long-horizon
ex. forecasting regressions

Autocorrelation

Recall from our discussion of heteroskedasticity that all we need to proceed w/OLS is (A1)-(A3) plus $E(\varepsilon\varepsilon'|X) = \Omega$ and a way to estimate Ω.

For serial correlation, we usually make simplifying assumps so we don't have to estimate $T\times T$ params of Ω:

stationarity: Ω_{ts} is a function of $|t-s|$ only

low-dimension: error follows AR(1) or something similar

AR(1) errors:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \qquad E(u_t)=0 \qquad E(u_t^2)=\sigma_u^2 \qquad E(u_t u_s)=0 \ \forall t \neq s$$

what does Ω look like here?

$$\text{var}(\varepsilon_t) = \rho^2 \text{var}(\varepsilon_{t-1}) + \sigma_u^2$$

and by stationarity

$$\sigma_\varepsilon^2 = \rho^2\sigma_\varepsilon^2 + \sigma_u^2 \Rightarrow \sigma_\varepsilon^2 = \frac{\sigma_u^2}{1-\rho^2}$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t-1}) = E(\varepsilon_t\varepsilon_{t-1}) = E[(\rho\varepsilon_{t-1}+u_t)\varepsilon_{t-1}] = \rho\sigma_\varepsilon^2$$

and

$$\text{cov}(\varepsilon_t, \varepsilon_{t-s}) = E\left[\left(\rho^s\varepsilon_{t-s} + \sum_{i=0}^{s-1}\rho^i u_{t-i}\right)\varepsilon_{t-s}\right] = \rho^s\sigma_\varepsilon^2$$

so the autocorrelation is ρ^s. This means that

$$\Omega = \frac{\sigma_u^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & & & \vdots \\ \vdots & & \ddots & & \rho \\ \rho^{T-1} & & \cdots & \rho & 1 \end{bmatrix}$$

Recall from before that w/ strict exogeneity

$$\mathrm{var}(\hat\beta) = (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

when we relax strict exogeneity, asymptotics depend on

$$Q_T = \frac{1}{T} X'X \quad \text{and} \quad Q_T^* = \frac{1}{T} X'\Omega X$$

we need these to converge, but we don't have independence like we did before.

Turns out the main thing we need is

<u>Assump</u>

$[x_t, \varepsilon_t]$ is a jointly stationary and ergodic process

If we don't know the form of Ω, the key to using OLS is to pin down $Q_T^* = \frac{1}{T} X'\Omega X$

Expand this to form: $Q_T^* = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \gamma_{ts} \underset{\sim}{x}_t \underset{\sim}{x}_s'$ where x_t is the t^{th} row of X

The issue is how to estimate this quantity $\gamma_{ts} = \mathrm{cov}(\varepsilon_t, \varepsilon_s)$

Recall that if errors are not autocorrelated, then this reduces to the heteroskedasticity situation and we can use White's estimator

$$S_0 = \frac{1}{T} \sum_{t=1}^{T} e_t^2 x_t x_t'$$

with $\mathrm{avar}(\hat\beta) = T(X'X)^{-1} S_0 (X'X)^{-1}$

In the more general case, you might think to use

$$\hat{Q}_T^* = \frac{1}{T} \sum_t \sum_s e_t e_s x_t x_s'$$

but there's a practical problem: this need not be p.d.

p.206

Newey-West (1987) overcome this w/a particular set of linearly declining weights:

$$\hat{Q}_* = S_0 + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T} w_l e_t e_{t-l}(x_t x_{t-l}' + x_{t-l} x_t')$$

where $w_l = 1 - \dfrac{l}{L+1}$

last problem is how big L should be (how many lags?) in practice

errors AR(1) → ~~__~~ corrs lags decline exponentially but are never zero

but we have to cut it off ~~__~~ somewhere

MA(1) → 1 lag is enough

overlap → depends on degree of overlap (ex. use dyld Forecast regressio NB Hansen-Hodrick SE's better

~~__~~ Outside finance, often $L \approx T^{1/4}$, but not std in finance

~~__~~ L too big better than L too small

Testing for autocorrelation

regress $e_t = r e_{t-1} + v_t$ test $H_0: r = 0$

or use more lags and test all = 0

equiv. to Box-Pierce Q stat

$$Q = T\sum_{j=1}^{p}\hat{\rho}_j^2 \sim \chi_p^2 \text{ under } H_0 \qquad \hat{\rho}_j = \mathrm{Corr}(e_t, e_{t-j})$$

or the Ljung-Box refinement

$$Q = T(T+2)\sum_{j=1}^{p}\frac{\hat{\rho}_j^2}{T-j} \sim \chi_p^2$$

what if Ω is known (or we're willing to assume it)? GLS

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) \qquad \hat{\mathrm{var}}(\hat{\beta}) = \hat{\sigma}_\varepsilon^2\left[X'\Omega^{-1}X\right]^{-1}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{(y - X\hat{\beta})'\Omega^{-1}(y - X\hat{\beta})}{T}$$

Or equiv Can apply transformation to recover ideal conds.

AR(1) case:

$$y^* = \begin{bmatrix} \sqrt{1-\rho^2}\, y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} \qquad X^* = \begin{bmatrix} \sqrt{1-\rho^2}\, X_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_T - \rho x_1 \end{bmatrix}$$ and apply OLS

what if we don't know ρ? FGLS

1. OLS
2. residual regression to est $\hat{\rho}$
3. transform data, run OLS, use OLS s.e.'s

called Prais-Winsten or Cochrane-Orcutt

no point in iterating, since asymp. correct

Again, be careful: GLS inconsistent if the model of Ω is wrong
normally better to use OLS w/adjusted std errors

Classical Tests of Asset Pricing Models (wk. 10)

These are older regression-based tests of CAPM and factor models.

Important to look at older tests because some methods are useful in lots of contexts outside CAPM.

There are lots of ways to derive the CAPM, and I won't go through them again, except to refer you to CLM 5.2 and Cochrane Chap. 9.1

but also, in general, if

- all agents agree that $\underset{N\times 1}{\underset{\sim}{r}} \sim (\underset{\sim}{\mu}, \Sigma)$
- all agents hold mean-variance efficient pfls
- perfect capital markets exist

then
the wealth pfl or mkt pfl is m-v efficient

Roll (1977) shows that the SML follows directly if the mkt pfl is efficient

Sharpe-Lintner CAPM assumes $\exists\ r_f$ for borrowing and lending

$$\Rightarrow \quad ER_i = R_f + \beta_{im}(ER_m - R_f) \qquad \beta_{im} = \frac{Cov(R_i, R_m)}{Var(R_m)}$$

or more compactly in excess returns (lower case)

$$Er_i = \beta_{im} Er_m \qquad \beta_{im} = \frac{Cov(r_i, r_m)}{Var(r_m)}$$

(if we assume the riskless rate is nonstochastic)

If we estimate

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it} \qquad \text{CAPM} \Rightarrow \alpha_i = 0 \quad \forall i$$

Note that CAPM is a single-period model, so we have to m some further assumptions if we're going to use data.

The classic assumptions: CAPM holds period by period, and return "unconditional"

Fama-MacBeth (1973) - 1^{st} use of cross-sectional regs to test CAPM

very important; because this technique has been adapted to lots of other apps.

Basic Idea: each period, run CSR $r_{it} = \gamma_{0t} + \gamma_{1t}\beta_{i} + \varepsilon_{it}$ $i=1,\dots,N$

Then take avgs of the coefs over time:

$$\hat{\gamma}_0 = \frac{1}{T}\sum_{t=1}^{T} \hat{\gamma}_{0t} \qquad \hat{\gamma}_1 = \frac{1}{T}\sum_{t=1}^{T} \gamma_{1t}$$

They also propose that if returns are indep over time, so are $\hat{\gamma}_{0t}$ & $\hat{\gamma}_{1t}$, so

$$var(\hat{\gamma}_0) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\gamma}_{0t} - \hat{\gamma}_0)^2 \qquad var(\hat{\gamma}_1) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\gamma}_{1t} - \hat{\gamma}_1)^2$$

H_0: Sharpe-Lintner CAPM $\Rightarrow$ $\gamma_0 = 0$ $\gamma_1 > 0$

Note: we've assumed independence over time, but we don't really have to. We could calculate an autocorr-consistent std. error and be on our way. This is particularly important when applying this technique to, say, corp. finance data.

How do we implement this? Need assets / pfls and betas

~~FMV suggest~~

Problem: we're always using $\hat{\beta}$, so there's a potential EIV problem
To minimize, use portfolios. But how should we ~~choose~~ sort into pfls?

One idea: assign pfls randomly.
Problem: all β's are close to one, so no RHS variability $\Rightarrow$ no power

Two idea: rank on $\hat{\beta}$ in indiv sec. mkt model time-series regression
Problem: high $\hat{\beta}$'s probably overest. true β "regression phenomenon"

Three idea: 3 periods
ranking (on $\hat{\beta}_i$ in mkt model)
estimation (of β_p's)
CSR

They use: ~~estimation of~~

ranking (4 yrs of indiv-sec. monthly data)
estimation (next 5 yrs of pfl. returns in mkt model)

Then 9 yrs in you can start the CSR's (they use 1935–1968 CRSP data)

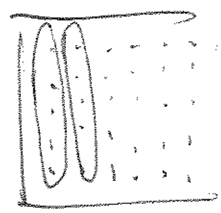
Advantage of F/M approach: you can also throw in other vars.
F/M examine idiosyncratic variance and squared beta
CLM mention size

e.g. we could test $\gamma_2 = \gamma_3 = 0$ in

$$r_i = \gamma_0 + \gamma_1 \beta_i + \gamma_2 \sigma_i + \gamma_3 \, \text{size}_i + \varepsilon_i$$

what do F/M find? $\hat{\gamma}_1 > 0$
no evidence that σ_i, β_i^2 matter
$\hat{\gamma}_0 > 0$ $\Leftarrow$ reject <u>Sharpe</u>-Lintner version

<u>More on</u> what F/M are doing by averaging CSR coef ests...



Think of ~~this~~ the overall data as a big panel: $i = 1, \ldots, N$ $t = 1, \ldots, T$

$$r_{it} = \gamma_0 + \gamma_1 \beta_{it} + \varepsilon_{it}$$

You could just pool everything together and estimate $\hat{\gamma}_0, \hat{\gamma}_1$.
But... $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) \neq 0$ $\forall$ combos of i, j, s, t

so if you did this w/o correcting std. errors, you'd be doing <u>bad</u> things
This is clearly wrong.

F/M assume that $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0$ $\forall s \neq t$ pricing errors uncorr'd over time
and allow arbitrary C-S correlation of errors

Thought experiment: if the errors are perfectly correlated in each CSR, there's really only 1 obs. per period, and F/M approach is efficient

If CSR errors aren't perfectly correlated, there's some info there that we throw out by doing F/M. But we don't have to try to model it if we use F/M

Cochrane 12.3.1 has a nice discussion of F/M compared to other ways of aggregating the data.

He points out one further advantage of F/M: it can capture risk-return w/ time-variation in β's ; pure CSR's (discussed below) cannot.

A couple of other notes: you don't need 9 yrs of data to get started. In Fama-French (1992), they showed that you just need a way of assigning portfolios that avoids the regression phenomenon.

So, for instance, you can rank based on size or another char or previous β and then just use the whole sample to calculate β_p's (if you assume they are constant over time).

form pfls/rank ~~rank~~ using $[t-60, t-1]$, calc r_{pt}

next months: rank using $[t-59, t]$, calc $r_{p,t+1}$

and so on

Black (1972) CAPM

The F/M evidence $\Rightarrow \hat{\gamma}_0 > 0$, which violates S-L CAPM

But this is fine in the absence of a riskless asset. W/o a riskless asset,

$$ER_i = ER_z + \beta_{im}(ER_m - ER_z)$$

where R_z is the return on the min-var pfl that is uncorrelated with m

(mathematics of the efficient set show there is one such pfl, and if CAPM is true, it's held in net amount zero)

$$\Rightarrow ER_i = ER_z(1-\beta_{im}) + \beta_{im} ER_m$$

if we estimate a mkt model

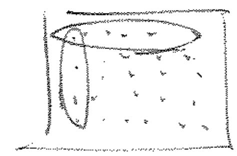
$$R_{it} = \alpha_i + \beta_{im} R_{mt} + \varepsilon_{it}$$

H_0: Black CAPM $\Rightarrow \alpha_i = \mu_z(1-\beta_{im}) \quad \forall i$, which imposes a c-s restriction on the α_i's

But before we go off to test the Black model, let's look at the other approach to testing S-L CAPM

Time-series regressions — Black, Jensen & Scholes (1972); Gibbons, Ross, Shanken (1989)

instead of doing CSR's a la F/M, you might instead do TSR's for each pfl



$$r_{it} = \alpha_i + \beta_{im} r_{mt} + \varepsilon_{it} \quad , \quad t = 1, \dots, T$$

Test H_0: $\alpha_i = 0 \quad \forall i$

BJS assume ideal conditions for ε_{it}, but this can be relaxed in the usual way

what BJS do is test $\alpha_i = 0$ individually for 20 pfls (similar construction techniques to F/M), i.e. 20 t-tests

what we really want is a joint test of $\alpha_i = 0 \ \forall i$

so we need to set up a system of equations

$$\begin{aligned} r_{1t} &= \alpha_1 + \beta_1 r_{mt} + \varepsilon_{1t} \\ &\ \vdots \\ r_{Nt} &= \alpha_N + \beta_N r_{mt} + \varepsilon_{Nt} \end{aligned}$$

and figure out the effect of $\text{cov}(\varepsilon_{it}, \varepsilon_{jt}) \quad \forall i \neq j$

If we assume that returns are iid MVN, then we can set this up as a MLE problem rather than as a system of OLS eqns and derive the result that:

$$\hat{\underset{\sim}{\alpha}} \sim N\left(\alpha, \frac{1}{T}\left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right] \Sigma \right)$$

where

$$\hat{\mu}_m = \bar{r}_m \qquad \hat{\sigma}_m^2 = \frac{1}{T}\sum_{t=1}^{T}(r_{mt} - \bar{r}_m)^2 \qquad \Sigma = E(\varepsilon_t \varepsilon_t')$$

for which ~~and~~ we can use the estimate $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t'$

Intuition behind this formula:

what if no market term? then $\text{var}(\hat{\alpha}) = \frac{1}{T}\Sigma$

but the sampling variability in $\hat{\beta}$ means a little extra variability in $\text{var}(\hat{\alpha})$

— exact same idea that we had in the event study

So this tells us how to do the test of $\alpha_i = 0 \quad \forall i$

We can just do a Wald test, which just tests the sum of squared α's scaled by var

$$\hat{\alpha}' \, Var(\hat{\alpha})^{-1} \hat{\alpha} \sim \chi^2_N$$

$$T \, \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \sim \chi^2_N$$

This is only valid in large samples, because we don't account for the fact that $\hat{\Sigma} \neq \Sigma$ except in the limit

(in fact, it can be shown that $T\hat{\Sigma} \sim W_N(T-2, \Sigma)$, i.e. a Wishart dist'n with $T-2$ d.f.'s and cov matrix Σ, which is a multivariate generalization of the χ^2 dist'n, but that's more than you need here)

if we take the MVN assumption seriously, then we can do an F-test rather than just an asymptotic test (just like in a simple regression)

It turns out that $\dfrac{T-N-1}{N} \left[1 + \dfrac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{N, T-N-1}$

(see CLM which refers to a multivariate text by my man Muirhead)

This is just the Gibbons Ross Shanken (1989) GRS test statistic

Also, be aware that you can do a likelihood-ratio test by also estimating the model with $\alpha_i = 0$ (constrained)
see CLM pp. 193-195 if you're interested

Recall: from Roll, single beta representation $\iff$ r_m is m-v. efficient

So this test is a test of whether r_m is close enough to the ex-post m-v frontier that it can be explained by sampling variability

GRS show that their test statistic can also be written as

$$\frac{T-N-1}{N} \; \frac{\left(\dfrac{\hat{\mu}_q^2}{\hat{\sigma}_q^2} - \dfrac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right)}{1 + \dfrac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}}$$

where q is the ex-post tangency pfl. using N assets + mkt return

Intuition: how far is the ex-post Sharpe ratio on the mkt. from the max ex-post Sharpe ratio (that's the tangency pfl q)

Empirical results (GRS): reject S-L CAPM ($\alpha_i \neq 0$)
mostly $\alpha_i > 0$

so we're led back to thinking about the Black version

recall that
H_0: Black CAPM $\Rightarrow$ $\alpha_i = \gamma(1-\beta_i) \quad \forall i$

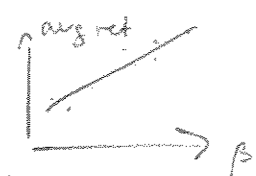
How do we test this? Harder, because it's a nonlinear restriction

In vector notation $\quad \underset{\sim}{\alpha} = \gamma(\underset{\sim}{1} - \underset{\sim}{\beta})$

In an MLE context, there are 2 approaches:
1. linearize the non-linear constraint and do a Wald test
2. estimated the restricted model (this isn't easy either) and do an LR test

For more details, see Gibbons (1982), Shanken (1985), CLM 5.3.2
but that's enough for us

Let's loop back and look at a "pure CSR" (not Fama-MacBeth)



Again, the basic idea is to regress
and test the intercept (for S-L)
or potentially throw in other variables

Technique: 1 estimate β's using a TSR on the whole sample
2 estimate one big CSR $\quad \bar{R}_i = \gamma_0 + \gamma_1 \hat{\beta}_i + \varepsilon_i$

Sometimes called a two-pass regression for obvious reasons

As before, the observations aren't independent, and $\Sigma = E(\underset{\sim}{\varepsilon}_t \underset{\sim}{\varepsilon}_t')$

$$Y = X\beta + u \qquad E(uu') = \Omega$$

in this more general case, you may recall that

$$\text{cov}(\hat\beta) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$, which reduces if $\Omega = \sigma^2 I$

We can use that to do testing, of course.
But you might be tempted to use GLS

$$\hat\beta_{GLS} = (X'\hat\Omega^{-1}X)^{-1}X'\hat\Omega^{-1}Y \qquad \text{var}(\hat\beta_{GLS}) = (X'\hat\Omega^{-1}X)^{-1}$$

This is more efficient asymptotically ($\Rightarrow$ more power)
But is less robust, especially if Ω is big (lots of assets)

Current thinking almost never advocates GLS in financial work (but see below)

Shanken (1992) addresses this and also the fact that β is estimated

He provides correction factors for var($\hat\gamma_0$) and var($\hat\gamma_1$)
(they're bigger than the OLS formulas)

Cochrane gives them in 12.2.3 and also discusses their magnitudes in typical estimation environments.

But the main point: you need to make these corrections in any CSR, whether a pure CSR or a Fama-MacBeth CSR

This would seem to argue for the TSR, which has none of these problems. (ok, the statistic still isn't trivial, but it's straightforward)
still have to think in terms of a system of eqns

The main disadvantage of the TSR: you need a factor that's also a return
Obviously this is fine for the CAPM, as long as you ID a proxy for the mkt pfl
but it can be an issue elsewhere

Cochrane 12.2.4 shows that a TSR is equivalent to a GLS pure CSR that also includes r_m as a test asset.

How should we choose pfls?

Early literature ranked on β. This probably helps tests of CAPM.

Anomalies literature basically did these same tests with judiciously chosen portfolios sorted on the characteristic of interest.

Turns out that these pfls as a group have a tangency pfl that is much better than the mkt proxy (ex post, anyway)

Biggest one is size (market cap), Banz (1981)
small firms earn more than their β would predict

(tho' it's interesting to note that there hasn't been a size premium since 1980)

The basic drill now is to:
- find a characteristic that you think is assoc'd with Er's and sort stocks based on this characteristic
- compute betas for the pfls, and ck if $Er - \beta$ is well-fit by a straight line
- if not, voila! an anomaly is born

This is the idea behind Fama-French (1992, 1993) who form 25 pfls sorted by size and book-to-mkt ratios

The Fama-French 25 have become very popular as test assets

They find:
$E(r) - \beta$ relation is basically flat in more recent data
"value" stocks do much better than their β suggests (high B/M)
(low M/B)

they also find the size premium again

Stack all the obs in a system of equations

$$\underset{mT\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}} = \underset{mT\times 2m}{\begin{bmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_m \end{bmatrix}} \underset{2m\times 1}{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}} + \underset{mT\times 1}{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}}$$

See Greene Chap. 14

In the case of testing CAPM, we have $X_i = X = [\,\underset{\sim}{1} \;\; r_m\,] \;\; \forall i$

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} X'X & & & 0 \\ & X'X & & \\ & & \ddots & \\ 0 & & & X'X \end{bmatrix}^{-1} \begin{bmatrix} Xy_1 \\ \vdots \\ Xy_m \end{bmatrix}$$

$$= \begin{bmatrix} (X'X)^{-1} & & 0 \\ & \ddots & \\ 0 & & (X'X)^{-1} \end{bmatrix} \begin{bmatrix} Xy_1 \\ \vdots \\ Xy_m \end{bmatrix}$$

So $\hat{\beta}_{OLS}$ for the system is the same as estimating OLS eqn by eqn

$\hat{\beta}$: what's the dist'n? Depends on dist'n of $\varepsilon_{\cdot} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}$

Assume

$$E[\,\varepsilon \mid X\,] = 0 \qquad E[\,\varepsilon_{it}\varepsilon_{js} \mid X\,] = \begin{cases} \sigma_{ij} & \text{if } t=s \\ 0 & \text{otherwise} \end{cases}$$

arbitrary c-s correl.

indep. over time

This can be written as

$$E[\,\varepsilon\varepsilon' \mid X\,] = \Omega = \begin{bmatrix} \sigma_{11}I & \cdots & \sigma_{1m}I \\ \vdots & & \\ \sigma_{m1}I & & \sigma_{mm}I \end{bmatrix} = \Sigma \otimes I$$

So you can see this doesn't satisfy the ideal conditions unless $\sigma_{ij} = 0 \;\; \forall i \neq j$ and $\sigma_{ii} = \sigma_{jj} \;\; \forall i, j$

Normally, we expect our returns to be correlated across portfolios even after taking out a common factor, so we expect $\sigma_{ij} \neq 0$.

But this is fine — we can still conduct inference. Recall...

$$\text{var}(\hat\beta) = (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

In the case of our stacked regression, we have $\text{var}(\hat\beta) =$

$$\begin{bmatrix} (X'X)^{-1} & & \\ & \ddots & \\ & & (X'X)^{-1} \end{bmatrix} \begin{bmatrix} X' & & 0 \\ & \ddots & \\ 0 & & X' \end{bmatrix} \begin{bmatrix} \sigma_{11} I & \cdots & \sigma_{1m} I \\ \vdots & & \\ \sigma_{m1} I & & \sigma_{mm} I \end{bmatrix} \begin{bmatrix} X & & \\ & \ddots & \\ & & X \end{bmatrix} \begin{bmatrix} (X'X)^{-1} & & \\ & \ddots & \\ & & (X'X)^{-1} \end{bmatrix}$$

which reduces to

$$\text{var}(\hat\beta) = \begin{bmatrix} \sigma_{11}(X'X)^{-1} & \cdots & \sigma_{1m}(X'X)^{-1} \\ \vdots & & \vdots \\ \sigma_{m1}(X'X)^{-1} & & \sigma_{mm}(X'X)^{-1} \end{bmatrix}$$

and all we have to do is replace σ_{ij} with $\hat\sigma_{ij} = \frac{1}{T} e_i' e_j$

and we can then do inference using asymptotic test statistics at least (and maybe even finite-sample tests as well)

In the case of testing the CAPM, we want to test $\hat a_i = 0 \ \forall i$

(recall that $\hat\beta = \begin{bmatrix} \hat a_1 \\ \hat b_1 \\ \vdots \\ \hat a_m \\ \hat b_m \end{bmatrix}$ here), and this is fairly straightfwd to implement

where does the GRS statistic come from? Are they the same?

$$(X'X)^{-1} = \begin{bmatrix} 1'1 & r_m' 1 \\ 1' r_m & r_m' r_m \end{bmatrix}^{-1} = \begin{bmatrix} T & T\hat\mu_m \\ T\hat\mu_m & T(\hat\mu_m^2 + \hat\sigma_m^2) \end{bmatrix}^{-1} = \frac{1}{T\hat\sigma_m^2} \begin{bmatrix} \hat\mu_m^2 + \hat\sigma_m^2 & \hat\mu_m \\ \hat\mu_m & 1 \end{bmatrix}$$

and the $(1,1)^{\text{th}}$ element of this matrix is $\frac{1}{T}\left[1 + \frac{\hat\mu_m^2}{\hat\sigma_m^2}\right]$

so the GRS statistic is just the Wald stat $\hat\alpha' \text{Var}(\hat\alpha)^{-1} \alpha$