errors
2. GLS would be BLUE, but we don't know $\sigma_i^2$
3. Feasible GLS – "asymptotically Best"

```
proc reg data=new2;
  model wage = yrsed exp expsq fe expfe/acov;
  test yrsed=0, exp=0, expsq=0, fe=0, expfe=0;
run;
```

asymptotic cov.

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 20635 | 4126.91882 | 16.07 | <.0001 |
| Error | 404 | 103781 | 256.88486 | | |
| Corrected Total | 409 | 124416 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 16.02763 | R-Square | 0.1659 |
| Dependent Mean | 16.11876 | Adj R-Sq | 0.1555 |
| Coeff Var | 99.43465 | | |

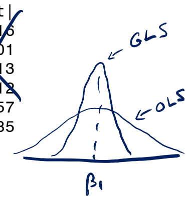
ok - unbiased

OLS st. errors

wrong!! ☹

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -16.68316 | 5.23960 | -3.18 | 0.0016 |
| YRSED | 1 | 2.00215 | 0.27844 | 7.19 | <.0001 |
| exp | 1 | 0.46415 | 0.24737 | 1.88 | 0.0613 |
| expsq | 1 | -0.00322 | 0.00457 | -0.70 | 0.4812 |
| fe | 1 | -4.69790 | 3.38288 | -1.39 | 0.1657 |
| expfe | 1 | -0.06215 | 0.13623 | -0.46 | 0.6485 |




---Heteroscedasticity Consistent---

| Variable | Label | DF | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 13.52539 | -1.23 | 0.2181 |
| YRSED | YRSED | 1 | 0.80913 | 2.47 | 0.0138 |
| exp | | 1 | 0.23057 | 2.01 | 0.0448 |
| expsq | | 1 | 0.00455 | -0.71 | 0.4787 |
| fe | | 1 | 2.69616 | -1.74 | 0.0822 |
| expfe | | 1 | 0.15507 | -0.40 | 0.6888 |

proper std. errors for OLS

## Test for the Regression:

Test 1 Results for Dependent Variable wage

| Source | DF | Mean Square | F Value | Pr > F |
|---|---|---|---|---|
| Numerator | 5 | 4126.91882 | 16.07 | <.0001 |
| Denominator | 404 | 256.88486 | | |

Dependent Variable: wage
Test 1 Results using ACOV estimates

✓

| DF | Chi-Square | Pr > ChiSq |
|---|---|---|
| 5 | 39.97 | <.0001 |

$H_0: \beta_1 = \beta_2 = \beta_3 = \delta = \gamma = 0$; $H_a$: at least one $\neq 0$

Reject $H_0$:

```
proc glm data=new2;
  model wage = yrsed exp expsq fe expfe;
  weight yrsed;
run;
```

FGLS

$\sigma_i^2 = \frac{\sigma^2}{yrsed}$

SAS multiplies by $\sqrt{yrsed}$ to transform data

The GLM Procedure

Dependent Variable: wage

**Weight: wt2sq**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 449128.991 | 89825.798 | 17.49 | <.0001 |
| Error | 404 | 2074667.013 | 5135.314 | | |
| Corrected Total | 409 | 2523796.003 | | | |

| R-Square | Coeff Var | Root MSE | wage Mean |
|---|---|---|---|
| 0.177958 | 414.1508 | 71.66111 | 17.30314 |

correct

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Intercept | -27.48868249 | 6.42718367 | -4.28 | <.0001 |
| YRSED | 2.50557801 | 0.32449684 | 7.72 | <.0001 |
| exp | 0.76424120 | 0.30933638 | 2.47 | 0.0139 |
| expsq | -0.00798781 | 0.00589328 | -1.36 | 0.1760 |
| fe | -3.70722012 | 4.07196043 | -0.91 | 0.3631 |
| expfe | -0.11840736 | 0.16850020 | -0.70 | 0.4826 |

# Correcting for Heteroskedasticity – OLS

**FGLS Estimates**

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Intercept | -27.48868249 | 6.42718367 | -4.28 | <.0001 |
| YRSED | 2.50557801 | 0.32449684 | 7.72 | <.0001 |
| exp | 0.76424120 | 0.30933638 | 2.47 | 0.0139 |
| expsq | -0.00798781 | 0.00589328 | -1.36 | 0.1760 |
| fe | -3.70722012 | 4.07196043 | -0.91 | 0.3631 |
| expfe | -0.11840736 | 0.16850020 | -0.70 | 0.4826 |

| Variable | DF | Estimate | Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -16.68316 | 5.23960 | -3.18 | 0.0016 |
| YRSED | 1 | 2.00215 | 0.27844 | 7.19 | <.0001 |
| exp | 1 | 0.46415 | 0.24737 | 1.88 | 0.0613 |
| expsq | 1 | -0.00322 | 0.00457 | -0.70 | 0.4812 |
| fe | 1 | -4.69790 | 3.38288 | -1.39 | 0.1657 |
| expfe | 1 | -0.06215 | 0.13623 | -0.46 | 0.6485 |

---Heteroscedasticity Consistent--

| Variable | Label | DF | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | Intercept | 1 | ~~13.52539~~ | -1.23 | 0.2181 |
| YRSED | YRSED | 1 | 0.80913 | 2.47 | 0.0138 |
| exp | | 1 | 0.23057 | 2.01 | 0.0448 |
| expsq | | 1 | 0.00455 | -0.71 | 0.4787 |
| fe | | 1 | 2.69616 | -1.74 | 0.0822 |
| expfe | | 1 | 0.15507 | -0.40 | 0.6888 |

# Group Heteroskedasticity

```
** Goldfeld-Quandt for Male/Female groups;

Proc reg data=new2; where fe = 1;  title 'wages for females';
  model wage = yrsed exp expsq ;
run;

proc reg data=new2; where fe = 0;  title 'wages for males';
  model wage = yrsed exp expsq ;
run;
quit;
```

$\Rightarrow \hat{\sigma}_f^2$

$\Rightarrow \hat{\sigma}_m^2$

what do need?

$H_0: \sigma_f^2 = \sigma_m^2$

$H_a: \sigma_f^2 \neq \sigma_m^2$

$F_{calc} = \dfrac{\hat{\sigma}_i^2}{\hat{\sigma}_j^2}$

wages for females

The REG Procedure
Dependent Variable: wage

Number of Observations Used 207

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr >F |
|---|---|---|---|---|---|
| Model | 3 | 2972.68463 | 990.89488 | 11.19 | <.0001 |
| Error | 203 | 17969 | 88.51654 $= \hat{\sigma}_f^2$ | | |
| Corrected Total | 206 | 20942 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 9.40832 | R-Square | 0.1420 |
| Dependent Mean | 12.46035 | Adj R-Sq | 0.1293 |
| Coeff Var | 75.50608 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -11.19235 | 4.14088 | -2.70 | 0.0075 |
| YRSED | 1 | 1.30235 | 0.25468 | 5.11 | <.0001 |
| exp | 1 | 0.39238 | 0.18732 | 2.09 | 0.0374 |
| expsq | 1 | -0.00403 | 0.00360 | -1.12 | 0.2643 |

**wages for males**
The REG Procedure
Dependent Variable: wage
Number of Observations Used 203

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 13234 | 4411.42857 | 10.37 | <.0001 |
| Error | 199 | 84645 | 425.35034 $= \hat{\sigma}_m^2$ | | |
| Corrected Total | 202 | 97879 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 20.62402 | R-Square | 0.1352 |
| Dependent Mean | 19.84924 | Adj R-Sq | 0.1222 |
| Coeff Var | 103.90331 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -24.82416 | 8.87010 | -2.80 | 0.0056 |
| YRSED | 1 | 2.50202 | 0.46857 | 5.34 | <.0001 |
| exp | 1 | 0.54066 | 0.46042 | 1.17 | 0.2417 |
| expsq | 1 | -0.00420 | 0.00890 | -0.47 | 0.6375 |

# Goldfeld-Quandt Test:

$$F_{calc} = \frac{\hat{\sigma}_m^2}{\hat{\sigma}_f^2} = \frac{425.35}{88.52} = \underline{4.81}$$

$$F_{(0.05,\ 199,\ 203)} = 1.25$$

Reject $H_0: \sigma_f^2 = \sigma_m^2$

## Correcting Group Heteroskedasticity: GLS

```
data grouphet; set new2;
 pfe = (fe*88.52)**0.5;
 pm  = (m*425.35)**0.5;
 wtp = 1/(pfe + pm);  p2=wtp**2;
 edstar = yrsed*wtp; expstar = exp*wtp; expsqstar = expsq*wtp;
 festar = fe*wtp; expfestar = expfe*wtp;
 wagestar = wage*wtp;
run;
proc reg data=grouphet;
 model wagestar = edstar expstar expsqstar festar expfestar;

run;

proc glm data=grouphet;
 model wage = yrsed exp expsq fe expfe;
 weight p2;
run;
```

$\frac{1}{\hat{\sigma}_f^2 \cdot fe + (1-fe)\hat{\sigma}_m^2}$

to get the correct with each type of person

by hand

my intercept

model wagestar = wtp edstar expstar expsqstar festar expfestar / noint;

Dependent Variable: wagestar

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 6 | 610.71764 | 101.78627 | 101.02 | <.0001 |
| Error | 404 | 407.07442 | 1.00761 | | |
| Uncorrected Total | 410 | 1017.79206 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 1.00380 | R-Square | 0.6000 |
| Dependent Mean | 1.14517 | Adj R-Sq | 0.5941 |
| Coeff Var | 87.65508 | | |

Parameter Estimates

Intercept

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| wtp | 1 | -10.27326 | 4.81967 | -2.13 | 0.0336 |
| edstar | 1 | 1.57443 | 0.22439 | 7.02 | <.0001 |
| expstar | 1 | 0.45561 | 0.20990 | 2.17 | 0.0305 |
| expsqstar | 1 | -0.00353 | 0.00334 | -1.06 | 0.2913 |
| festar | 1 | -4.79260 | 3.50195 | -1.37 | 0.1719 |
| expfestar | 1 | -0.06881 | 0.14105 | -0.49 | 0.6259 |

The GLM Procedure
Dependent Variable: wage
Weight: p2

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 81.2533689 | 16.2506738 | 16.13 | <.0001 |
| Error | 404 | 407.0744209 | 1.0076100 | | |
| Corrected Total | 409 | 488.3277898 | | | |

| R-Square | Coeff Var | Root MSE | wage Mean |
|---|---|---|---|
| 0.166391 | 7.320180 | 1.003798 | 13.71275 |

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Intercept | -10.27326243 | 4.81966589 | -2.13 | 0.0336 |
| YRSED | 1.57442994 | 0.22438623 | 7.02 | <.0001 |
| exp | 0.45561219 | 0.20989609 | 2.17 | 0.0305 |
| expsq | -0.00353339 | 0.00334406 | -1.06 | 0.2913 |
| fe | -4.79260329 | 3.50194991 | -1.37 | 0.1719 |
| expfe | -0.06881051 | 0.14105228 | -0.49 | 0.6259 |

results are the same "by hand"
or
by SAS