Greene 2000, Ch. 10 — Robust Estimation of $Cov(\hat{\beta})$

A. White's heteroskedasticity consistent estimator
B. Newey & West's autocorrelation consistent estimator
of $\widehat{Cov}(\hat{\beta})$

The Problem: $\epsilon \sim N(0, \Omega)$

where Ω is a symmetric matrix $(n \times n)$
with unequal elements above/below diagonal.

1. Fact: If you knew Ω, GLS is best.

But you don't know Ω.
Can assume $\Omega = f(\theta)$

$$\boxed{\begin{array}{lll} \text{e.g. Cone; } & Var(\epsilon_i) = x_{kt}\,\sigma^2 & \text{for heterosked.} \\ \text{or } AR(1); & \epsilon_t = \rho\,\epsilon_{t-1} + u_t & \text{for autocorrel.} \end{array}}$$

and then do Generalized Least Squares.

This is an attempt to estimate Ω.

2. Fact: If your estimate, $\hat{\Omega}$, is consistent for Ω
then GLS has desireable properties.

But how can you know if your $\hat{\Omega}$
is truly consistent, or good?

3. Recall: OLS $\hat{\beta} \sim N(\beta, (X'X)^{-1}(X'\Omega X)(X'X)^{-1})$

OLS $\hat{\beta}$ is unbiased & consistent,

but OLS covariance matrix, $\sigma^2(X'X)^{-1}$, bad,
so that t & F statistics are bad.

White's approach: Use OLS $\hat{\beta}$,
and estimate $Cov(\hat{\beta}) \Rightarrow (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$

If Ω is unknown,

(1★) → but its structure is known

(2★) → and we can estimate $\hat{\Omega}$ from sample information,

the choice of using $\hat{\Omega}$ in GLS, versus OLS, is not clearcut.

In many cases, $\hat{\Omega}$ in GLS will be better than OLS.

But, What if we don't know about (1★) and (2★) ?
— what if Ω is completely unknown;
— both its ~~structure and its~~ elements?

Go back to OLS $\hat{\beta}$!

If you knew Ω, then you'd ESTIMATE $Cov(\hat{\beta})$ with

$$\left[\widehat{Cov}(\hat{\beta})\right] = \frac{1}{n}\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\Omega X\right)\left(\frac{1}{n}X'X\right)^{-1}$$

★ Consider the middle product, $\boxed{\left(\frac{1}{n}X'\Omega X\right)}$. ← Ω is $n \times n$; $(X'\Omega X)$ is $K \times K$

It might seem that you need to estimate all $n(n+1)/2$ elements in Ω. (impossible with n obs!).

Not true! You only need to estimate $K(K+1)/2$ elements in the matrix $\left(\frac{1}{n}X'\Omega X\right)$, or:

$$\text{plim}\, Q^* = \text{plim}\,\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\, x_i x_j' \; ; \; i,j=1,\dots,K$$

<u>Point:</u> Q^* contains sums of squares & cross-products that involve σ_{ij}, $i=1,\dots,K$ and $j=1,\dots,K$, and the rows of X.

INSERT (A)

Consider $\frac{1}{n}(X'\Omega X)$

$$(K\times n)\,(n\times n)\,(n\times K)$$

$$K\times K$$

$$= \underset{K\times n}{\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ & \vdots & & \\ X_{K1} & X_{K2} & \cdots & X_{Kn} \end{bmatrix}} \underset{n\times n}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ & \vdots & & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}} \underset{n\times K}{\begin{bmatrix} X_{11} & X_{21} & \cdots & X_{K1} \\ X_{12} & X_{22} & \cdots & X_{K2} \\ & \vdots & & \\ X_{1n} & X_{2n} & \cdots & X_{Kn} \end{bmatrix}}$$

$$= \underset{K\times n}{\begin{bmatrix} \sum_{j=1}^{n} X_{1j}\sigma_{j1} & \sum_{j=1}^{n} X_{1j}\sigma_{j2} & \cdots & \sum_{j=1}^{n} X_{1j}\sigma_{jn} \\ \sum_{j=1}^{n} X_{2j}\sigma_{j1} & \sum_{j=1}^{n} X_{2j}\sigma_{j2} & \cdots & \sum_{j=1}^{n} X_{2j}\sigma_{jn} \\ \vdots & & & \\ \sum_{j=1}^{n} X_{Kj}\sigma_{j1} & \sum_{j=1}^{n} X_{Kj}\sigma_{j2} & \cdots & \sum_{j=1}^{n} X_{Kj}\sigma_{jn} \end{bmatrix}} \underset{n\times K}{\begin{bmatrix} X_{11} & X_{21} & \cdots & X_{K1} \\ X_{12} & X_{22} & \cdots & X_{K1} \\ \vdots & & & \\ X_{1n} & X_{2n} & \cdots & X_{Kn} \end{bmatrix}}$$

$$= \begin{bmatrix} \left(X_{11}\sum X_{1j}\sigma_{j1} + X_{12}\sum X_{1j}\sigma_{j2} + \cdots + X_{1n}\sum X_{1j}\sigma_{jn}\right) & \cdots \\ \left(X_{11}\sum X_{2j}\sigma_{j1} + X_{12}\sum X_{2j}\sigma_{j2} + \cdots + X_{1n}\sum X_{2j}\sigma_{jn}\right) & \cdots \\ \vdots & \\ \left(X_{11}\sum X_{Kj}\sigma_{j1} + X_{12}\sum X_{Kj}\sigma_{j2} + \cdots + X_{1n}\sum X_{Kj}\sigma_{jn}\right) & \cdots \end{bmatrix}$$

1^{st} col.

(K columns like this)

Easier way to write this,
in terms of the outer products of columns of X.

Outer product of i^{th} and j^{th} columns of X:

$$\underset{(n\times n)}{\underset{(n\times 1)\ (1\times n)}{\underset{(i^{th}\text{ col.})(j^{th}\text{ col.})'}{(X_i\, X_j')}}} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{bmatrix} \begin{bmatrix} x_{j1} & x_{j2} & \cdots & x_{jn} \end{bmatrix}$$

$$= \underset{(n\times n)}{\begin{bmatrix} x_{i1}\,x_{j1} & x_{i1}\,x_{j2} & \cdots & x_{i1}\,x_{jn} \\ x_{i2}\,x_{j1} & x_{i2}\,x_{j2} & \cdots & x_{i2}\,x_{jn} \\ \vdots & & & \\ x_{in}\,x_{j1} & x_{i2}\,x_{j2} & \cdots & x_{in}\,x_{jn} \end{bmatrix}}$$

The $(i,j)^{th}$ element of $(X'\Omega X)$
~~can be written as a double sum:~~

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}\left[(i,j)^{th}\text{ element of } X_i X_j'\right].$$

To see this, consider the first column of $X'\Omega X$ $(i=1)$
as written on the previous page bottom.

How to estimate $Q^* \Rightarrow \frac{1}{n}(X'\Omega X) \Rightarrow Cov(\hat{\beta})$

OLS $\hat{\beta}$ is consistent for β;

$\therefore$ Residuals from OLS are "pointwise consistent" estimators of ϵ.

General approach: Use X and e to estimate Q^*.

A. Heteroskedasticity Case. $Cov(\epsilon) = \Omega \leftarrow$ (diagonal matrix)

$$Q^* = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 x_i x_i'$$

[Note: Ω is diagonal; $\therefore$ only n terms to sum.

White (1980) shows that under general conditions,

$$S_0 = \frac{1}{n} \sum_{i=1}^{n} e_i^2 x_i x_i' \text{ is consistent for } Q^*.$$

IMPORTANT POINT: Q^* is not a parameter matrix, really. Q^* is a weighted sum of the outer products of the columns of X.
$\therefore$ We don't really need to "estimate" Q^*; we need to find a function of the sample that will come arbitrarily close to this function of the population parameters as $n \to \infty$.

We are not really "estimating" $(X'\Omega X)$; we want to construct a matrix from the sample that will behave the same way as this behaves $\to Q^*$.

[If we knew the true ϵ, we'd use $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \epsilon_i^2 x_i x_j'$.]
$\longrightarrow$ Replace ϵ_i with e_i !

<u>White's approach:</u> use OLS $\hat{\beta}$,

and estimate $\operatorname{Cov}(\hat{\beta}) = (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$

with $\left[\widehat{\operatorname{Cov}(\hat{\beta})}\right] = \frac{1}{n}\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i'\right)\left(\frac{1}{n}X'X\right)^{-1}$

This may be more desire-able than GLS;
transforming your whole model
with some assumption about $\Omega = f(\theta)$,
which would only improve on OLS
if assumptions were true.

OLS $\hat{\beta}$ appealling!

Interpretation is clear!

Interpretation of GLS is <u>different</u>;

→ impact of (VX) on (VY)

→ In order to fix heteroskedastic errors,
you changed your model!

(OK if transformation fixes
true problem with error.)

<u>BUT</u> Interpretation may not be clear!

<u>Note:</u> White's heteroskedasticity consistent est. uses OLS $\hat{\beta}$.
Residuals based on ANY consistent estimator of β
will work in place of OLS e.

OLS residuals are the obvious choice.

White's result is extremely important!

— w/o actually specifying the type of heterosked., can still make appropriate inferences from OLS $\hat{\beta}$.

Especially useful if we are unsure of precise nature of the heteroskedasticity. (most of time!)

Practical result: Be careful using OLS std. errors!

Under general conditions, the elements in $\sigma^2(X'X)^{-1}$ are too small relative to $(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$

Thus t-ratios are too large, using OLS $Cov(\hat{\beta})$.

White's $\widehat{\left[Cov(\hat{\beta})\right]}$ results in more accurate (bigger) standard errors, and thus, appropriate tests.

B. Autocorrelation Case $\quad Cov(\epsilon) = \Omega$ (not diagonal)

Again, $Cov(\hat{\beta}) = (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$

But now Ω is not diagonal; $(X'\Omega X)$ is messier.

With heteroskedasticity,

$$Q^* = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 x_i x_j' \qquad (n \text{ terms to } \Sigma)$$

& White suggested:

$$S_0 = \frac{1}{n} \sum_{i=1}^{n} e_i^2 x_i x_j'$$

With autocorrelation,

$$Q^* = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j' \qquad (n^2 \text{ terms to } \Sigma)$$

The natural counterpart to White's S_0 would be:

$$\boxed{\hat{Q}^*} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} e_i e_j x_i x_j'$$

But $\exists$ two problems with this estimator.

One is theoretical, and applies to Q^* as well;
One is practical, and is specific to $\hat{Q}^*$ only.

Autocorrelation

1. Theoretical Problem with Q^* and $\boxed{\hat{Q}^*}$.

With heteroskedasticity, $Q^* = \frac{1}{n}(\Sigma\ n \text{ terms})$;
easy to conclude this will converge as $n \to \infty$.

With autocorrelation, $Q^* = \frac{1}{n}(\Sigma\ n^2 \text{ terms})$;

NOT easy to conclude it will converge!

Consider assumption of first order autocorrelation;

$$\epsilon_t = \rho\epsilon_{t-1} + u_t$$

$$Cov(\epsilon) = \Omega = \begin{bmatrix} \rho & \rho^2 & \rho^3 & \cdots \\ \rho^2 & \rho & \rho^2 & \cdots \\ \rho^3 & \rho^2 & \rho & \cdots \\ & \vdots & & \end{bmatrix}$$

Suppose $\rho = 1$. Then Ω contains all 1's.

Implications: ϵ_t is random walk.

$Cov(\epsilon_t, \epsilon_{t-s}) = \rho^s$ does not $\downarrow$ as $s \uparrow$

This process (ϵ_t) is nonstationary.

$\boxed{\Sigma\ n^2 \text{ terms blows up!}}$ $\Rightarrow$ OLS $\hat{\beta}$ is not consistent in this case;
$\exists$ too much correlation across ϵ_t & ϵ_{t-s}.

Generally, if ϵ is autocorrelated,
must assume that the covariances further from
the diagonal of Ω become smaller as $s \uparrow$.
-- [likely true in reality!]

Think of the sum of weights in Q^*, rather than just the number of terms in the sum.

Then, as n grows larger (dimension of $\Omega \uparrow$),

this sum of weights must fall off for terms further off the diagonal,

so that this sum is of order n rather than n^2.

Thus we achieve convergence of Q^* if we assume:
(i) the X's behave, and
(ii) the correlations diminish with increasing separation in time.

2. Practical Problem with $\boxed{\hat{Q}^*}$

$\boxed{\hat{Q}^*}$ need not be positive definite. ← ~~This violates what~~ we know must be true $\forall$ cov. matrix!

Newey & West (1987) devised an estimator that overcomes this difficulty:

$$\hat{Q}^* = S_0 + \frac{1}{n}\sum_{\ell=1}^{L}\sum_{t=\ell+1}^{n} \omega_\ell e_t e_{t-\ell}\left(x_t x_{t-\ell}' + x_{t-\ell} x_t'\right)$$

where $\omega_\ell = \dfrac{\ell}{L+1}$

Note: L = maximum lag $[\mathrm{Cov}(\epsilon_t, \epsilon_{t-L})]$ incorporated.
Must be determined in advance; large enough that $\rho^\ell \approx 0 \; \forall \ell > L$.
See Davidson & MacKinnon (1993).

Because of this need to determine L,
the Newey-West estimator for $Cov(\hat{\beta})$ — autocorrelation
is not as "clean" as the White estimator — heteroskedasticity

Still Newey-West estimator is surprisingly simple to use.

What about hypothesis tests?

- Need "distribution" of $\hat{Q}^*$ for White & Newey-West.
- depends on OLS $\hat{\beta}$ → e.

We know OLS $\hat{\beta}$ is asymptotically normal,
& White or Newey-West $\hat{Q}^*$ is appropriate
asymptotic covariance matrix.

But, have not specified distribution of disturbances.
[ie. if ϵ not normal.]

⇒ F is approximate at best.

⇒ don't have likelihood ratio test (w/o normality assumption)

This leaves the Wald test statistic,
including asymptotic t-ratios
as main tool for testing hypotheses

☆☆ White & Newey-West estimators for $Cov(\hat{\beta})$
are now ubiquitous — everywhere!

Represent major advances in our set of tools.
Use in SAS: PROC REG; MODEL Y = X1 X2 / ACOV;

(*)

Obtaining Newey – West Standard Errors

To obtain Newey-West heteroskedasticity & autocorrelation-adjusted standard errors requires some work. Instead of using proc reg, we'll use GMM & Proc Model. Here is an example.

proc model data=a;	(notice the proc model statement instead of proc reg)
parms b0 b1 b2;	(here you define your parameters)
instruments x1 x2;	(here list your explanatory variables)
y = b0 + b1*x1 + b2*x2;	(your model)
fit y / gmm kernel = (bart, ***L+1***, 0);	(don't ask; if you really want to know see Ted Juhl)

```
test b1=0,b2=0;
test b1+b2=.5;
```

Here L = maximum lag length allowed for autocorrelation across errors.
Note1: Do NOT insert this exact notation ---
insert a number, the max lag you want to allow + 1.
For example, if you choose L+1 = 5 (i.e., put 5 in the above command), then you allow a maximum of 4 nonzero lags of autocorrelated errors in Ω.

Note2: If you choose L+1 = 1, then you don't allow any autocorrelated errors. Then this procedure only corrects for heteroskedasticity (the White-adjustment).

An alternative way to get (just) White-adjusted std errors (**)
[not as good as (*)]

Using White-adjusted standard errors to obtain asymptotically valid tests on the OLS parameter estimates.

```
title 'OLS regression with White-adjusted standard errors';
title2 'to correct for influence of possible heteroskedasticity';
title3 'on the standard errors of the OLS parameter estimates.'
title4 'Use the option, acov, to get White-adjusted covariance matrix, Q*';
title5 'Then use the Test command to test any linear restrictions desired';
title5 'Test commands with 1 degree of freedom give White-adjusted t-ratios';
title6 'Note:  test is really distributed asymptotically chi-square, rather than t';
title7 'because White-adjusted standard errors are asymptotic, not exact';
title8 'See Hal White paper, Econometrica, 1980, pp. 820-821.';
```

Example:

```
proc reg data=a1;
        model y = x1 x2 x3/acov;
        test x1=0;
        test x2=0;
        test x3=0;
run;
```