UCI Paul Merage School of Business

Leadership for a Digitally Driven World™

MFIN 290: Financial Econometrics

Lecture 3-1

Last Time

Violations of GM Assumptions:

$$E(X'e) \neq 0$$

Endogeneity

Omitted Variables

Measurement Error

Selection Bias

Last Time

Violations of GM Assumptions:

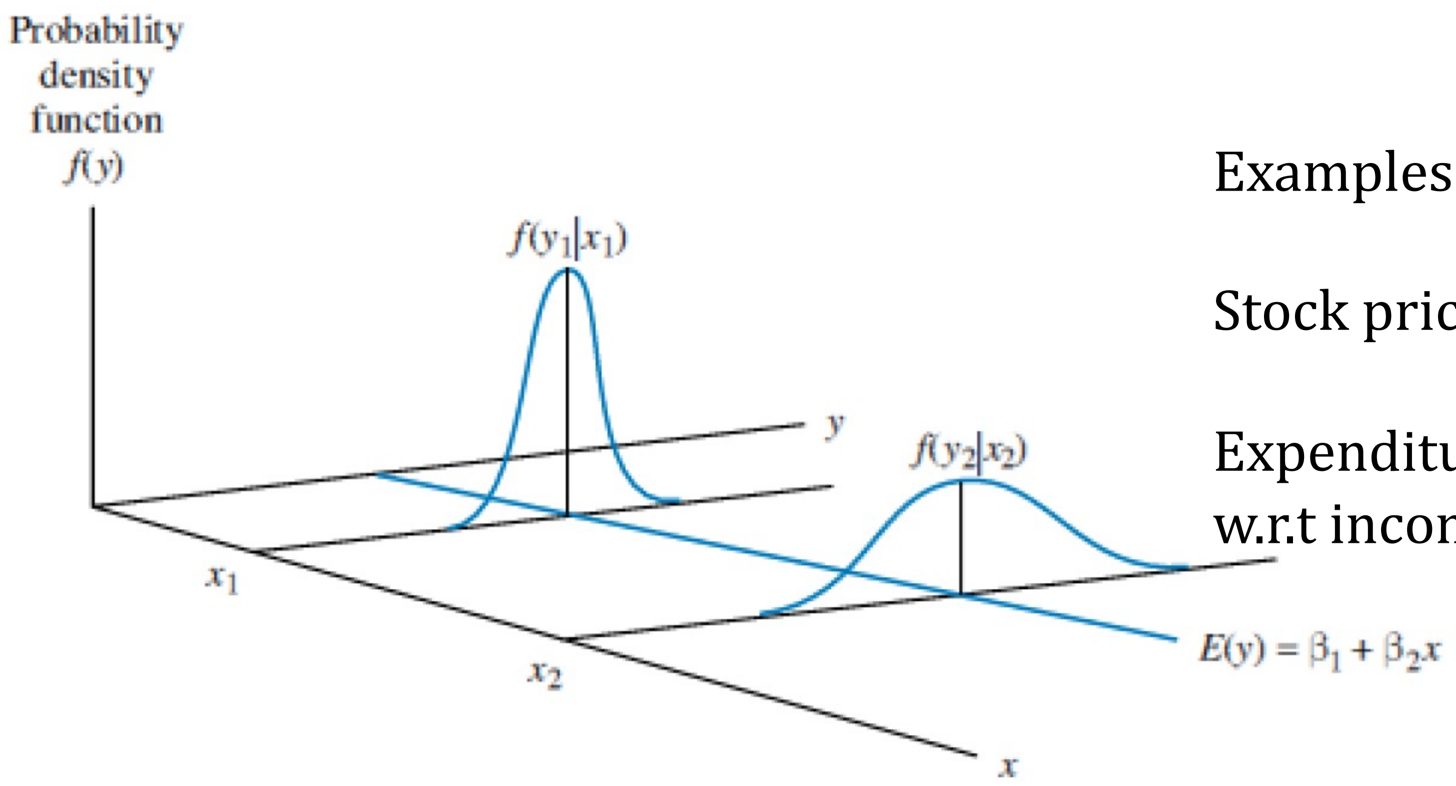
$$E(X'e) \neq 0$$

Violation here gives you bias. Your coefficient is not showing what you want it to show you.

Can happen for intuitive reasons, measurement or interpretation issues, can be fixed with instruments…

Today’s issue is also a violation, but does not give bias.

Heteroskedasticity



Examples:

Stock price returns

Expenditure on food w.r.t income

Heteroskedasticity and LS Assumptions

When there is heteroskedasticity, one of the least squares assumptions is violated. We still have that
$E(\epsilon|X) = 0$

Which was the key assumption for unbiasedness

But now, the assumption that $var(\epsilon) = \sigma^2 \mathrm{I}$, is replaced with $var(\epsilon) = \mathbf{\Omega}$ some generic covariance matrix. We used this when deriving the variance of our OLS estimator.

For example, $var(\epsilon_i) = h(x_i)$ where $h(x_i)$ is some function of x_i

Where to find heteroskedasticity

- Heteroskedasticity is often encountered when using cross-sectional data, but also in other types of data (like economic or financial time series!)

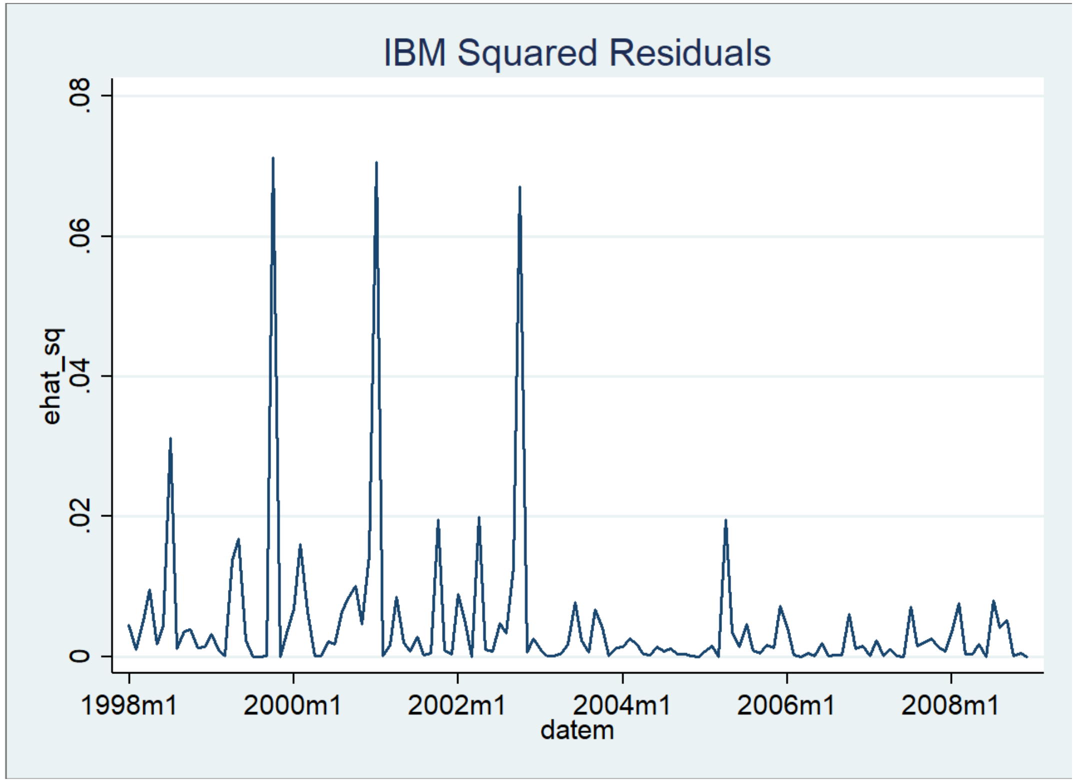
- Cross-sectional data invariably involve observations on economic units of varying sizes. This means that as the size of the economic unit becomes larger, there is more uncertainty associated with the outcomes.

- This greater uncertainty is modeled by specifying an error variance that is larger, the larger the size of the economic unit (variance may increase with income, for example).
- Always a good idea to think about this when you care about inference.

CAPM Example

```
. gen ibm_excess = ibm-riskfree

. gen mkt_excess = mkt - riskfree

. reg ibm_excess - mkt_excess

      Source |       SS           df       MS      Number of obs   =       132
-------------+----------------------------------   F(1, 130)       =     88.32
       Model |  .432585943         1  .432585943   Prob > F        =    0.0000
    Residual |  .636721741       130  .00489786   R-squared       =    0.4045
-------------+----------------------------------   Adj R-squared   =    0.4000
       Total |  1.06930768       131  .008162654   Root MSE        =    .06998

------------------------------------------------------------------------------
  ibm_excess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  mkt_excess |   1.188208   .1264327     9.40   0.000     .9380763     1.43834
       _cons |   .0058513   .0060914     0.96   0.339    -.0061999     .0179024
------------------------------------------------------------------------------
```

CAPM Example




```
*get predicted residuals & square
predict ehat,resid
gen ehat_sq = ehat*ehat

*format dataset as time series
gen year = floor(date/10000)
gen month = floor((date-year*10000)/100)
gen datem = ym(year, month)
tsset datem, month

*draw graph
tsline ehat_sq, title("IBM Squared
Residuals")
```

Heteroskedasticity and LS Assumptions

$$var(\hat{\beta} - \beta) = E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\middle|X\right)$$

$$= E\left((X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}|X\right)$$

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

$$\neq \sigma^2 I(X'X)^{-1}$$

Heteroskedasticity and LS Assumptions

There are two implications of heteroskedasticity:

1. The least squares estimator is still a linear and unbiased estimator, but it is no longer the "best" estimator. In fact, there is another linear unbiased estimator with a smaller variance

2. The typical standard errors computed for the least squares estimator are incorrect. Confidence intervals and hypothesis tests that use these standard errors may be misleading

Generalized Least Squares

1. GLS

Want to find a different estimator with lower variance than OLS when we have $var(\epsilon) = \mathbf{\Omega}$

Generalized Least Squares

1. GLS

Want to find a different estimator with lower variance than OLS when we have $var(\epsilon) = \mathbf{\Omega}$

Idea is that if we know the form of the heteroskedasticity, we can transform X and Y into a homoscedastic model and estimate that… OLS will be minimum variance on the transformed data!

GLS

To motivate the GLS estimator we impose some structure on σ_i^2.

For example:

$$Var(\epsilon_i) = \sigma_i^2 = \sigma^2 x_i$$

GLS

To make the GLS estimator we impose some structure on σ_i^2.

For example:

$$Var(\epsilon_i) = \sigma_i^2 = \sigma^2 x_i$$

$$Var(\epsilon_i) = \sigma_i^2 = \sigma^2 x_i \Rightarrow Var\left(\frac{\epsilon_i}{\sqrt{x_i}}\right) = \sigma^2$$

GLS

So instead of regressing

$$y = x\beta + \epsilon$$

We transform

$y^* = \frac{y}{\sqrt{x}}$;

$x^* = \frac{x}{\sqrt{x}}$ (we transform all of the columns.. Including the constant term)

And regress

$$y^* = x^*\beta + \epsilon^*$$

GLS

One way of viewing the generalized least squares estimator is as a *weighted* least squares estimator

Minimizing the sum of squared transformed errors:

$$\sum_{i=1}^{N} e_i^{*2} = \sum_{i=1}^{N} \frac{e_i^2}{x_i} = \sum_{i=1}^{N} \left(x_i^{-1/2} e_i\right)^2$$

That is, the errors are weighted by $\frac{1}{\sqrt{x_i}}$… the inverse square root of our volatility function!

Multivariate GLS

$$Y = X\beta + \epsilon;\ var(\epsilon) = \mathbf{\Omega}$$

Since $\mathbf{\Omega}$ is positive definite, it has a matrix square root:

$$\Rightarrow\ \mathbf{\Omega} = PP'$$

Multivariate GLS

$$\Rightarrow\ \mathbf{\Omega} = PP' = \mathbf{\Omega}' = P'P$$

Premultiply $y = x\beta + \epsilon$ by P^{-1}:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}\epsilon$$

$$Y^* = X^*\beta + \epsilon^*$$

$$var(P^{-1}\epsilon) = E\left((P^{-1}\epsilon)(P^{-1}\epsilon)'\right) = P^{-1}\mathbf{\Omega}P^{-1'}$$

Multivariate GLS

$$var(P^{-1}\epsilon) = E\left((P^{-1}\epsilon)(P^{-1}\epsilon)'\right) = P^{-1}\mathbf{\Omega}P^{-1'}$$

$$var(P^{-1}\epsilon) \ = P^{-1}\mathbf{\Omega}P^{-1\prime}$$

And we know $\mathbf{\Omega} = PP'$

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

$= \sigma^2 I$ with $\sigma = 1$

$\Rightarrow$ Gauss Markov Assumptions hold with OLS of Y^*on X^*

$\Rightarrow$ GLS is BLUE

Multivariate GLS

If we estimated this by OLS, we would be unbiased $(E(P^{-1}\epsilon|X) = 0)$ with

$$var(\hat{\beta}^*) = \left((P^{-1}X)'P^{-1}X\right)^{-1}(P^{-1}X)'P^{-1}\epsilon(P^{-1}\epsilon)'P^{-1}X\left((P^{-1}X)'P^{-1}X\right)^{-1})$$

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

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

Multivariate GLS

- If we transform/“rotate” the data by premultiplying by P^{-1} = matrix square root of Ω , then we are all set and back to our efficient estimator with the usual inference if we estimate via OLS on the transformed data!

- Note that this estimator includes regular OLS as a special case when $P = I$; this shares the same BLUE properties.

Multivariate GLS

- What does $\hat{\beta}_{GLS}$ look like?
- $P^{-1\prime}Y = P^{-1\prime}X\beta + P^{-1\prime}\varepsilon$
- $\hat{\beta}_{GLS} = (X'P^{-1}P^{-1\prime}X)^{-1}X'P^{-1}P^{-1\prime}Y$
- $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$

GLS and the Hausman Test

- Since OLS and GLS are both unbiased, they should give us the same coefficients (in a statistical sense). Under the null hypothesis that there is heteroskedasticity AND GLS corrects it properly, GLS is an efficient estimator, OLS is in efficient, and they have the same limit.

- Can run a Hausman test!

- Will identify when the GLS "fix" isn't the right fix in the sense that the coefficients are different (which means we have introduced bias).
- Remember: GLS shouldn't change the coefficients: have seen this problem!

"Feasible" GLS

- If we know $\mathbf{\Omega}$, then we can transform the data, run GLS, and recover a minimum variance estimator.

- But it is very unlikely that we know $\mathbf{\Omega}$… we don't know β after all.

- So we have to estimate $\mathbf{\Omega}$ with $\widehat{\mathbf{\Omega}}$

FGLS

- If we have a consistent estimator of $\mathbf{\Omega}$, $\widehat{\mathbf{\Omega}}$ then we asymptotically can transform the data as needed for GLS.

- Note that we don't need an *efficient* estimator of $\mathbf{\Omega}$, just a consistent one.

- So how can we estimate $\widehat{\mathbf{\Omega}}$?

FGLS

- So how can we estimate $\widehat{\boldsymbol{\Omega}}$?

- Can assume a form and estimate the parameters directly (through ML or other methods)

- Or use a two step procedure: remember OLS is still unbiased (and consistent), just inefficient. This means that our estimated errors, $e = y - x\hat{\beta}$ are unbiased and consistent.

FGLS

- 1) Estimate OLS
- 2) Recover estimated coefficients, predicted error terms
- 3) Estimate $\widehat{\boldsymbol{\Omega}}$ given the error term. We still need some structure here as there are too many unknowns ($\frac{(n-1)^2}{2} + n$) in the variance covariance matrix of the epsilons…
- Note: this will be consistent, not unbiased. We estimate $\widehat{\boldsymbol{\Omega}}$, but need $\widehat{\boldsymbol{\Omega}}^{-1}$. Expectations are only closed under linear operators.

FGLS Example: CAPM

```
. gen ibm_excess = ibm-riskfree

. gen mkt_excess = mkt - riskfree

. reg ibm_excess   mkt_excess

      Source |       SS           df       MS      Number of obs   =       132
-------------+----------------------------------   F(1, 130)       =     88.32
       Model |  .432585943         1  .432585943   Prob > F        =    0.0000
    Residual |  .636721741       130  .00489786    R-squared       =    0.4045
-------------+----------------------------------   Adj R-squared   =    0.4000
       Total |  1.06930768       131  .008162654   Root MSE        =    .06998

------------------------------------------------------------------------------
  ibm_excess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  mkt_excess |   1.188208   .1264327     9.40   0.000     .9380763     1.43834
       _cons |   .0058513   .0060914     0.96   0.339    -.0061999    .0179024
------------------------------------------------------------------------------
```

FGLS Example: CAPM

Let's assume that $\mathbf{\Omega}$ has off diagonal values all equal to zero (why might this be reasonable?)

FGLS Example: CAPM

Let's assume that $\boldsymbol{\Omega}$ has off diagonal values all equal to zero

(why might this be reasonable?)

We proceed in two steps:

1. Use e_i^2 as our estimate of the squared predicted residuals on each date:
2. Divide each observation by the std dev at that point.
   - Don't forget to replace the constant term/vector of ones with a new one and suppress the usual constant

FGLS Example: CAPM

```
gen mkt_excess_gls = mkt_excess/sqrt(ehat_sq)
gen ibm_excess_gls = ibm_excess/sqrt(ehat_sq)
gen const_gls = 1/sqrt(ehat_sq)
reg ibm_excess_gls mkt_excess_gls const_gls, noc
```

```
. reg ibm_excess_gls mkt_excess_gls const_gls,noc

      Source |       SS           df       MS      Number of obs   =       132
-------------+----------------------------------   F(2, 130)       >  99999.00
       Model |  537846006          2  268923003   Prob > F        =    0.0000
    Residual | 130.229031        130 1.00176178   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total |  537846137        132 4074591.94   Root MSE        =    1.0009

------------------------------------------------------------------------------
ibm_excess_gls |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
mkt_excess_gls |   1.186588   .0021162   560.72   0.000     1.182401    1.190774
     const_gls |   .0058874   .0000453   129.96   0.000     .0057978     .005977
------------------------------------------------------------------------------
```

Note the coefficient is basically identical, but look at all the extra power! Unneeded here, sure…

Option 2: Correcting Inference

- If we don't care about efficient (or maybe we don't know how to correct the heteroskedasticity as we have failed a Hausman test), the other option would be to correct the standard errors.

- We know the variance of $\hat{\beta}$ is given by: $(X'X)^{-1}X'\mathbf{\Omega}X(X'X)^{-1}$

- So we need an estimate of $X'\mathbf{\Omega}X$… Note that this is k x k, a smaller matrix than $\mathbf{\Omega}$

Let's continue with the assumption that the off-diagonals are zero for now

Option 2: Correcting Inference

- $X'\Omega X = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega^{i,j}\, x_i x_j'$ for the $i, j\, th$ elements
- $X'\Omega X = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega^{i,j}\, x_i x_j' = \sum_{i=1}^{n} \omega^{i,i^2} x_i x_i'$ since $\omega^{i,j} = 0\ \forall\ i \neq j$
- $\sum_{i=1}^{n} \omega^{i,i^2} x_i x_i' = \sum_{i=1}^{n} \varepsilon_i{}^2 x_i x_i'$
-

White Robust Standard Errors

- $Var[\hat{\beta}|X] = (X'X)^{-1}X'\Omega X(X'X)^{-1}$
- If the off-diagonals are zero, we can estimate this with:

- $\widehat{X'\Omega X} = \sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2} x_i x_i'$

- We are just missing a division by n to get familiar terms in each block:
- $(X'X)^{-1}X'\Omega X(X'X)^{-1} = \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}[\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2} x_i x_i'] \left(\frac{X'X}{n}\right)^{-1}$

White Robust Standard Errors

$$\frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^{\,2}x_ix_i'\right]\left(\frac{X'X}{n}\right)^{-1}$$

- This is the White estimator for standard errors (often called "White Robust"). Allows for the diagonal elements of Ω to not all be identical. We still restrict the off-diagonals to be zero here, but this is sometimes reasonable, especially in finance (but not always!)
- Note that there is no guarantee that White standard errors will be "larger" than spherical standard errors… but this is almost always the case.
-

White Robust Standard Errors

- Will have numerically identical coefficients, just different standard errors/different inference.
- The estimate of the coefficient is still inefficient, but the estimator’s standard error is consistent (if the heteroskedasticity is as assumed!).

```
. reg ibm_excess mkt_excess

      Source |       SS           df       MS      Number of obs   =       132
-------------+----------------------------------   F(1, 130)       =     88.32
       Model |  .432585943         1  .432585943   Prob > F        =    0.0000
    Residual |  .636721741       130  .00489786    R-squared       =    0.4045
-------------+----------------------------------   Adj R-squared   =    0.4000
       Total |  1.06930768       131  .008162654   Root MSE        =    .06998

------------------------------------------------------------------------------
  ibm_excess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  mkt_excess |   1.188208   .1264327     9.40   0.000     .9380763     1.43834
       _cons |   .0058513   .0060914     0.96   0.339    -.0061999    .0179024
------------------------------------------------------------------------------

. reg ibm_excess mkt_excess, robust

Linear regression                               Number of obs     =        132
                                                F(1, 130)         =      83.71
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4045
                                                Root MSE          =     .06998

------------------------------------------------------------------------------
             |               Robust
  ibm_excess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  mkt_excess |   1.188208   .1298694     9.15   0.000     .9312772     1.44514
       _cons |   .0058513   .0060981     0.96   0.339    -.0062131    .0179156
------------------------------------------------------------------------------
```

Testing for Heteroskedasticity

- Suppose that we have the model
- $y = x\beta + \epsilon$
- Which we estimate through OLS, and recover the fitted residuals:
- $e = y - x\hat{\beta}$

- There are many possible forms of heteroskedasticity, but one might be that the variance in the residuals depends on X…

Testing for Heteroskedasticity

- We can estimate this!

Square the fitted residuals, and regress these on the x's

$$e_i^2 = x_i\gamma + u_i$$

Null Hypothesis = no heteroskedasticity => no relationship => γ are all zero

How do we test if any of the regressors are significant in a regression?

Testing for Heteroskedasticity

- We can estimate this!

Square the fitted residuals, and regress these on the x's

$$e_i^2 = x_i\gamma + u_i$$

nR^2 is distributed as a chi squared with P-1 degrees of freedom

This is called the Breusch Pagan test.

It has a Chi-Squared distribution with p-1 degrees of freedom (where p is the dimension of γ). It is also called the Lagrange multiplier test.

Testing for Heteroskedasticity

- Unsurprisingly, we reject the null of constant variance on the untransformed CAPM data.

```
. qui reg ibm_excess mkt_excess

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ibm_excess

         chi2(1)      =     6.84
         Prob > chi2  =   0.0089
```

- Obvi eteroskedasticity (b/c we never "accept the r
- Ther to you to be able to understand how the tests ble based on the test result.

Non-Diagonal Heteroskedasticity

$$var(\epsilon) = \mathbf{\Omega} = E(\epsilon\epsilon') = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_T \end{bmatrix} [\epsilon_1 \quad \epsilon_2 \quad \cdots \quad \epsilon_T]$$

$$= \begin{bmatrix} \epsilon_1^2 & \cdots & \epsilon_T\epsilon_1 \\ \vdots & \ddots & \vdots \\ \epsilon_1\epsilon_T & \cdots & \epsilon_T^2 \end{bmatrix}$$

Off diagonal elements in $\mathbf{\Omega}$ imply some sort of dependence between the errors of different observations (not just that the variance changes by observation). Can have common shocks for groups – called "clustered" errors, or with a time series or panel dataset, this implies there is some relationship between different values over time.

Non-Diagonal Heteroskedasticity

We usually think of financial time series as uncorrelated but there can be exceptions

Who can think of example of assets where say, year over year returns are correlated (hint: not stocks, but…)

What about longer horizons (CAPE?)?

What about private or illiquid assets?

Newey West Standard Errors

If we know the form of the off diagonals (say they decay exponentially?), we can estimate the resulting parameters and get consistent estimates of Ω, but if we don't, we have to do something else…

If we know that the autocorrelations beyond a certain lag are "small" (either they are zero or go to zero asymptotically), we reduce the number of parameters we have to estimate to something more tractable.

This is what Newey West Standard Errors do.

Newey West Standard Errors

$$\begin{bmatrix} \epsilon_1^2 & \cdots & \epsilon_T\epsilon_1 \\ \vdots & \ddots & \vdots \\ \epsilon_1\epsilon_T & \cdots & \epsilon_T^2 \end{bmatrix}$$

Within some band, allow the numbers to be non-zero. Beyond the band, do not.

Is this reasonable?

Trade off: flexibility vs. parameters.

People default to $T^{\frac{1}{4}}$ as the band, but that's driven by simulation (footnote in the original paper) .

Newey West Standard Errors

Begin with the White Estimator:

$$\frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^{\,2}x_ix_i'\right]\left(\frac{X'X}{n}\right)^{-1}$$

Call the estimate of the diagonal terms $S_0 = [\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^{\,2}x_ix_i']$ and add to this estimates of the off diagonals weighted by the corresponding X's

$$\widehat{\boldsymbol{\Omega}} = S_0 + \frac{1}{n}\sum_{l=1}^{L}\sum_{t=l+1}^{n} w_l e_t e_{t-l}\,(x_t x_{t-l}' + x_{t-l}x_t')$$

With weights $w_l = 1 - \frac{l}{L+1}$

Newey West Standard Errors

Newey West estimates the off-diagonal terms in a band of a certain length, beyond which they become zero. The idea is that if the width of this band increases with the number of observations (but slowly enough so that the terms can all be identified), we will recover a flexible estimate of any $\boldsymbol{\Omega}$ *asymptotically*.

This is not a magic bullet – but is often treated like one. First, we are assuming that the autocorrelations go away over time. Second, that we can precisely estimate a lot of parameters in $X'\Omega X$ in our finite sample.

Benefits from NW errors rely heavily on asymptotics – *the lags have to go to infinity as well, just slower than the observation counts.*

Newey West Standard Errors

```
. newey ibm_excess mkt_excess, lag(3)

Regression with Newey-West standard errors          Number of obs     =        132
maximum lag: 3                                      F(  1,        130) =      67.58
                                                    Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Newey-West
  ibm_excess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  mkt_excess |   1.188208   .1445344     8.22   0.000     .9022643    1.474152
       _cons |   .0058513   .0060736     0.96   0.337    -.0061646    .0178672
------------------------------------------------------------------------------
```

Example: CAPM and Heteroskedasticity

- PS2 example.csv, .xls, .do