Lecture 3 & 4
OLS: Data Problems, Bootstrapping and Testing

Brooks (4th edition): Chapters 4 & 5

 1

Review: Maximum Likelihood Estimation

• We get an *independent* sample (X_1, X_2,... , X_N). We assume we known where this sample is drawn from: A distribution with pdf $f(\boldsymbol{X}|\theta)$ where θ are k parameters.

The joint pdf is given by:

$$L(\boldsymbol{X}|\theta) = f(X_1,\ X_2, ..., X_N|\theta) = f(X_1|\theta) * f(X_2|\theta) * \cdots * f(X_N|\theta)$$
$$= \prod_{i=1}^{N} f(X_i|\theta)$$

• $L(\boldsymbol{X}|\theta)$: **Likelihood function**. It represents how likely it is to get a particular sample from the model.

• We maximize $L(X|\ \theta)$ w.r.t. θ to get ML estimates: $\hat{\theta}_{MLE}$

• It is easier to work with the **Log of the likelihood** function:

$$\ln L(X|\theta) = \sum_{i=1}^{N} \ln f(X_i|\theta)$$

Review: ML Estimation – Properties

• ML estimators (MLE) have very appealing properties:

(1) **Efficiency.** Lowest Variance of any estimator of θ.

(2) **Consistency:** $\hat{\theta}_{MLE} \xrightarrow{p} \theta$

(3) **Asymptotic Normality**: $\hat{\theta}_{MLE} \xrightarrow{a} N(\theta, \mathbf{I}(\theta|X)^{-1})$

where $\mathbf{I}(\theta|X)$ is the information matrix for the whole sample.

$$E\left[\left(\frac{\partial \log L}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log L}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}}\right] = \mathbf{I}(\theta|X)$$

(4) **Invariance.** If $\hat{\theta}_{MLE}$ is the MLE of $\theta \Rightarrow g(\hat{\theta}_{MLE})$ is the MLE of $g(\theta)$.

Review: ML Estimation – Linear Model

• Suppose we assume, using the usual notation:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I}_T)$$

where we have k explanatory, exogenous variables, $\boldsymbol{x}_i$'s , that we treat as numbers. $\boldsymbol{\beta}$ is a kx1 vector of unknown parameters.

Then, the joint likelihood function becomes:

$$L = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-T/2} \prod_{i=1}^{T} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right)$$

• Taking logs to get the log likelihood function, a function of $(\sigma^2, \boldsymbol{\beta})$:

$$\ln L = -\frac{T}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{T}\varepsilon_i^2 = -\frac{T}{2}\ln 2\pi\sigma^2 - \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}$$

$$= -\frac{T}{2}\ln 2\pi\sigma^2 - \frac{\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}}{2\sigma^2}$$

Review: ML Estimation – Linear Model

- After taking f.o.c. and solving for $\hat{\boldsymbol{\beta}}_{MLE}$ & $\hat{\sigma}^2_{MLE}$:

$$\hat{\boldsymbol{\beta}}_{MLE} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$$

$$\hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^{T} e_i^2}{T} = \frac{\sum_{i=1}^{T}(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}_{MLE})^2}{T}$$

- Under (**A5**) –i.e., normality for the errors–, we have that $\hat{\boldsymbol{\beta}}_{MLE}$ = **b**.

- It can be shown (see notes) that Var[$\hat{\boldsymbol{\beta}}_{MLE}$] = $\hat{\sigma}^2_{MLE}$ $(X'X)^{-1}$

Note: $\hat{\sigma}^2_{MLE}$ is biased, but as T gets bigger, the differences between $\hat{\sigma}^2_{MLE}$ and s^2_{OLS} become very small. Thus, with a big T (& normality) the difference between Var[$\hat{\boldsymbol{\beta}}_{MLE}$] & Var[$\boldsymbol{b}$] should be minor.

Review: ML Estimation – Linear Model

Example: We estimate the 3 F-F factor model for IBM with ML and OLS.

- Summary: OLS vs MLE

	OLS		**MLE**	
	Coeff. (1)	S.E.	Coeff. (2)	S.E.
Intercept	-0.00509	0.00238	-0.00509	0.00237
Mkt_RF	0.86761	0.05425	0.86761	0.05406
SMB	-0.68159	0.08045	-0.68159	0.08017
HML	-0.22842	0.08100	-0.22842	0.08071

Same as expected — Not so different

Review: Data Problems

- Three data problems (in general, exogenous to data user):

(1) **Missing Data** – very common, especially in corporate finance.
  - Detection: blanks, NA, etc. We know if the data has this issue.
  - Usual solutions: Impute values (fill in blanks), use inverse weights (give more weight to "unrepresented" observations).

(2) **Outliers** - unusually high/low observations.
  - Detection: Visual (graphs for standardized residuals), Formal measures for Leverage and for influence (Dif Beta & Cook's D).
  - Usual solutions: Transform data, remove outliers, Winsorization.

(3) **Multicollinearity** - High correlation in the explanatory variables.
  - Detection: VIF, Condition Index.
  - Usual solutions: Not much to do. Be aware of problem.

Bootstrapping (Again!): Review

Idea: We use the data at hand -the empirical distribution (ED)- to estimate the variation of statistics that are themselves computed from the same data. Recall that, for large samples drawn from F, the ED approximates the CDF of F very well.

- Bootstrap choice for an approximating distribution: The ED.

  ⇒ The ED becomes a "*fake population*."

John Fox (2005, UCLA): *"The population is to the sample as the sample is to the bootstrap samples."*

- Suppose we have a dataset with N *i.i.d.* observations drawn from F:

$$\{x_1, x_2, \ldots, x_N\}$$

This sample becomes the *"fake population."*

Bootstrapping: Review - Resampling

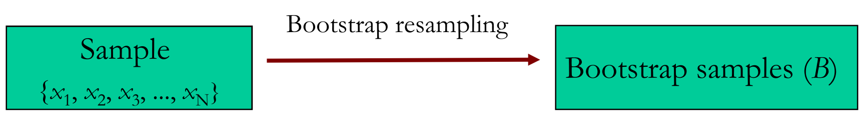
- We have a dataset with N *i.i.d.* observations drawn from F:

$$\{x_1, x_2, \ldots, x_N\}$$ -"*fake population.*"

From the ED, ***F****, we sample with *replacement* N observations, say:

$$\{x_1^*{=}x_1,\ x_2^*{=}x_1,\ x_3^*{=}x_7,\ \ldots,\ x_N^*{=}x_{N-10}\}$$ - a bootstrap sample

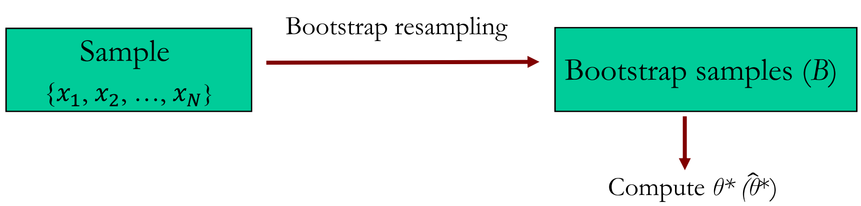
This is an *empirical bootstrap sample*, which is a resample of the same size N as the original data, drawn from ***F****. We will resample many times.



- For any statistic θ computed from the original sample data, we can define a statistic θ^* by the same formula, but using the resampled data.

Bootstrapping: Review – Resampling B Times

- We resample ***B*** times from F^*.



- We compute ***B*** $\hat{\theta}^*$, by resampling ***B*** times from F^*.

$\Rightarrow$ We have a collection of $\hat{\theta}^*$'s: $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \ldots, \hat{\theta}_B^*\}$.

From this collection of $\hat{\theta}^*$'s, we learn about statistic θ: Compute moments, C.I.'s, etc.

Bootstrapping: Review - Empirical Bootstrap

• Bootstrap Steps:

1. From the original sample, draw random sample with size N.
2. Compute statistic θ from the resample in 1: $\hat{\theta}_1^*$.
3. Repeat steps 1 & 2 ***B*** times ⇒ Get B statistics: $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, ..., \hat{\theta}_B^*\}$
4. Compute moments, draw histograms, etc. for these ***B*** statistics.

• Results:

1. With a large enough ***B***, the LLN allows us to use the $\hat{\theta}^*$'s to estimate the distribution of $\hat{\theta}$, $F(\hat{\theta})$.
2. The variation in $\hat{\theta}$ is well approximated by the variation in $\hat{\theta}^*$.

Result 2 is the one we used in Lecture 2-d to estimate the size of a C.I.

Bootstrapping: Review - Why?

• Q: Why do we need a bootstrap?

- N is "small," asymptotic assumptions do not apply.
- DGP assumptions are violated.
- Distributions are complicated.

• Advantages and Disadvantage:

- Only *consistent* results, no finite sample results.
- Main appeal: Simplicity.

• The most common econometric applications are situations where you have a consistent estimator of a parameter of interest, but it is hard or impossible to calculate its standard error or its C.I.

Bootstrapping: Simple correlation example

• You are interested in the correlation between IBM's returns (**X**) and S&P 500 returns (**y**). You have monthly data from 1973 (N = 588). You estimate the correlation coefficient, ρ, with its sample counterpart, r = **0.5894632**. You find the correlation to be lower than expected.

Q: How reliable is this result? The distribution of r is complicated. You decide to use a bootstrap to study the distribution of r. Note that to compute r, we need to bootstrap pairs, then, we use a *paired bootstrap*.

• Randomly construct a sequence of B samples (all with N = 588). Say,

$B_1 = \{(x_1, y_1), (x_3, y_3), (x_6, y_6), (x_6, y_6), ..., (x_{1458}, y_{1458})\} \Rightarrow \hat{\theta}^*_1 = r_1$

$B_2 = \{(x_5, y_5), (x_7, y_7), (x_{11}, y_{11}), (x_{12}, y_{12}), ..., (x_{1486}, y_{1486})\} \Rightarrow \hat{\theta}^*_2 = r_2$

. . . .

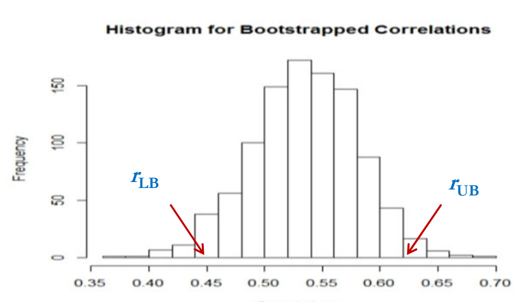
$B_B = \{(x_2, y_2), (x_2, y_2), (x_2, y_2), (x_3, y_3), ..., (x_{1499}, y_{1499})\} \Rightarrow \hat{\theta}^*_B = r_B$

Bootstrapping: Simple correlation example

• We have a collection of estimators of ρ_i's: $\{r_1, r_2, r_3, ..., r_B\}$.

• Below, we do a histogram to get an approximation of the probability distribution and, then, we build an empirical C.I., defined by $[r_{LB}, r_{UB}]$

$\Rightarrow \rho \in [r_{LB}, r_{UB}]$ with some probability



Bootstrapping: Simple correlation example

Remarks:

- We rely on the ED –i.e., observed data. We take it as our "*fake population*" and we sample from it ***B*** times.
- We have a collection of *bootstrap subsamples*.
- The sample size of each bootstrap subsample is the same (N). Some elements are repeated.

Bootstrapping: Estimating the correlation, ρ

Example: We bootstrap the correlation between the returns of IBM & the S&P 500, using monthly data 1973-2020, with ***B* = 1,000**.

```
sim_size = 1000
lr_sp <- log(x_sp[-1]/x_sp[-T])
dat_spibm <- data.frame(lr_sp, lr_ibm)
library(boot)
# function to obtain the correlation coefficient from the data
cor_xy <- function(data, i) {
        d <-data[i,]
        return(cor(d$lr_sp,d$lr_ibm))
}
# bootstrapping with sim_size replications
boot.samps <- boot(data=dat_spibm, statistic=cor_xy,   R=sim_size)

# view stored bootstrap samples and compute mean
boot.samps                  # Print original ρ, bias and SE of bootstraps
mean(boot.samps$t)          # our estimate of the correlation
```

Bootstrapping: Estimating the correlation, ρ

Example (continuation): Output from R:

```
ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = dat_spibm, statistic = cor_xy, R = sim_size)

Bootstrap Statistics :
    original       bias          std. error
t1* 0.5894632 -0.001523914  0.03406313

> boot.samps$t[1:10]                    # show first 10 bootstrapped correlations coeff
 [1] 0.5863186 0.5898572 0.6473122 0.6473249 0.5311525 0.5734280 0.6241236 0.5790740
 [9] 0.5790095 0.5932918
> mean(boot.samps$t)                    # our estimate of the correlation
[1] 0.5879392
> sd(boot.samps$t)                      # SD of the correlation estimate
[1] 0.03406313
```

Bootstrapping: Histogram for ρ

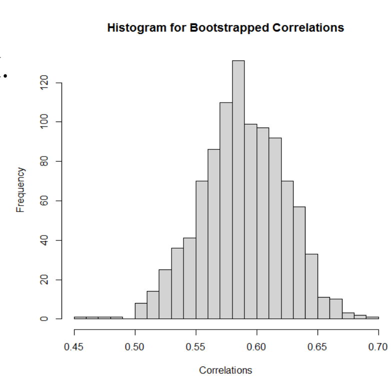
Example (continuation): Output from R:

```
> # Elegant histogram
> hist(boot.samps$t,main="Histogram for Bootstrapped Correlations",
+      xlab="Correlations", breaks=20)
```

- Simple 95% **percentile method** C.I.

```
> new <- sort(boot.samps$t)
> new[25]
[1] 0.5151807
> new[975]
[1] 0.6495722
```

Note: You get same results using boot.ci(boot.samps, type = "perc")



Bootstrapping: 95% Confidence Interval for ρ

Example (continuation): Output from R:

• 95% C.I using **empirical bootstrap method** (preferred method.)

```
> boot.ci(boot.samps, type="basic")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
boot.ci(boot.out = boot.samps, type = "basic")
Intervals :
Level     Percentile
95%   ( 0.5293, 0.6637 )
Calculations and Intervals on Original Scale
```

Bootstrapping: How many bootstraps?

• Not clear. There are many theorems on asymptotic convergence, but there are no clear rules regarding *B*. There are some suggestions, from ***B*** = 100 (or even ***B*** = 25!) to ***B*** = 2,400.

Rule of thumb: Start with ***B* = 100**, then, try ***B* = 1,000**, and see if your answers have changed by much. Increase bootstraps until you get stability in your answers.

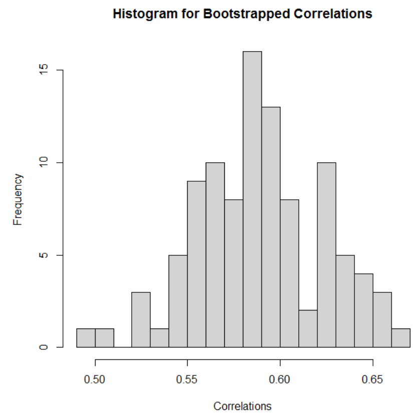
Bootstrapping: How many bootstraps?

Example: We bootstrap the correlation between IBM returns and S&P 500 returns, using $B = 100$.

```
sim_size <- 100
> # view bootstrap results
> boot.samps
ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = dat_spibm, statistic = cor_xy, R =
sim_size)

Bootstrap Statistics :
    original      bias     std. error
t1* 0.5898636 -0.00115623  0.03449216
> mean(boot.samps$t)
[1] 0.5887074
> sd(boot.samps$t)
[1] 0.02885868
```



- Results do not change that much.

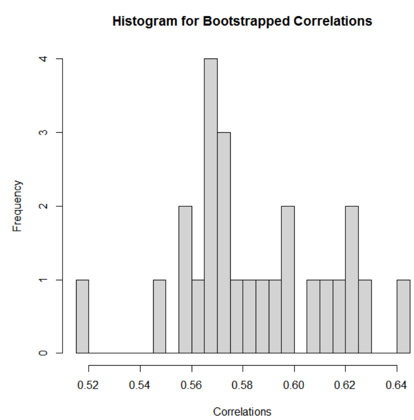
Bootstrapping: How many bootstraps?

Example: We bootstrap the correlation between IBM returns and S&P 500 returns, using $B = 25$.

```
sim_size <- 25
> # view bootstrap results
> boot.samps
ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = dat_spibm, statistic = cor_xy, R =
sim_size)

Bootstrap Statistics :
    original      bias     std. error
t1* 0.5898636 -0.00115623  0.03449216
> mean(boot.samps$t)
[1] 0.5847676
> sd(boot.samps$t)
[1] 0.03449216
```



- Results do not change that much.

Bootstrapping: Linear Model – Var[b]

• Some assumptions in the CLM are not reasonable, say, (**A3**) or normality (**A5**). By assuming (**A5**), we also assume the sampling distribution of **b**. But if data is not normal, results are only asymptotic.

• We use a bootstrap to estimate the sampling distribution of **b**. It can give us a better idea of the small sample distribution. Then, we estimate the Var[**b**].

• Monte Carlo (MC=repeated sampling) method:

1. Estimate model using full sample (of size T) ⇒ get **b**
2. Repeat B times:
   - Draw T observations from the sample, *with replacement*
   - Estimate $\boldsymbol{\beta}$ with OLS **b**. (We have a vector of b's, **b**(r).)
3. Estimate variance with

$$\mathbf{V}_{\mathbf{boot}} = (1/B)\ [\mathbf{b}(r) - \mathbf{b}][\mathbf{b}(r) - \mathbf{b}]'$$

Bootstrapping: Linear Model – Var[b]

• In the case of one parameter, say $\mathbf{b}_1$: Estimate variance with

$$\text{Var}_{\mathbf{boot}}[\mathbf{b}_1] = (1/B) \sum_{r=1}^{B} (\mathbf{b}_{1,r} - \mathbf{b}_1)^2$$

• You can also estimate Var[$\mathbf{b}_1$] as the variance of $\mathbf{b}_1$ in the bootstrap

$$\text{Var}_{\mathbf{boot}}[\mathbf{b}_1] = (1/B) \sum_{r=1}^{B} (\mathbf{b}_{1,r} - \text{mean}(\mathbf{b}_1(r)))^2$$

$$\text{mean}(\mathbf{b}_1(r)) = (1/B) \sum_{r=1}^{B} \mathbf{b}_1(r)$$

Note: Obviously, this method for obtaining standard errors of parameters is most useful when no formula has been worked out for the standard error (SE), or the formula is complicated –for example, in some 2-step estimation procedures– or the assumption behind the formula are not realistic.

Bootstrapping: Linear Model – Var[b]

Example: We bootstrap the SE for **b** for IBM returns using the 3 FF Factor Model. We use the R package *lmboot.* (Install it first!)

```
library(lmboot)                        # need to run before install.packages("lmboot")
y <- ibm_x
x <- cbind(x0, Mkt_RF, SMB, HML)
dat_yx <- data.frame(y, x)             # lmboot needs an R data frame. We make one.

sim_size = 1000
ff3_b <- paired.boot(y ~ x-1, data=dat_yx, B=sim_size)

ff3_b$origEstParam                     # print OLS results ("original estimates")

# Mean values for b
mean(ff3_b$bootEstParam[,1])           # print mean of bootstrap samples for constant
mean(ff3_b$bootEstParam[,2])           # print mean of bootstrap samples for Mkt_RF
mean(ff3_b$bootEstParam[,3])           # print mean of bootstrap samples for SMB
mean(ff3_b$bootEstParam[,4])           # print mean of bootstrap samples for HML
```

Bootstrapping: Estimating Var[b]

Example (continuation):

```
# Statistics for sampling distribution of b
summary(ff3_b$bootEstParam)            # distribution of b

# SD of parameter vector b
sd(ff3_b$bootEstParam[,1])             # print SD of bootstrap samples for constant
sd(ff3_b$bootEstParam[,2])             # print SD of bootstrap samples for Mkt_RF
sd(ff3_b$bootEstParam[,3])             # print SD of bootstrap samples for SMB
sd(ff3_b$bootEstParam[,4])             # print SD of bootstrap samples for HML

# bootstrap bias
ff3_b$origEstParam[1] - mean(ff3_b$bootEstParam[,1])
ff3_b$origEstParam[2] - mean(ff3_b$bootEstParam[,2])
ff3_b$origEstParam[3] - mean(ff3_b$bootEstParam[,3])
ff3_b$origEstParam[4] - mean(ff3_b$bootEstParam[,4])
```


Bootstrapping: Estimating Var[b]

Example (continuation):

```
> ff3_b$origEstParam
             [,1]
x        -0.005088944
xMkt_RF   0.908298898
xSMB     -0.212459588
xHML     -0.171500223

> summary(ff3_b$bootEstParam)
        x                  xMkt_RF            xSMB                 xHML
 Min.   :-0.012159   Min.   :0.7115   Min.    :-0.5175   Min.   :-0.4699
 1st Qu.:-0.006731   1st Qu.:0.8669   1st Qu. :-0.2890   1st Qu.  :-0.2362
 Median :-0.005074   Median :0.9087   Median  :-0.2185   Median  :-0.1690
 Mean   :-0.005008   Mean   :0.9068   Mean    :-0.2125   Mean    :-0.1710
 3rd Qu.:-0.003273   3rd Qu.:0.9492   3rd Qu. :-0.1415   3rd Qu. :-0.1086
 Max.   : 0.002293   Max.   :1.0854   Max.    : 0.1909   Max.    : 0.2477

> sd(ff3_b$bootEstParam[,1])
[1] 0.002493708
```

Bootstrapping: Estimating Var[b]

```
> ff3_b$bootEstParam[1:10,]              # print the first 10 of B=1,000 bootstrap samples

              x         xMkt_RF        xSMB          xHML
 [1,] -6.109007e-03 0.9186830 -0.1299534100 -0.163421636
 [2,] -1.757503e-03 0.8333006 -0.2067565390 -0.147604991
 [3,] -3.907573e-03 0.9746878 -0.2870744815 -0.169189619
 [4,]  1.596103e-03 0.9185157 -0.2937731120 -0.296972497
 [5,] -8.409239e-03 0.7309406 -0.0681714313 -0.149883639
 [6,] -1.998929e-03 0.9133751 -0.3001713380 -0.315913280
 [7,] -6.289286e-03 0.9441856 -0.2276894034 -0.058924929
 [8,] -5.533354e-03 0.8210057 -0.2221866298 -0.078512341
 [9,] -6.152301e-03 1.0389917 -0.2592958758 -0.237930809
[10,] -3.778058e-03 0.9544829 -0.1859554067 -0.217702583
```

- From the B samples, we compute variances and SD as usual.

Bootstrapping: Estimating Var[b]

```
> sd(ff3_b$bootEstParam[,2])
[1] 0.06132218
> sd(ff3_b$bootEstParam[,3])
[1] 0.1108
> sd(ff3_b$bootEstParam[,4])
[1] 0.09729972
>
```

Bootstratp has higher SE, more **conservative** tests: less H_0 rejections

- Comparing OLS and Bootstrap

	OLS		Bootstrap		Bias
	Coeff. (1)	S.E.	Coeff. (2)	S.E.	(2)-(1)
x	-0.00509	0.00249	-0.00501	0.00249	8.0765e-05
xMkt_RF	**0.90829**	**0.05672**	**0.90684**	**0.06132**	-0.0014571
xSMB	**-0.21246**	**0.08411**	**-0.21245**	**0.11080**	1.9914e-06
xHML	**-0.17150**	**0.08468**	**-0.17099**	**0.09730**	0.0005133

OLS Subject to Linear Restrictions

- Restrictions: Theory imposes certain restrictions on parameters and provide the foundation of several tests. In this Lecture, we only consider linear restrictions, written as $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$.

Dimensions:

$\mathbf{R}$: Jxk - J = # of restrictions & k = # of pars.

$\boldsymbol{\beta}$: kx1

$\mathbf{q}$: kx1

- We consider the following restrictions:
  - (1) Dropping variables from model ($\beta_{SMB} = 0$).
  - (2) Adding up conditions ($\beta_{SMB} + \beta_{HML} = 1$).
  - (3) Equality restrictions ($\beta_{SMB} = \beta_{HML} = 0$).

OLS Subject to Linear Restrictions

Examples: Linear restrictions, written as $\mathbf{R}\beta = \mathbf{q}$.

(1) Dropping variables from the equation. That is, certain coefficients in $\boldsymbol{\beta}$ are forced to equal 0. For example, in the 3-factor Fama-French factor model we force $\beta_{SMB} = \beta_{HML} = 0$, that is, we fit the traditional CAPM).

Using the above notation:

$$\mathbf{R}\beta = \mathbf{q} \qquad \Rightarrow \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_{Mkt} \\ \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = \begin{bmatrix} \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

We have two restrictions (J=2): $\beta_{SMB} = 0$ & $\beta_{HML} = 0$.

$\Rightarrow$ $\mathbf{R}$ is a 2x4 matrix, $\boldsymbol{\beta}$ is a 4x1 vector, and $\mathbf{q}$ is a 2x1 vector.

OLS Subject to Restrictions

Examples (continuation):

(2) Adding up conditions: Sums of certain coefficients must equal fixed values. In a CAPM setting, the sum of all cross-sectional β_i's should be equal to 1. For example, in the 3 Fama-French factor model, we force $\beta_{SMB} + \beta_{HML} = 1$.

$$\mathbf{R}\beta = \mathbf{q} \qquad \Rightarrow [0 \quad 0 \quad 1 \quad 1] * \begin{bmatrix} \beta_1 \\ \beta_{Mkt} \\ \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = \beta_{SMB} + \beta_{HML} = 1.$$

Note: From a theory point of view, not a very interesting restriction!

OLS Subject to Restrictions

Examples (continuation):

(3) Equality restrictions: Certain coefficients must equal other coefficients. Using real vs. nominal variables in equations. For example, in the 3 FF factor model, we force $\beta_{SMB} = \beta_{HML}$.

$$\mathbf{R}\beta = \mathbf{q} \qquad \Rightarrow [0 \quad 0 \quad 1 \quad -1] * \begin{bmatrix} \beta_1 \\ \beta_{Mkt} \\ \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = 0.$$

Note: From a theory point of view, not a very interesting restriction!

• Common formulation: We minimize the error sum of squares, subject to the linear restrictions. That is,

$$\text{Min}_b \; \{S(x_i, \theta) = \sum_{i=1}^{n} e_i^{\,2} = \boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - \mathbf{X}\beta)' (\boldsymbol{y} - \mathbf{X}\beta)\} \quad \text{s.t. } \mathbf{R}\beta = \mathbf{q}$$

Restricted Least Squares

• In many situations, restrictions can usually be imposed by solving them out. Suppose we have the following model:

$$y_i = \beta_1 \, x_{i1} + \beta_2 \, x_{i2} + \beta_3 \, x_{i3} + \varepsilon_i$$

(1) Dropping variables –i.e., force a coefficient to equal zero, say β_3.

Problem: $\text{Min}_\beta \; \sum_{i=1}^{n} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3})^2 \quad s.t. \; \beta_3 = 0$

$\Rightarrow \qquad \text{Min}_\beta \; \sum_{i=1}^{n} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2})^2$

(2) Adding up. We impose: $\beta_1 + \beta_2 + \beta_3 = 1 \qquad (\Rightarrow \beta_3 = 1 - \beta_1 - \beta_2)$

Then, substituting in model:

$$(y_i - x_{i3}) = \beta_1 (x_{i1} - x_{i3}) + \beta_2 (x_{i2} - x_{i3}) + \varepsilon_i.$$

Problem: $\text{Min}_\beta \; \sum_{i=1}^{n} ((y_i - x_{i3}) - \beta_1 (x_{i1} - x_{i3}) - \beta_2 (x_{i2} - x_{i3}))^2$

Restricted Least Squares

(3) Equality. Suppose we impose: $\beta_2 = \beta_3$.
Then,

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_2 x_{i3} + \varepsilon_i = \beta_1 x_{i1} + \beta_2 (x_{i2} + x_{i3}) + \varepsilon_i$$

Problem: $\text{Min}_\beta \sum_{i=1}^{n} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3})^2 \quad s.t.\ \beta_2 = \beta_3$

$\text{Min}_\beta \sum_{i=1}^{n} (y_i - \beta_1 x_{i1} - \beta_2 (x_{i2} + x_{i3}))^2$

Restricted LS: One Restriction, $r\beta = q$

- Before setting the general restricted LS problem, we look at the simplest case: one explanatory variable (x) and one restriction ($r\beta - q$).

First, we set the Lagrangean (values of Lagrange λ play no role):

$$\min_{\beta,\lambda} L(\beta, \lambda) = \sum_{i=1}^{T} (y_i - x_i \beta)^2 + 2\lambda (r\beta - q)$$

Second, take f.o.c.:

$$\Rightarrow \frac{\partial L(\beta, \lambda)}{\partial \beta} = 2 \sum_i^T (y_i - x_i \beta)(-x_i) + 2\lambda r$$

$$\frac{\partial L(\beta, \lambda)}{\partial \lambda} = 2 (r\beta - q)$$

Then, the f.o.c. are:

$$-\sum_i^T (y_i - x_i b^*)(x_i) + \lambda r = 0 \qquad \Rightarrow \sum_i^T (y_i x_i - x_i^2 b^*) = \lambda r$$

$$2 (r b^* - q) = 0 \qquad \Rightarrow r b^* - q = 0$$

Restricted LS: One Restriction, $r\beta = q$

- From the 1st equation:

$$\sum_i^T y_i x_i - b^* \sum_i^T x_i^2 = \boldsymbol{x'y} - b^*(\boldsymbol{x'x}) = \lambda\, r$$

$$\Rightarrow b^* = (\boldsymbol{x'x})^{-1}\, \boldsymbol{x'y} - (\boldsymbol{x'x})^{-1} \lambda\, r$$

$b^* = b - r\,(\boldsymbol{x'x})^{-1}\,\lambda \quad \Rightarrow$ Restricted OLS = OLS + "*correction*"

- Finally, solve for λ. Premultiply both sides by r and then subtract q:

$$r\, b^* - q = r\, b - r^2\,(\boldsymbol{x'x})^{-1}\lambda - q$$

$$0 = -r^2\,(\boldsymbol{x'x})^{-1}\lambda + (rb - q)$$

Solving for λ $\quad \Rightarrow \quad \lambda = [r^2\,((\boldsymbol{x'x})^{-1}]^{-1}\,(rb - q)$

- Substituting in b^* $\quad \Rightarrow \quad b^* = b - (\boldsymbol{x'x})^{-1}\, r\,[r^2\,(\boldsymbol{x'x})^{-1}]^{-1}\,(rb - q)$

This is the Restricted OLS estimator:

Restricted OLS = Unrestricted OLS + correction

37

Restricted LS: One Restriction – Properties

- $b^* = b - (\boldsymbol{x'x})^{-1}\, r\,[r^2(\boldsymbol{x'x})^{-1}]^{-1}\,(rb - q)$

- Properties of Restricted OLS.

Property 1. Taking expectations of b^*:

$$\text{E}[b^*\,|\,\mathbf{X}] = \text{E}[b\,|\,\mathbf{X}] - (\boldsymbol{x'x})^{-1}\, r\,[r^2(\boldsymbol{x'x})^{-1}]^{-1}\,\text{E}[(rb - q)\,|\,\mathbf{X}]$$

$$= \beta - (\boldsymbol{x'x})^{-1}\, r\,[r^2(\boldsymbol{x'x})^{-1}]^{-1}\,(r\beta - q)$$

Implications:

If the restriction is *true* –i.e., $(r\beta = q)$ $\quad \Rightarrow \text{E}[b^*\,|\,\mathbf{X}] = \beta$

If the restriction is *not true* –i.e., $(r\beta \neq q)$ $\quad \Rightarrow \text{E}[b^*\,|\,\mathbf{X}] \neq \beta$

- Then, if theory imposes a correct restriction, then, b^* is *unbiased:*

$$\text{E}[b^*\,|\,\mathbf{X}] = \beta$$

In practice, if restriction is true, the restricted and unrestricted estimators should be similar.

38

Restricted LS: One Restriction – Properties

- Recall the LM: $\lambda = [r^2\,(\mathbf{x}'\mathbf{x})^{-1}]^{-1}\,(rb - q)$

Interpretation: If theory is correct, the expected shadow price is 0!

$$E[\lambda\,|\,\mathbf{X}] = [r^2\,(\mathbf{x}'\mathbf{x})^{-1}]^{-1}\,E[(rb - q)\,|\,\mathbf{X}] = 0$$

That is, you would pay nothing to release the restriction.

Property 2. We can also compute the Var[b^*]. It can be shown that

$$\text{Var}[b^*\,|\,\mathbf{X}] = \text{Var}[b\,|\,\mathbf{X}] - \sigma^2\,(\boldsymbol{x}'\boldsymbol{x})^{-1} r\,[r^2(\boldsymbol{x}'\boldsymbol{x})^{-1}]^{-1}\,r\,(\boldsymbol{x}'\boldsymbol{x})^{-1}$$

$$\Rightarrow \text{Var}[b\,|\,\mathbf{X}] - \text{Var}[b^*\,|\,\mathbf{X}] > 0.$$

⇒ The restricted OLS estimator is more efficient!

39

Restricted LS: One Restriction – Properties

Remark from Properties 1 and 2: It is common to select an estimator based on the MSE (=RSS/T). The one with the lowest MSE is said to be more "*precise*."

We can decompose the MSE of an estimator, $\hat{\theta}$, as:

$$\text{MSE}[\hat{\theta}] = \text{Variance}[\hat{\theta}] + \text{Squared bias}[\hat{\theta}]$$

For an unbiased estimator, like $\mathbf{b} \Rightarrow \text{MSE}\,[\mathbf{b}] = \text{Var}[\mathbf{b}\,|\,\mathbf{X}]$

- Back to $\mathbf{b}^*$. Suppose the theory is incorrect ⇒ $\mathbf{b}^*$ is biased.

There may be situations (small bias, but much lower variance) where $\mathbf{b}^*$ is more "precise" (lower MSE) than $\mathbf{b}$.

It is possible that a practitioner may prefer imposing a wrong H_0 to get a better MSE.

40

Restricted LS: General case, $R\beta = q$

• All the results for the one variable case, can be extended for the general case, with J restrictions. We have a programming problem:

Minimize wrt β $\quad L^* = (y - X\beta)'(y - X\beta) \quad$ s.t. $R\beta = q$

• Form the Lagrangean, L^* (the 2 multiplying λ is for convenience).

$$\text{Min}_{\beta,\lambda} \quad L^* = (y'y - 2\beta'X'y + \beta'X'X\beta) + 2\,\lambda\,(R\beta - q)$$

f.o.c.:

$$\partial L^*/\partial \beta' = -2X'y + 2X'Xb^* + 2R'\lambda = 0 \Rightarrow -X'(y - Xb^*) + R'\lambda = 0$$

$$\partial L^*/\partial\lambda = 2(Rb^* - q) = 0 \qquad \Rightarrow (Rb^* - q) = 0$$

where b^* is the restricted OLS estimator and λ is the Jx1 vector of Lagrange multipliers.

After (a lot of algebra) we get:

$$b^* = b - (X'X)^{-1}R'\,[R(X'X)^{-1}R']^{-1}(Rb - q)$$

Restricted LS – Properties

Restricted LS estimator: $b^* = b - (X'X)^{-1}R'\,[R(X'X)^{-1}R']^{-1}(Rb - q)$

$= b +$ correction

• Properties:

1. Unbiased?
   - Yes, if Theory is correct! $\quad E[b^* \mid X] = \beta$
   - No, if Theory is incorrect: $\quad E[b^* \mid X] \neq \beta$.
2. Efficiency?

   $\text{Var}[b^* \mid X] = \text{Var}[b \mid X] - \sigma^2 (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1} R(X'X)^{-1}$

   $\Rightarrow \text{Var}[b^* \mid X] < \text{Var}[b \mid X]$
3. A biased b^* may be more "precise," using metric MSE (=RSS/T), recall that MSE = $(\text{Bias})^2$ + Variance

Restricted LS – Interpretation

1. $\mathbf{b}^* = \mathbf{b} - \mathbf{Cm}$, $\mathbf{m}$ = the "discrepancy vector" $\mathbf{Rb} - \mathbf{q}$.
$\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}$

Note: If $\mathbf{m} = \mathbf{0} \Rightarrow \mathbf{b}^* = \mathbf{b}$.

2. We can show that RSS never decreases with restrictions:

$$\boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - \mathbf{Xb})'(\boldsymbol{y} - \mathbf{Xb}) \leq \boldsymbol{e}^{*\prime}\boldsymbol{e}^* = (\boldsymbol{y} - \mathbf{Xb}^*)'(\boldsymbol{y} - \mathbf{Xb}^*)$$

$\Rightarrow$ RSS $\leq$ RSS* -RSS* = RSS_R (Restricted RSS)

$\Rightarrow R^2 \geq R^{2*}$ (Recall, $R^2 = 1 -$ RSS/TSS)

That is, restrictions cannot increase R^2.

Restricted LS - Interpretation

- Implications of restrictions (Theory). Two cases:
  - Case 1: Theory is correct: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ (restrictions hold).
    $\mathbf{b}^*$ is unbiased & $\text{Var}[\mathbf{b}^* | \mathbf{X}] \leq \text{Var}[\mathbf{b} | \mathbf{X}]$
  - Case 2: Theory is incorrect: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} \neq \mathbf{0}$ (restrictions do not hold).
    $\mathbf{b}^*$ is biased & $\text{Var}[\mathbf{b}^* | \mathbf{X}] \leq \text{Var}[\mathbf{b} | \mathbf{X}]$.

- Interpretation

- The theory gives us information.
  Bad information produces bias (away from "the truth.")
  Any information, good or bad, makes us more certain of our answer. In this context, *any* information reduces variance.

Review – Significance Testing

• Fisher's *significance testing* procedure relies on the *p-value*: the probability of observing a result at least as extreme as the test statistic, under H_0.

• Fisher's Idea

1. Form H_0 & decide on a *significance level* (α%) to compare your test results.
2. Find *T(X)*. Know (or derive) the distribution of *T(X)* under H_0.
3. Collect a sample of data $X = \{x_1, x_2, \dots, x_T\}$.
   Compute the test-statistics *T(X)* used to test H_0 ⇒ Report its *p-value*.

4. Rule: If *p-value* < α (say, 5%) ⟹ test result is *significant:* Reject H_0.
   If the results are "*not significant*," no conclusions are reached (no learning here). Go back gather more data or modify model.

Review – Testing Only One Parameter

• We are interested in testing a hypothesis about one parameter in the linear model: $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

1. Set H_0 and H_1 (about only one parameter): H_0: $\beta_k = \beta_k^0$
   H_1: $\beta_k \neq \beta_k^0$

2. Appropriate T(*X*): *t-statistic*. Under H_0:
   If (**A5**) $t_k = (b_k - \beta_k^0)/s_{b,k}\,|\mathbf{X} \sim t_{T-k}$
   Otherwise, $t_k \xrightarrow{d} N(0,1)$

3. Compute t_k, $\hat{t}$, using b_k, β_k^0, s, and $(\mathbf{X'X})^{-1}$. Get *p-value*($\hat{t}$).

4. Rule: Set an α level. If *p-value*($\hat{t}$) < α ⇒ Reject H_0: $\beta_k = \beta_k^0$
   Alternatively, if $|\hat{t}| > \boldsymbol{t_{T-k,1-\alpha/2}}$ ⇒ Reject H_0: $\beta_k = \beta_k^0$.

Review – Testing Only One Parameter

- Special case: H_0: $\beta_k = 0$
  H_1: $\beta_k \neq 0$.

Then,

$$t_k = \frac{b_k}{\text{sqrt}\{s^2(\mathbf{X'X})^{-1}]_{kk}} = \frac{b_k}{\text{SE}[b_k]} = \textit{t-value or t-ratio.}$$

- Usual α levels and $t_{T-k,1-\alpha/2}$ –when $T > 30$, $t_{T-k,1-\alpha/2} \approx z_{1-\alpha/2}$

  $\alpha = 5\%$, then $z_{1-\alpha/2} = \mathbf{1.96}$ -in R, $z_{1-.05/2}$ = qnorm(0.975).

  $\alpha = 2\%$, then $z_{1-\alpha/2} = \mathbf{2.33}$ -in R, $z_{1-.02/2}$ = qnorm(0.99).

  $\alpha = 1\%$, then $z_{1-\alpha/2} = \mathbf{2.58}$ -in R, $z_{1-.01/2}$ = qnorm(0.995).

Testing: The Expectation Hypothesis (EH)

Example: EH states that forward/futures prices are good predictors of future spot rates: $E_t[S_{t+T}] = F_{t,T}$

Implication of EH: $S_{t+T} - F_{t,T}$ = unpredictable.

That is, $E_t[S_{t+T} - F_{t,T}] = E_t[\varepsilon_t] = 0$!

Empirical tests of the EH are based on a regression:

$$(S_{t+T} - F_{t,T})/S_t = \alpha + \beta Z_t + \varepsilon_t, \qquad \text{(where } E_t[\varepsilon_t] = 0)$$

where Z_t represents any economic variable that might have power to explain S_t, for example, interest rate differentials, $(i_d - i_f)$.

Then, under EH, H_0: $\alpha = 0$ and $\beta = 0$.
vs H_1: $\alpha \neq 0$ and/or $\beta \neq 0$.

Testing: The Expectation Hypothesis (EH)

Example (continuation): We will informally test EH using exchange rates (USD/GBP), 3-mo forward rates and 3-mo interest rates.

```
SF_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/SpFor_prices.csv",
head=TRUE, sep=",")
summary(SF_da)
x_date <- SF_da$Date
x_S <- SF_da$GBPSP
x_F3m <- SF_da$GBP3M
i_us3 <- SF_da$Dep_USD3M
i_uk3 <- SF_da$Dep_UKP3M
T <- length(x_S)
prem <- (x_S[-1] - x_F3m[-T])/x_S[-1]
int_dif <- (i_us3 - i_uk3)/100
y <- prem
x <- int_dif[-T]
fit_eh <- lm( y ~ x)
```

Testing: The Expectation Hypothesis (EH)

Example (continuation): We do two *individual* t-tests on α & β.

```
> summary(fit_eh)
Call:
lm(formula = y ~ x)

Residuals:
 Min       1Q    Median      3Q      Max
-0.125672 -0.014576 -0.000439  0.017356  0.094283

Coefficients:
              Estimate   Std. Error  t value Pr(>|t|)
(Intercept)  -0.0001854  0.0016219  -0.114  0.90906     ⇒ constant not significant (|t|<2)
x            -0.2157540  0.0731553  -2.949  0.00339 ** ⇒ slope is significant (|t|>2). ⇒ Reject H0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02661 on 361 degrees of freedom
Multiple R-squared:  0.02353,   Adjusted R-squared:  0.02082
F-statistic: 8.698 on 1 and 361 DF,  p-value: 0.003393
```

Testing: The Expectation Hypothesis (EH)

Example (continuation): 95% C.I. for β_k:

$C_n = [b_k \pm t_{T-k,1-\alpha/2} * \text{Estimated SE}(b_k)]$

Then,

$C_n = [\mathbf{-0.215754} - \mathbf{1.96} * \mathbf{0.0731553},\ \mathbf{-0.215754} + \mathbf{1.96} * \mathbf{0.0731553}]$

$= [\mathbf{-0.3591384}, \mathbf{-0.07236961}]$

Since $\beta = 0$ is not in C_n with 95% confidence $\Rightarrow$ Reject H_0: $\beta_1 = 0$ at 5% level.

Note: The EH is a **joint hypothesis**, it should be tested with a joint test!

Testing a Hypothesis: Wald Statistic

- Most of our test statistics, including joint tests, are Wald statistics.

Wald = normalized distance measure.

One parameter: $t_k = (b_k - \beta_k^0)/s_{b,k}$ = distance/unit

More than one parameter.

Let **z** = (random vector – hypothesized value) be the distance

$W = \mathbf{z}' [\text{Var}(\mathbf{z})]^{-1} \mathbf{z}$ -a quadratic form, produces a number

- Distribution of W? We have a quadratic form. Since σ^2 is unknown, we use s^2, then:

– If $\mathbf{z}$ is normal, $W \sim F_{J,T-k}$

– If $\mathbf{z}$ is not normal, we rely on asymptotic theory

$W \xrightarrow{d} \chi^2_{rank[Var(z)]}$

Abraham Wald (1902–1950, Hungary)

The General Linear Hypothesis: H_0: $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$

• Suppose we are interested in testing J joint hypotheses.

Example: We want to test that in the 3 FF factor model that the SMB and HML factors have the same coefficients, $\beta_{SMB} = \beta_{HML} = \beta^0$.

We can write linear restrictions as H_0: $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$, where $\mathbf{R}$ is a Jxk matrix and $\mathbf{q}$ a Jx1 vector.

In the above example (J=2), we write:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_{Mkt} \\ \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = \begin{bmatrix} \beta^0 \\ \beta^0 \end{bmatrix}$$

The General Linear Hypothesis: H_0: $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$

• Q: Is $\mathbf{Rb} - \mathbf{q}$ close to $\mathbf{0}$? There are two different approaches to this question. Both have in common the property of unbiasedness for $\mathbf{b}$.

Approach (1): We base the answer on the discrepancy vector:

$$\mathbf{m} = \mathbf{Rb} - \mathbf{q}.$$

Then, we construct a Wald statistic:

$$W = \mathbf{m}' \, (\text{Var}[\mathbf{m}|\mathbf{X}])^{-1} \, \mathbf{m}$$

to test if $\mathbf{m}$ is different from 0.

Approach (2): We base the answer on a model loss of fit when restrictions are imposed: RSS must increase and R^2 must go down. Then, we construct an F test to check if the unrestricted RSS (RSS_U) is different from the restricted RSS (RSS_R).

Wald Test Statistic for H_0: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$

Approach (1): Test H_0 with $W = \mathbf{m}' (\text{Var}[\mathbf{m}|\mathbf{X}])^{-1}\,\mathbf{m}$

Based on unrestricted OLS estimation we compute:

$\mathbf{m} = \mathbf{Rb} - \mathbf{q}$ (under (**A5**) & H_0: $\mathbf{m} \sim N(\mathbf{0}, \text{Var}[\mathbf{m}])$)

$\text{Var}[\mathbf{m}|\mathbf{X}] = \mathbf{R}\,[\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}]\,\mathbf{R}'$

Then, we compute the Wald statistic:

$$W = (\mathbf{Rb} - \mathbf{q})' \{\mathbf{R}[\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'\}^{-1} (\mathbf{Rb} - \mathbf{q})$$

Under H_0 and assuming (**A5**) & estimating σ^2 with $s^2 = \boldsymbol{e}'\boldsymbol{e}/(T-k)$:

$$W^* = (\mathbf{Rb} - \mathbf{q})' \{\mathbf{R}[s^2 (\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}\}^{-1} (\mathbf{Rb} - \mathbf{q})$$

$$F = W^*/J \sim F_{J,T-k}.$$

If (**A5**) is not assumed, the results are only asymptotic: $J * F \xrightarrow{d} \chi^2_J$

Wald Test Statistic for H_0: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$

- Under H_0, assuming (**A5**) & estimating σ^2 with $s^2 = \boldsymbol{e}'\boldsymbol{e}/(T-k)$:

$$W^* = (\mathbf{Rb} - \mathbf{q})' \{\mathbf{R}[s^2 (\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}\}^{-1} (\mathbf{Rb} - \mathbf{q})$$

$$F = W^*/J \sim F_{J,T-k}.$$

Technical note: Why the F distribution?

The F-distribution is a ratio of two independent χ^2_J and χ^2_T RV divided by their degrees of freedom: $F = \frac{\chi^2_J/J}{\chi^2_T/T} \sim F_{J,T}$

(1) Numerator: $W = (\mathbf{Rb} - \mathbf{q})' \{\mathbf{R}[\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'\}^{-1} (\mathbf{Rb} - \mathbf{q}) \sim \chi^2_J$

(2) Denominator: $(T-k) * s^2/\sigma^2 \sim \chi^2_{T-k}$

$$F = \frac{\chi^2_J/J}{\chi^2_T/T} = \frac{[(\mathbf{Rb} - \mathbf{q})' \{\mathbf{R}[\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'\}^{-1} (\mathbf{Rb} - \mathbf{q})\,]/J}{[(T-k)*s^2/\sigma^2]/(T-k)} \sim F_{J,T-k}.$$

Wald Test Statistic for H_0: $R\beta - q = 0$

Example: We test in the 3 FF factor model for IBM returns (T=569). Steps

1. H_0: $\beta_{SMB} = 0.2$ and $\beta_{HML} = 0.6$.
   H_1: $\beta_{SMB} \neq 0.2$ and/or $\beta_{HML} \neq 0.6$. $\Rightarrow J = 2$

We define **R** (2x4) below and write $\mathbf{m} = \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_{Mkt} \\ \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}$$

2. Test-statistic: $F = W^*/J = (\mathbf{Rb} - \mathbf{q})' \{\mathbf{R}[s^2 (\mathbf{X'X})^{-1}]\mathbf{R}'\}^{-1} (\mathbf{Rb} - \mathbf{q})$

   Distribution under H_0: Exact: $F = W^*/2 \sim F_{2,T-4}$

   Asymptotic: $2 * F \xrightarrow{d} \chi^2_2$

Wald Test Statistic for H_0: $R\beta - q = 0$

Example (continuation):

3. Get OLS results, compute F, $\hat{F}$.

4. <u>Decision Rule</u>: $\alpha = 0.05$ level. We reject H_0 if *p-value*($\hat{F}$) < .05.
   Or, reject H_0, if $\hat{F} > F_{J=2,T-4,.05}$.

Step 1. Define **R** (2x4) and **q**. write $\mathbf{m} = \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$:

```
J <- 2                                  # number of restriction
R <- matrix(c(0,0,0,0,1,0,0,1), nrow=2) # matrix of restrictions
q <- c(.2, .6)                          # hypothesized values
```

Step 3. Do OLS and compute compute F, $\hat{F}$.

```
fit_ibm_ff3 <- lm(ibm_x ~ Mkt_RF + SMB + HML)
b <- fit_ibm_ff3$coefficients           # Extract OLS coefficients
Var_b <- vcov(fit_ibm_ff3)              # Extract Var[b]
m <- R%*%b - q                          # m = Estimated R*Beta - q
```

Wald Test Statistic for H_0: $R\beta - q = 0$

Example (continuation):

Step 3. Do OLS and compute compute F, $\hat{F}$.

```
Var_m <- R %*% Var_b %*% t(R)          # Variance of m
det(Var_m)                             # check for non-singularity
W <- t(m)%*%solve(Var_m)%*%m           # W = m' Var[m] m
F_t <- as.numeric(W/J)                 # F-test statistic
> F_t
49.21676
F_t_asym <- as.numeric(J*F_t)          # Chi-square-test statistic (asymptotic)
> F_t_asym
98.433
```

Wald Test Statistic for H_0: $R\beta - q = 0$

Example (continuation):

Step 4. Decision rule.

```
qf(.95, df1=J, df2=(T - k))                  # exact distribution (F-dist) if e normal
[1] 3.011644                    F_t > 3.011644 ⇒ reject H0 at 5% level
p_val <- 1 - pf(F_t, df1=J, df2=(T - k))     # p-value(F_t) under e normal
[1] 0                           very low chance H0 is true.

> p_val <- 1 - pchisq(F_t_asym, df=J)        # p-value(F_t) under asymptotic distrib.
> p_val
[1] 0                           very low chance H0 is true.
```


Wald Test Statistic for H_0: $R\beta - q = 0$

Example (continuation): You can use the R package *car* to test linear restrictions (linear H_0).

```
install.packages("car")
library(car)
linearHypothesis(fit_ibm_ff3, c("SMB = 0.2","HML = 0.6"), test="F") # "F": exact test

Linear hypothesis test

Hypothesis:
SMB = 0.2
HML = 0.6

Model 1: restricted model
Model 2: ibm_x ~ Mkt_RF + SMB + HML

Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    567 2.2691
2    565 1.9324  2   0.33667 49.217 < 2.2e-16 ***
```

⇒ reject H_0 at 5% level

Wald Test Statistic for H_0: $R\beta - q = 0$

Example (continuation): The asymptotic test uses test="Chisq".

```
> linearHypothesis(fit_ibm_ff3, c("SMB = 0.2","HML = 0.6"), test="Chisq")     # Asymptotic F
Linear hypothesis test

Hypothesis:
SMB = 0.2
HML = 0.6

Model 1: restricted model
Model 2: ibm_x ~ Mkt_RF + SMB + HML

  Res.Df    RSS Df Sum of Sq  Chisq Pr(>Chisq)
1    567 2.2691
2    565 1.9324  2   0.33667 98.433  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

qf(.95, df1=J, df2=(T - k))                 # asymptotic distribution (Chi-square-distribution)
[1] 5.991465
```

⇒ reject H_0 at 5% level

F_t_asym > 5.991465 ⇒ reject H_0 at 5% level

Wald Test Statistic for H_0: Does EH hold?

Example: Now, we do a joint test of the EH. H_0: $\alpha = 0$ and $\beta = 0$.

Using the R car package, but with **fit_eh**:

```
> linearHypothesis(fit_eh,c("(Intercept) = 0","x = 0"), test="F")       # "F": exact test, with F-distrib
Linear hypothesis test

Hypothesis:
(Intercept) = 0
x = 0

Model 1: restricted model
Model 2: y ~ x

  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1    363 0.27033
2    361 0.26432  2 0.0060075 4.1024 0.01731 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 qf(.95, df1=J, df2=(T - k))                          # exact distribution (F-dist) if errors normal
[1] 3.020661                                          F_t > 3.020661 ⇒ reject H0 at 5% level
```

The F Test: H_0: $R\beta - q = 0$

Approach(2): We know that imposing the restrictions leads to a loss of fit. R^2 must go down. Does it go down a lot? –i.e., significantly?

Recall (i) $\boldsymbol{e}^* = (\boldsymbol{y} - \mathbf{Xb}^*) = \boldsymbol{y} + (\mathbf{Xb} - \mathbf{Xb}) - \mathbf{Xb}^* = \boldsymbol{e} - \mathbf{X}(\mathbf{b}^* - \mathbf{b})$

(ii) $\mathbf{b}^* = \mathbf{b} - (\mathbf{X'X})^{-1}\mathbf{R'}[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}]^{-1}(\mathbf{Rb} - \mathbf{q})$

$$\Rightarrow \quad \boldsymbol{e}^{*\prime}\boldsymbol{e}^* = \boldsymbol{e}'\boldsymbol{e} + (\mathbf{b}^* - \mathbf{b})'\,\mathbf{X'X}\,(\mathbf{b}^* - \mathbf{b})$$

Replacing $(\mathbf{b}^* - \mathbf{b})$ from (ii) in the above formula, we get:

$$\boldsymbol{e}^{*\prime}\boldsymbol{e}^* - \boldsymbol{e}'\boldsymbol{e} = (\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}]^{-1}(\mathbf{Rb} - \mathbf{q})$$

Note: $\boldsymbol{e}^{*\prime}\boldsymbol{e}^* - \boldsymbol{e}'\boldsymbol{e}$ is a quadratic form, then we can use a lot of results for quadratic forms to derive its asymptotic distribution.

- Recall, the F-distribution is a ratio of two independent χ^2_J and χ^2_T RV divided by their degrees of freedom: $F = \frac{\chi^2_J\,/J}{\chi^2_T\,/T} \sim F_{J,T}$

The F Test: H_0: $R\beta - q = 0$

Then, to get to the F-test, we rely on two results:

– $W = (\mathbf{Rb} - \mathbf{q})'\{\mathbf{R}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'\}^{-1}(\mathbf{Rb} - \mathbf{q}) \sim \chi^2_J$ (if σ^2 is known)

– $\boldsymbol{e}'\boldsymbol{e}/\ \sigma^2 \sim \chi^2_{T-k}$.

$$\Rightarrow F = \frac{(\boldsymbol{e}^{*\prime}\boldsymbol{e}^{*} - \boldsymbol{e}'\boldsymbol{e})/J}{[\boldsymbol{e}'\boldsymbol{e}/(T-k)]} \sim F_{J,T-k}.$$

- We can write the F-test in terms of R^2's. Let

  R^2 = unrestricted model = 1 – RSS/TSS

  R^{*2} = restricted model fit = 1 – RSS*/TSS

Then, dividing and multiplying F by TSS we get:

$$F = \frac{(1-R^{*2}) - (1-R^2)/J}{(1-R^2)/(T-k)} \sim F_{J,T-k}$$

or

$$F = \frac{(R^2 - R^{*2})/J}{(1-R^2)/(T-k)} \sim F_{J,T-k}.$$

The F Test: H_0: F-test of Goodness of Fit

- In the linear model, with a constant ($\mathbf{X}_1 = \boldsymbol{i}$):

$$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon = \beta_1 + \mathbf{X}_2\,\beta_2 + \mathbf{X}_3\,\beta_3 + \ldots + \mathbf{X}_k\,\beta_k + \varepsilon$$

- We want to test if the slopes of $\mathbf{X}_2, \ldots, \mathbf{X}_k$ are equal to zero. That is,

  H_0: $\beta_2 = \cdots = \beta_k = 0$

  H_1: at least one $\beta_k \neq 0$ $\Rightarrow J = k - 1$

- We can write H_0: $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ $\Rightarrow \begin{bmatrix} 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ \ldots \\ 0 \end{bmatrix}$

- We have $J = k - 1$. Then,

$$F = \{\ (R^2 - R^{2*})/(k-1)\ \}\ /\ [(1 - R^2)/(T-k)] \sim F_{k-1,T-k}.$$

- For the restricted model, $R^{2*} = 0$.

10

The F Test: H_0: F-test of Goodness of Fit

Then, $F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)} \sim F_{k-1,T-k}$.

- Recall ESS/TSS = R^2 & RSS/TSS = $(1 - R^2)$, we compute F:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)} = \frac{\frac{ESS}{TSS}/(k-1)}{\frac{RSS}{TSS}/(T-k)}$$

$$F = \frac{ESS/(k-1)}{RSS/(T-k)}$$

- This test statistic is called the *F-test of goodness of fit*. It is reported in all regression packages as part of the regression output. In R, the lm function reports it as "*F-statistic*."

10

The F Test: H_0: F-test of Goodness of Fit

Example: We want to test if all the FF factors (Market, SMB, HML) are significant (J=3), using monthly data 1973 – 2020 (T=569).

```
T <- length(ibm_x)
k <- 4
e <- fit_ibm_ff3$residuals                    # Extract residuals
y <- ibm_x - mean(ibm_x)
RSS <- sum(e^2)
R2 <-  1 - RSS/sum(y^2)              #R-squared
> R2
[1] 0.338985
> F_goodfit <- (R2/(k-1))/((1-R2)/(T-k))                 # F-test of goodness of fit.
> F_goodfit
[1] 96.58204
```

⇒ F_goodfit > $F_{3,565,.05}$ = **2.62068** ⇒ Reject H_0.

10

The F Test: General Case – Example

- In the linear model

$$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon = \beta_1 + \mathbf{X}_2\,\beta_2 + \mathbf{X}_3\,\beta_3 + \mathbf{X}_4\,\beta_4 + \varepsilon$$

- We want to test if the slopes $\mathbf{X}_3, \mathbf{X}_4$ are equal to zero. That is,

H_0: $\beta_3 = \beta_4 = \mathbf{0}$

H_1: $\beta_3 \neq \mathbf{0}$ or $\beta_4 \neq \mathbf{0}$ or both β_3 and $\beta_4 \neq \mathbf{0}$

- We can use, $F = (\boldsymbol{e}^{*\prime}\boldsymbol{e}^{*} - \boldsymbol{e}'\boldsymbol{e})/J \,/\, [\boldsymbol{e}'\boldsymbol{e}\,/(T-k)] \sim F_{J,T-k}$.

- Define
$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon = \beta_1 + \mathbf{X}_2\,\beta_2 + \varepsilon$ $\quad(RSS_R)$
$\mathbf{y} = \beta_1 + \mathbf{X}_2\,\beta_2 + \mathbf{X}_3\,\beta_3 + \mathbf{X}_4\,\beta_4 + \varepsilon$ $\quad(RSS_U)$

$$F(k_U - k_R, T-k) = \frac{\frac{RSS_R - RSS_U}{(k_U - k_R)}}{\frac{RSS_U}{(T-k_U)}}$$

32

The F Test: Are SMB and HML Priced Factors?

Example: We want to test if the additional FF factors (SMB, HML) are significant, using monthly data 1973 – 2020 (T=569).

Unrestricted Model:

(U) $(r_{IBM,t} - r_f) = \beta_0 + \beta_1\,(r_{m,t} - r_f) + \beta_2\,SMB_t + \beta_3\,HML_t + \varepsilon_t$

Hypothesis: H_0: $\beta_2 = \beta_3 = 0$

H_1: $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$

Then, the Restricted Model:

(R) $(r_{IBM,t} - r_f) = \beta_0 + \beta_1\,(r_{m,t} - r_f) + \varepsilon_t$

Test: $F = \frac{(RSS_R - RSS_U)/J}{RSS_U/(T-k_u)} \sim F_{J,T-k}$, with $J = k_U - k_R = 4 - 2 = 2$

The F Test: Are SMB and HML Priced Factors?

Example (continuation): The unrestricted 3-factor FF model was already estimated (**fit_ibm_ff3**). Same for the restricted model (**fit_ibm_capm**):

```
e_u <- fit_ibm_ff3$residuals          # Unrestricted residuals
e_r <- fit_ibm_capm$residuals         # Restricted residuals
T <- length(ibm_x)
k <- 4
k_r <- 2

RSS <- sum(e_u^2)                     # RSS_U
RSS_r <- sum(e_r^2)                   # RSS_R
> RSS = 1.932442                      > RSS2 = 1.964844

J <- k – k_r                          # J = degrees of freedom numerator
F_test <- ((RSS_r - RSS)/J)/(RSS/(T-k))
```

The F Test: Are SMB and HML Priced Factors?

Example (continuation):

```
> F_test <- ((RSS2 - RSS)/J)/(RSS/(T - k))
> F_test
[1] 4.736834
> qf(.95, df1=J, df2=(T-k))                     # F_{2,565,.05} value (≈ 3)
[1] 3.011672                                    ⇒ Reject H0.
> p_val <- 1 - pf(F_test, df1=J, df2=(T-k))     # p-value of F_test
> p_val
[1] 0.009117494                                 ⇒ p-value is small ⇒ Reject H0.
```

The F Test: Are SMB and HML Priced Factors?

Example (continuation):

There is package in R, *lmtest,* that performs this test, *waldtest,* (and many others, used in this class). You need to install it first.

Note: The models need to be nested. For the *waldtest*, the default reports the *F-test* with the F distribution.

```
library(lmtest)
fit_wU <- lm (ibm_x ~ Mkt_RF + SMB + HML)
fit_wR <- lm (ibm_x ~ Mkt_RF)
waldtest(fit_wU, fit_wR)

Wald test

Model 1: ibm_x ~ Mkt_RF + SMB + HML
Model 2: ibm_x ~ Mkt_RF
  Res.Df Df      F   Pr(>F)
1    565
2    567 -2 4.7368 0.009117 **
```

⇒ p-value is small: Reject H_0