VI. Problems

A. **Multicollinearity**

B. **Heteroskedasticity**

C. **Autocorrelation**

For each "Problem" you should be able to answer the following questions:

- ***What's the problem? Define and Explain.***
- ***What are the consequences of the problem?***
- ***How do we diagnose the problem?***
- ***How can we fix it?***

VI. Problems

A. Multicollinearity

1. **Definition**: The presence of ***linear association*** among independent variables.
2. **Consequences:**
   - OLS estimators – remain unbiased.
   - Standard errors inflated ⇒ t_{calc} deflated.
   - Can't trust your hypothesis tests.
   - ***P(Type II Error) is high***.

3. Diagnosis (Multicollinearity)

How can we tell if we have this problem?

a. **Classic Signs:**

b. **Correlation Coefficients** – how strong are the pair-wise correlations between the Xs?

c. **Auxilliary Regressions** – regressions of one X on all the others.

d. **Variance Inflation Factors** (VIF). Same information as R^2 from the Auxilliary Regs.

4. Solutions – fixing the problem

- Sample data problem:
- ***Eliminate the offensive variable:***
- Linear Association – try a non-linear model.
- ***Use "non-sample" information:***
- ***Data transformations***:

U.S. Demand for Chicken

```
The regression equation is
lnqchik = 2.19 - 0.505 lnpchik + 0.149 lnppork
        + 0.091 lnpbeef + 0.343 lndinc

Predictor     Coef  SE Coef      T      P   VIF
Constant    2.1898   0.1557  14.06  0.000
lnpchik    -0.5046   0.1109  -4.55  0.000  17.5
lnppork    0.14855  0.09967   1.49  0.153  41.4
lnpbeef     0.0911   0.1007   0.90  0.378  42.3
lndinc     0.34255  0.08327   4.11  0.001  65.1

S = 0.0275911   R-Sq = 98.2%   R-Sq(adj) = 97.8%
```

$F_{calc} = 249.93$

Solution – use relative prices

```
The regression equation is
lnqchik = 2.38 - 0.612 lnrpchik + 0.157 lnrppork +
0.382 lnrdinc

Predictor      Coef    SE Coef       T       P    VIF
Constant     2.3831     0.1209   19.71   0.000
lnrpchik   -0.61246    0.09942   -6.16   0.000    9.2
lnrppork     0.1575     0.1055    1.49   0.152    3.8
lnrdinc     0.38228    0.08516    4.49   0.000    8.4

S = 0.02923     R-Sq = 97.9%     R-Sq(adj) = 97.6%

Analysis of Variance

Source          DF        SS         MS       F      P
Regression       3   0.75851    0.25284  295.85  0.000
Residual Error  19   0.01624    0.00085
Total           22   0.77475
```

B. Heteroskedasticity

1. **Definition**: One of our CRM assumptions is ***violated***. The disturbances do not have constant variances.

- Homoskedastic Disturbances (CRMA #4):

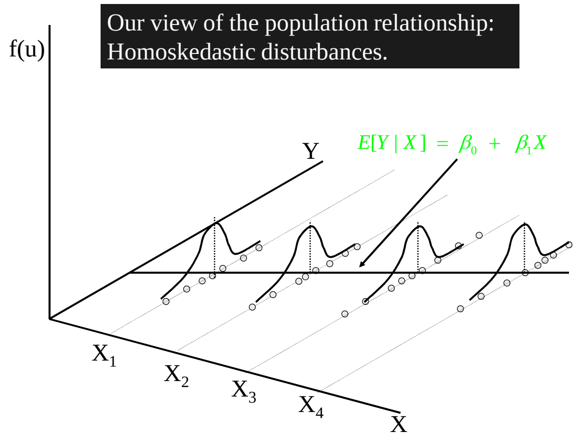
  $\mathrm{Var}(u_i) = E[u_i^2] = \sigma^2$

- Heteroskedastic Disturbances:

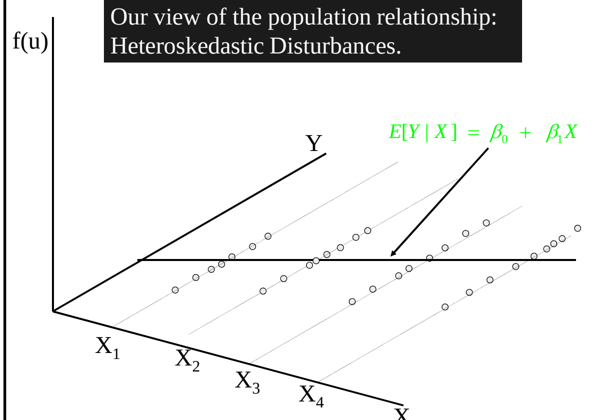
  $\mathrm{Var}(u_i) = E[u_i^2] = \sigma_i^2$

Example: $\mathrm{Var}(u_i) = \sigma^2 X_i$

Our view of the population relationship:
Homoskedastic disturbances.



Our view of the population relationship:
Heteroskedastic Disturbances.



2. Consequences

- CRMA#4 is violated. When was CRMA#4 used?

$$Var(\hat{\beta}_1) = \left[\sum_i k_i^2 E[u_i^2] + 2\sum_{i \neq j} k_i k_j E[u_i u_j] \right]$$

If: $E[u_i^2] = \sigma^2$	If: $E[u_i^2] = \sigma_i^2$
$s_{\hat{\beta}_1} = \sqrt{\dfrac{\sigma^2}{\sum x_i^2}}$	$s_{\hat{\beta}_1} = \sqrt{\dfrac{\sum x_i \sigma_i^2}{\sum x_i^2}}$

3. Diagnosis:

How can we tell if we have this problem?

i. ***Nothing on your printout tells you***. You gotta look.

ii. Plot: Errors (or e_i^2) vs. Xs, or Y.

iii. Regressions: $\mathbf{e_i^2}$ regressed on Xs and X^2s

- **Look for high R^2**
- **Statistical Test: $n \times R^2 \sim \chi^2_{(K)}$ (White's test)**

4. Solutions:

- ***Generalized Least Squares***: These are the BLUE estimators. Must know σ_i^2.
- "***Feasible Generalized Least Squares***"
  - Weight the dependent and independent variables.
  - Estimate using the weighted variables.
  - Example: $E[u_i^2] = X_i \sigma^2 = \sigma_i^2$

    Transform data:

    Estimate:

B. Autocorrelation

1. **Definition**: Another of our CRM assumptions is ***violated***. Disturbances related across observations. Usually time-series data.

- Non-autocorrelated Disturbances (CRMA #5):
- Autocorrelated Disturbances:
- Examples: First-order; Second-order.

2. Consequences

- CRMA#5 is violated. When was CRMA#5 used?

$$Var(\hat{\beta}_1) = \left[\sum_i k_i^2 E[u_i^2] + 2\sum_{i \neq j} k_i k_j E[u_i u_j] \right]$$

If: $E[u_i u_j] = 0$	If: $E[u_i u_j] \neq 0$
$s_{\hat{\beta}_1} = \sqrt{\dfrac{\hat{\sigma}^2}{\sum x_i^2}}$	$s_{\hat{\beta}_1} = \sqrt{\dfrac{\hat{\sigma}^2}{\sum x_i^2}} + \text{"}stuff\text{"}$

3. Diagnosis:

How can we tell if we have this problem?

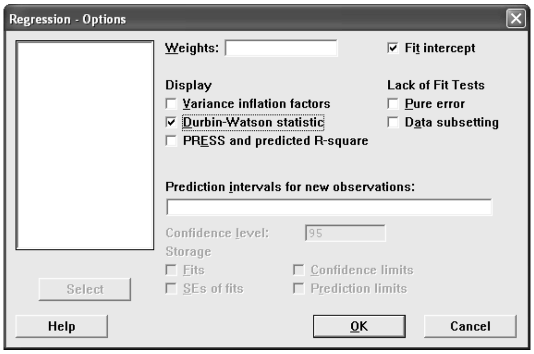
i. ***Durbin Watson Test Statistic***:

Positive Autocorrelation:

No Autocorrelation:

Negative Autocorrelation:

Durbin-Watson statistic – Minitab.




```
The regression equation is
qchik = 37.2 - 0.611 pchik + 0.198 ppork + 0.0695 pbeef +
0.00501 dinc

Predictor     Coef    SE Coef       T       P     VIF
Constant    37.232     3.718    10.01   0.000
pchik      -0.6112    0.1628    -3.75   0.001    18.9
ppork      0.19841   0.06372     3.11   0.006    29.1
pbeef      0.06950   0.05099     1.36   0.190    39.8
dinc      0.005011  0.004893     1.02   0.319    52.7

S = 1.953       R-Sq = 94.3%     R-Sq(adj) = 93.0%

Analysis of Variance
Source          DF        SS        MS       F        P
Regression       4   1127.26    281.81   73.87    0.000
Residual Error  18     68.67      3.81
Total           22   1195.93

Durbin-Watson statistic = 1.07
```

3. Diagnosis:

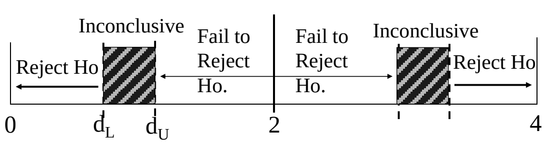
iii. Durbin Watson Test:

Durbin-Watson Table:

n – number of observations

k′ = number of explanatory variables.

Picture of the Test: $\mathbf{H_o}$: $\rho = 0$



4. Solutions:

- ***Generalized Least Squares***: BLUE – need ρ.
- "***Feasible Generalized Least Squares***"
  - Weight the dependent and independent variables.
  - Estimate using the weighted variables.
  - Example:
  - Estimate, transform data:

Estimate again using: