Given a set of estimates, we need to judge our confidence in

- ▶ estimates
- ▶ model
- ▶ Simple examples:
  - ▶ Univariate regressions, confidence intervals, testing
  - ▶ Multivariate regressions, confidence intervals, testing
- ▶ Confidence interval estimation
- ▶ Confidence region estimation
- ▶ Hypothesis tests, the trinity (in film parlance: the trilogy)
  - ▶ **Wald principle**: Estimate the model without imposing any restrictions, and then see *how far* the estimated parameters are from fulfilling the restrictions implied in the hypothesis.
  - ▶ **Lagrange multiplier principle**: Estimate the model *imposing* the restrictions, and then test how much relaxing the restrictions *improves* the fit of the model.
  - ▶ **Likelihood Ratio principle**: Estimate *both* the restricted and the unrestricted forms of the model, and then *compare the fit* of the two estimates.

Inference in (univariate) Regressions

The regression

$$\tilde{y} = a + bx + \tilde{e}$$

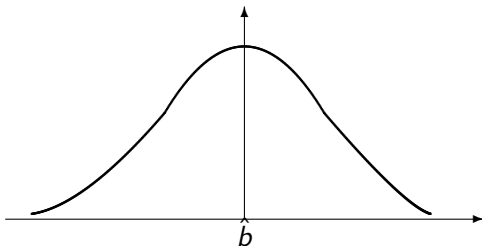Uncertainty centered in error term

$$\tilde{e} \sim (0, \sigma^2)$$

Other variables

$$a, b, x$$

think of as constants.

Statistical goal
Make probability statements about estimated parameters, e.g. $\hat{b}$.

To solve statistical problem:
Need estimates of

- ▶ Parameters
- ▶ Parameter standard deviations / standard errors.

Calculating standard error estimates

Function of uncertainty of $\tilde{e}$

$$\text{var}(e) = E[(e - E[e])^2]$$

Since $E[e] = 0$

$$\text{var}(e) = E[e^2]$$

Estimate from sample

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - (a + bx_i))^2$$

Given S^2, find

$$s_a = S\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$s_b = S\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Alternative calculation of standard errors.

$$s_a = \sqrt{\frac{1}{n-2} \frac{\left(\sum_{i=1}^n e_i^2\right)\left(\sum_{i=1}^n x_i^2\right)}{n\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$s_b = \sqrt{\frac{1}{n-2} \frac{\left(\sum_{i=1}^n e_i^2\right)}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Estimated standard errors s_a and s_b:
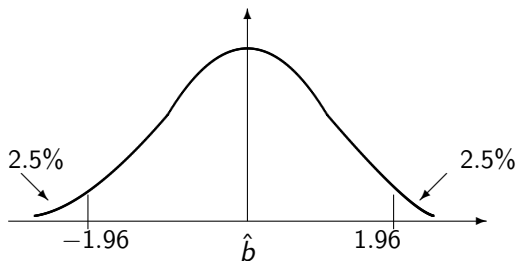Basis for inference.
Careful: Only one parameter at the time.

Confidence interval construction

Construct confidence intervals for a and b (one parameter at the time)

$$(\hat{a} - dS_a, \hat{a} + ds_a)$$

$$(\hat{b} - dS_b, \hat{b} + ds_b)$$

The constant d is a number that depends on the confidence level. Usual example, $\alpha = 5\%$

Normal distribution: $d = 1.96$.

Testing hypotheses about single parameters

Will test hypothesis of the types

$$a = a_0$$

and

$$b = b_0$$

Construct test statistics

$$z_a = \frac{\hat{a} - a_0}{S_a}$$

and

$$z_b = \frac{\hat{b} - b_0}{S_b}$$

Under the null mean zero unit variance.

The F test

Tests whether the nonconstant term of the regression is zero
Compares the two regression formulations

$$y = a + bx + e$$

to

$$y = a + e$$

In the univariate regression case the F statistic is calculated as

$$F = (n - 2)\frac{R^2}{1 - R^2}$$

Follows an F distribution $F(1, n)$.

Exercise

You want to apply the CAPM. The CAPM can be written as

$$E[r_{it}] - r_{ft} = \beta_i(E[r_{mt}] - r_{ft}$$

We rewrite this in excess return form, defining $er_{it} = r_{it} - r_{ft}$ and $er_{mt} = r_{mt} - r_{ft}$. The CAPM thus imposes

$$E[er_{it}] = \beta_{it}E[er_{mt}]$$

Consider now the regression

$$er_{it} = \alpha_i + \beta_i er_{mt} + \varepsilon_{it} \tag{1}$$

▶ Estimate this regression using historical data on the returns on IBM and a stock market index. Use monthly returns. Take data for 1995:1 to 2006:12.

▶ Discuss the fit of the model using the R^2 of the regression

▶ What restriction on the parameters of (1) does the CAPM impose? Test this restriction.

Use R to implement the estimations.

the input file contains data aligned on observation date.
The interest rate is observed looking forward, while returns are
looking backward from observation date.
Risk free rate is in annualzed percentages

```
> rets <- read.table("../data_set/ibm_sp500_95_06.txt",
                      header=TRUE,sep=";")
> n    <- dim(rets)[1]
> rf   <- rets[1:n-1,4]/1200
> rm   <- rets[2:n,2]
> ibm  <- rets[2:n,3]
```
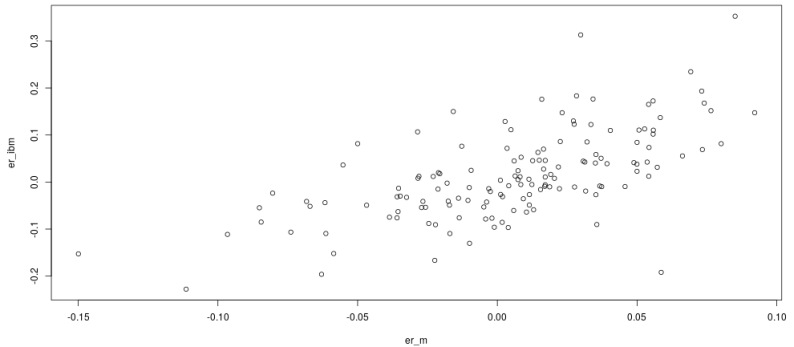
Calculating execess returns

$$er_{IBM,t} = r_{IBM,t} - r_{ft}$$

$$er_{m,t} = r_{m,t} - r_{ft}$$

```
> er_ibm <- ibm-rf
> er_m   <- rm-rf
```

Plotting IBM return against the market

Running the regression

$$er_i bm = a + ber_m + e$$

```
> mod=lm(er_ibm ~ er_m)
```

Showing the resulting estimates

```
> summary(mod)

Call:
lm(formula = er_ibm ~ er_m)

Residuals:
      Min       1Q    Median       3Q      Max
-0.280281 -0.042710 -0.007594  0.037619  0.264997

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.005879   0.006139   0.958     0.34
er_m        1.406461   0.145848   9.643   <2e-16 ***

Residual standard error: 0.07279 on 141 degrees of freedom
Multiple R-squared: 0.3974,Adjusted R-squared: 0.3931
F-statistic: 92.99 on 1 and 141 DF,  p-value: < 2.2e-16
```
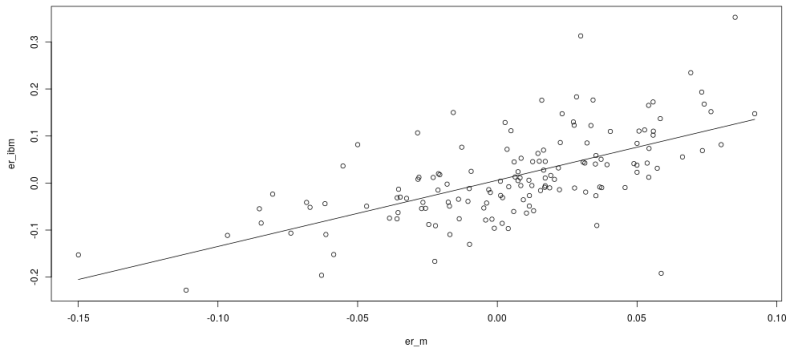
We estimate the parameters as

$$\hat{a} = 0.005$$

$$\hat{b} = 1.407$$

Next, plotting the observations and comparing it to the actual regrssion.
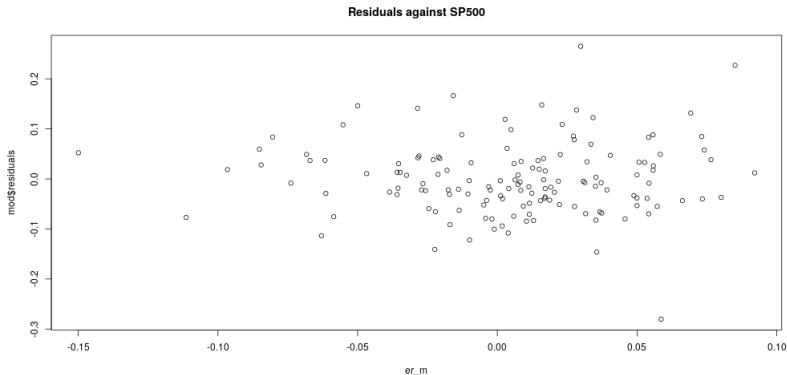
```
> plot(er_m,er_ibm,main="Regressing IBM on SP500")
> lines (er_m,mod$fitted.values)
```

Regressing IBM on SP500

Check for any obvious problems by calculating the residuals and plotting them against r_m:

```
> plot(er_m,mod$residuals,main="Residuals against SP500")
```


Residuals against SP500

What restrictions do the CAPM impose on the parameters?

That $a = 0$.

Set up a hypotesis test:
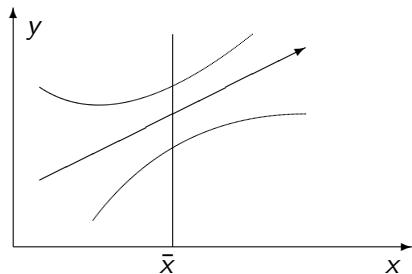
$$H_0 : \ a = a_0 = 0$$

$$H_A : \ a \neq 0$$

Here we can use the p-value in the R output:

The intercept, estimated as 0.005879, has a p-value of 0.34. We can thus not reject the null of a zero intercept.

Prediction of y

Confidence bands for prediction of y

The uncertainty of prediction increasing with distance from data used to generate estimates.

Summarizing

$$\tilde{y} = a + bx + \tilde{e}$$

Uncertainty: from error term

$$\tilde{e} \sim (0, \sigma^2)$$

$$\text{var}(e) = E[e^2]$$

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

$$s_a = S \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

$$s_b = S \sqrt{\frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Summarizing ctd

Confidence intervals

$$(\hat{a} - dS_a, \hat{a} + ds_a)$$

$$(\hat{b} - dS_b, \hat{b} + ds_b)$$

d depend on confidence level

Hypothesis test (of the form)

$$a = a_0$$

Construct test statistics

$$z_a = \frac{\hat{a} - a_0}{S_a}$$

$$z_b = \frac{\hat{b} - b_0}{S_b}$$

$$P(|z| > d) = \alpha$$

$$z_a \sim N(0, 1)$$

Summarizing ctd

Testing for whether x explains y

$$F = (n - 2)\frac{R^2}{1 - R^2}$$

Inference in multivariate regressions

Output overview
Typical output from regression:
Each explanatory variable

- Coefficient estimates
- Standard errors
- t-values
- p-values
    - Under a normal distribution
    - Under a t-distribution

The whole regression

- R^2 and Adjusted R^2.
- F test (with p-value)

Exercise

You want to measure the risk of an equity position on IBM. In a series of papers Eugene Fama and Ken French has introduced an alternative asset pricing model to the CAPM, the three factor model, where expected returns follows:

$$E[r_{it}] - r_{ft} = \beta_1(E[r_{mt}] - r_{ft} + b_2 SMB_t + b_3 HML_t$$

Defining $er_{it} = r_{it} - r_{ft}$ and $er_{mt} = r_{mt} - r_{ft}$, we consider the regression

$$er_{it} = \alpha_i + b_1 er_{mt} + b_2 SMB_t + b_3 HML_t + \varepsilon_{it} \qquad (2)$$

You have access to the time series of these variables on the course homepage. The data is for the period 1995:1 to 2005:12.

▶ Estimate this regression using historical data on the returns on IBM.

▶ Discuss the fit of the model using the R^2 of the regression.

▶ Which of the tree factors seem important for pricing of IBM?

Use R to implement the estimations.

Solution
Reading the data:

```
> rets <- read.table("../data_set/ibm_ff_95_05.txt",header=
> erIBM <- rets$erIBM
> RMRF <- rets$RMRF/100
> HML <- rets$HML/100
> SMB <- rets$SMB/100
```

```
Doing the analysis:

> mod <- lm( erIBM ~ RMRF + HML + SMB)


Call:
lm(formula = erIBM ~ RMRF + HML + SMB)


Residuals:
      Min        1Q    Median        3Q       Max
-0.306521 -0.047391 -0.007928  0.039860  0.267676


Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.007814   0.006986   1.118   0.2655
RMRF         1.170170   0.178451   6.557 1.22e-09 ***
HML         -0.422198   0.238618  -1.769   0.0792 .
SMB         -0.381568   0.184886  -2.064   0.0411 *


Residual standard error: 0.07642 on 128 degrees of freedom
Multiple R-squared: 0.387,Adjusted R-squared: 0.3726
F-statistic: 26.93 on 3 and 128 DF,  p-value: 1.433e-13
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0078	0.0070	1.12	0.2655
RMRF	1.1702	0.1785	6.56	0.0000
HML	-0.4222	0.2386	-1.77	0.0792
SMB	-0.3816	0.1849	-2.06	0.0411

Do the usual diagnostics, plot residuals against explanatory variables

```
> plot (mod$residuals,RMRF)
> plot (mod$residuals,HML)
> plot (mod$residuals,SMB)
```

Residual plot RMRF

Residual plot SMB

SMB

mod$residuals
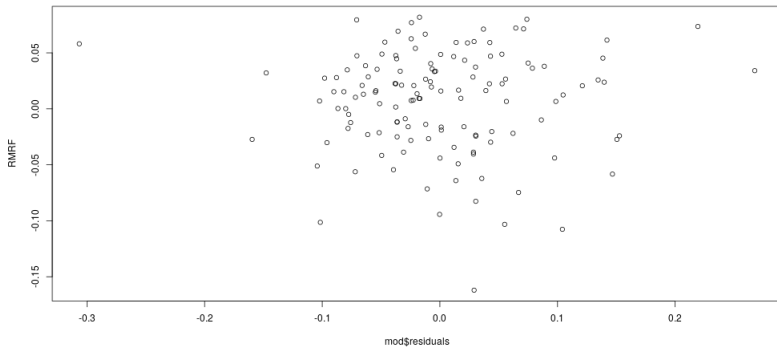
Residual plot HML
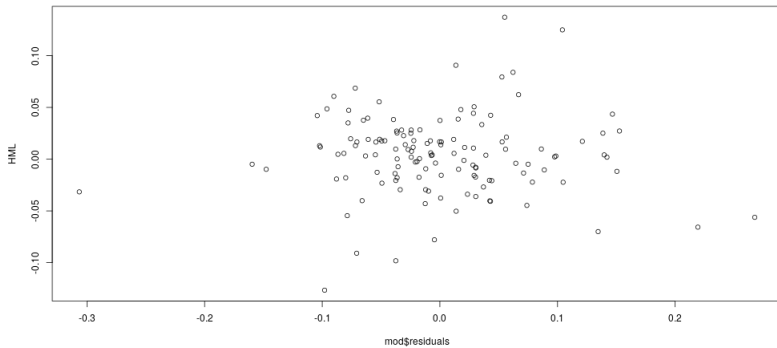
Observe the R^2 of the regression. $R^2 = 38.7\%$.

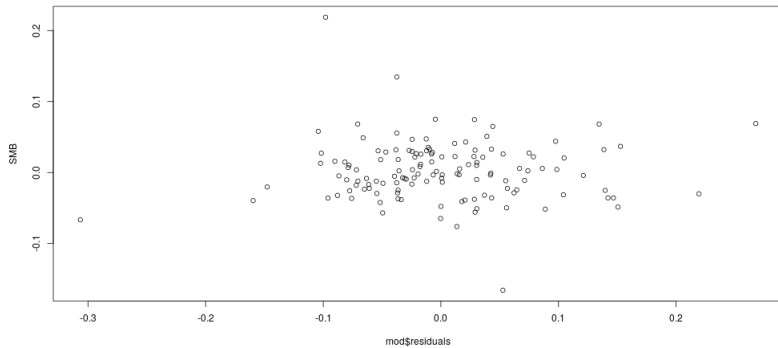To say something about what are the important factors relevant for the pricing of IBM, use the estimates of the standard deviations of the various parameters.

The t-values are the basis for most inference, and they are typically printed out by standard econometric packages.

Another typical way of reporting this is by finding the probability values

In terms of the results find that RMRF very significant. SMB significant at the 5% level, with a p value of 4.1%, while HML not significant, p value of 7.92%.

R^2

In the univariate regression we calculate

$$SSR = \sum_{i=1}^{n}(y_i - (a + bx_i))^2$$

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

and find

$$R^2 = 1 - \frac{SSR}{TSS}$$

Analogous calculation in the multivarate regression

$$SSR = \sum_{i=1}^{n}(y_i - (a + b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + \ldots + b_kx_{ki}))^2$$

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

and find

$$R^2 = 1 - \frac{SSR}{TSS}$$

The Adjusted R^2

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right)$$

F test

For a multivariate regression

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + e$$

The F test compares this to

$$y = a + e$$

i.e. it is a joint test of the null

$$H_0 : b_1 = b_2 = \ldots b_k = 0$$

"The" F test

For a multivariate regression

$$y = a + b_1 x_1 + b_2 x_2 + ... + b_k x_k + e$$

The F test compares this to

$$y = a + e$$

Joint test of the null

$$H_0 : b_1 = b_2 = \ldots b_k = 0$$

Calculated as

$$F = \frac{(n-k)}{r} \frac{R^2}{1 - R^2}$$

Exercise

You consider pricing of IBM stock using the Fama French three factor model.

$$er_{ibm,t} = a + b_1 er_{m,t} + b_2 SMB_t + b_3 HML_t + e_t$$

Test whether the thee explanatory factors $er_{m,t}$, SMB_t and HML_t has any explanatory power, i.e, test the hypothesis $b_1 = 0$, $b_2 = 0$ and $b_3 = 0$ jointly. Use monthly returns data on IBM and data for the three factors from the course homepage. The data is for 1995:1 to 2005:12. Use R to implement the estimations.

Solution This is often termed *the* F test, and is the result which is always printed in regression outputs, It is calculated as

$$F = \frac{(n-k)}{r} \frac{R^2}{1-R^2}$$

It has an F distribution $F(r, n-k)$.

```
Printout of the regression estimation in R:

> mod2 <- lm( erIBM ~ RMRF + HML + SMB )

Residuals:
      Min       1Q    Median       3Q       Max
-0.306521 -0.047391 -0.007928  0.039860  0.267676

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.007814   0.006986   1.118   0.2655
RMRF         1.170170   0.178451   6.557 1.22e-09 ***
HML         -0.422198   0.238618  -1.769   0.0792 .
SMB         -0.381568   0.184886  -2.064   0.0411 *

Residual standard error: 0.07642 on 128 degrees of freedom
Multiple R-squared: 0.387,Adjusted R-squared: 0.3726
F-statistic: 26.93 on 3 and 128 DF,  p-value: 1.433e-13
```

The F statistic in this regression is exactly the test we want. It is a test of the null that all coefficients except the intercept is zero.

The P-value of the F test is essentially zero. We reject the null of all coefficients being zero.

Note that this F statistic is *only* usable when we test whether *all* the coefficients are *jointly* zero. It can not be used when testing whether only some of the coeffients are zero.

Hypothesis testing in regressions

Testing hypotheses involving more than one parameter.
Example: Two alternative specifications

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + e \tag{3}$$

and

$$y = a + b_1 x_1 + b_2 x_2 + e \tag{4}$$

Testing whether (4) is sufficient:
Joint hypothesis that $b_3 = b_4 = 0$.

There are three general approaches to implementing hypotheses tests.

1. Wald principle: Estimate the unrestricted model, then ask how much it "hurts" to impose the null (in the example setting $b_3 = b_4 = 0$.)

2. Lagrange Multiplier principle: Estimate the restricted model, then ask how much we would "gain" if we allowed the parameters to move away from the null.

3. Likelihood Ratio principle: Estimate both the restricted and unrestriced models. Then compare the fit of the two models.

Likelihood ratio priciple:
Requires estimation of two regressions, one restricted and one unrestricted.
Calculate

$$LR = \frac{\frac{1}{r}\left(SSE\left(\tilde{b}\right) - SSE\left(\hat{b}\right)\right)}{\frac{1}{n-k}SSE\left(\hat{b}\right)}$$

where $\tilde{b}$ is the restricted estimate and $\hat{b}$ the unrestricted, r the number of restrictions, k the number of parameters, and n the number of observations.
$LR \sim F(r, n-k)$ in small samples, and $LR \sim \chi^2(r)$ in large samples.

In the example

$$r = 2$$

$$k = 5$$

$$\tilde{b} = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix}$$

$$\hat{b} = \begin{bmatrix} a \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

Exercise

You consider pricing of IBM stock. You have two alternative models in mind. One is a single factor model

$$er_{ibm,t} = \alpha + \beta er_{m,t} + e_t$$

where the excess return of IBM ($er_{ibm,t}$ only depends on one factor, the excess return on a market portfolio $er_{m,t}$. The other model you have in mind is Fama And French'es three factor model

$$er_{ibm,t} = a + b_1 er_{m,t} + b_2 SMB_t + b_3 HML_t + e_t$$

Using monthly returns data on IBM and data for the three factors from the course homepage, estimate the two models. Note that data is for 1995:1 to 2005:12.

Compare the R^2 of the two models.

Exercise ctd You wonder if the first model is "good enough", and you want to formally test whether the one factor model is a sufficient description of IBM's returns.

Do this as a hypothesis test, where the null hypothesis is that the first model is enough.

That is, you want to test $H_0 : b_2 = 0$ *and* $b_3 = 0$.

(Hint: This is a joint test, which can not be solved by looking at the coefficients b_1 and b_2 in isolation.)

Solution

```
> library(xtable)
> library(car)
> rets <- read.table("../data_set/ibm_ff_95_05.txt",
                      header=TRUE,sep=";")
> erIBM <- rets$erIBM
> RMRF <- rets$RMRF/100
> HML <- rets$HML/100
> SMB <- rets$SMB/100
```

```
> mod1 <- lm( erIBM ~ RMRF )
> summary(mod1)

Call:
lm(formula = erIBM ~ RMRF)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27138 -0.04877 -0.01114  0.04386  0.26441

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.004327   0.006810   0.635    0.526
RMRF        1.293020   0.150306   8.603 2.18e-14 ***

Residual standard error: 0.07731 on 130 degrees of freedom
Multiple R-squared: 0.3628,Adjusted R-squared: 0.3579
F-statistic:     74 on 1 and 130 DF,  p-value: 2.18e-14
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0043	0.0068	0.64	0.5263
RMRF	1.2930	0.1503	8.60	0.0000

```
> mod2 <- lm( erIBM ~ RMRF + HML + SMB )

Residuals:
     Min       1Q   Median       3Q      Max
-0.306521 -0.047391 -0.007928  0.039860  0.267676

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.007814   0.006986   1.118   0.2655
RMRF         1.170170   0.178451   6.557 1.22e-09 ***
HML         -0.422198   0.238618  -1.769   0.0792 .
SMB         -0.381568   0.184886  -2.064   0.0411 *

Residual standard error: 0.07642 on 128 degrees of freedom
Multiple R-squared: 0.387,Adjusted R-squared: 0.3726
F-statistic: 26.93 on 3 and 128 DF,  p-value: 1.433e-13
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0078	0.0070	1.12	0.2655
RMRF	1.1702	0.1785	6.56	0.0000
HML	-0.4222	0.2386	-1.77	0.0792
SMB	-0.3816	0.1849	-2.06	0.0411

Now the hypothesis test. Need to specify it in matrix form:

$$b_2 = 0$$

$$b_3 = 0$$

should be written as

$$Rb = r$$

where R and r are matrices.
In this case we use

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

```
> R = rbind(c(0,0,1,0),c(0,0,0,1))
> r=c(0,0)
> R = rbind(c(0,0,1,0),c(0,0,0,1))
> r=c(0,0)
> print(R)
     [,1] [,2] [,3] [,4]
[1,]    0    0    1    0
[2,]    0    0    0    1
> print(r)
[1] 0 0
```

And now finally performing the test:

```
> linearHypothesis(mod2,R,r)
Linear hypothesis test

Hypothesis:
HML = 0
SMB = 0

Model 1: restricted model
Model 2: erIBM ~ RMRF + HML + SMB

  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1    130 0.77693
2    128 0.74743  2  0.029508 2.5267 0.0839 .
```

At the 5% level we do not reject the null, and accept the one factor model as a sufficient explanation.

Inference - general

More generally.
classical OLS:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e},$$

and $\mathbf{e}$ is distributed normally with var$(\mathbf{e}) = \sigma_0 \mathbf{I}$
we found that

$$\widehat{\mathbf{b}}^{ols} \sim \mathcal{N}(\mathbf{b}, \sigma_0^2 (\mathbf{X}'\mathbf{X})^{-1})$$

Use this "classical case,"

Under less stringent assumptions

$$\sqrt{T}\widehat{\mathbf{b}}^{ols} \overset{D}{\to} \mathcal{N}(\mathbf{b}, \mathbf{M}^{-1}\mathbf{V}_T\mathbf{M}^{-1})$$

where

$$\mathbf{M} = E[\mathbf{X}_t\mathbf{X}_t']$$

$$\mathbf{V}_T = \text{var}\left(\sum_{t=1}^{T} x_t u_t\right),$$

Intuitively

$\mathbf{M}$: some kind of average of the x values.

$\mathbf{V}$: some kind of average of the covariance matrix of $\sum_{t=1}^{T} x_t u_t$.

In general, we estimate covariance matrices from

$$\mathbf{D}_T = \mathbf{M}^{-1}\mathbf{V}_T\mathbf{M}^{-1}$$

Confidence in parameter estimates

Two main techniques:
confidence intervals/regions
hypothesis testing.

Construction of confidence interval

Basic statistics: confidence intervals for a parameter such as a sample mean.

Similar for a coefficient in a regression.

α is the confidence level.

We assess the probability that the confidence interval covers the true parameters as $(1 - \alpha)$.

$$\widehat{\mathbf{b}}^{ols} \sim \mathcal{N}\left(b, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right) = \mathcal{N}\left(b, \widehat{\mathbf{V}}\right)$$

The standard deviation of one parameter, on diagonal of $\widehat{\mathbf{V}}$.

With two parameters b_1 and b_2

$$\widehat{\mathbf{V}} = \left[\begin{array}{cc} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 \end{array}\right],$$

and we would use the diagonal element to construct one parameter confidence intervals,
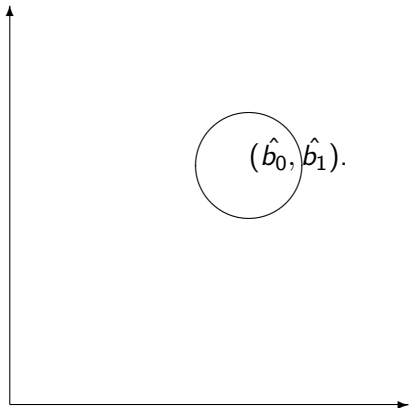
$$[b_1 - c\hat{\sigma}_1, b_1 + c\hat{\sigma}_1]$$

c is chosen based on confidence level.

A α percent *confidence region* - similar interpretation.
We look at two or more parameters jointly, and evaluate the
probability that the confidence region covers the true parameters.

CONFIDENCE REGIONS

$(\hat{b_0}, \hat{b_1}).$

Confidence regions summarize the probability of being in a *region*. We use the fact that the product of squared normals are chi-squared distributed to construct univariate random variables from the multivariate distributed $\widehat{\mathbf{b}}$,

$$\widehat{\mathbf{b}} \sim \mathcal{N}(b, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

Let $\mathbf{V}$ be the parameter covariance matrix (either exact or asymptotic). The quantity

$$(\widehat{\mathbf{b}} - \mathbf{b})'\mathbf{V}^{-1}(\widehat{\mathbf{b}} - \mathbf{b})$$

is chi-square distributed. To construct the confidence region, would look at what values of $\mathbf{b}$ would make this chi-square statistic less than some critical value.

Test of hypotheses

An alternative to checking whether a variable or a combination of variables lies in a confidence region is to construct hypothesis, and construct a test statistic that checks whether the data supports the hypothesis.

Exercise 1.

Suppose we have a time series of observations of variables x and y, indexed by time t:

$$\{(x_t, y_t)\}_{t=1}^{T}$$

Suppose we want to know whether we can use past observations of x and y to predict y.

1. Write down a linear regression with the current value of y, y_t as the dependent variable and the previous two periods observations of x and y as explanatory variables.

2. If our null hypothesis is that past values of x has no predictive power for y, write down the *null hypothesis* you will test.

Solution

1. The linear regression on lagged variables

   $$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 x_{t-1} \beta_2 x_{t-2} + \epsilon_t$$

   In this formulation we have used two *lags* (past observations). We can of course use more.

2. If our null hypothesis is that x has no predictive power for y, we would write a *null hypothesis*

   $$\mathcal{H}_0 : \ \beta_1 = \beta_2 = 0$$

   This particular test is one way to implement what is called the test of *Granger Causality*. A variable (x in this case) is said to *Granger Cause* another variable (y in this case) if we reject the null hypothesis, otherwise we conclude that x does not Granger cause y.

The three classical test statistics.

- ▶ **Wald principle**: Estimate the model without imposing any restrictions, and then see *how far* the estimated parameters are from fulfilling the restrictions implied in the hypothesis.
- ▶ **Lagrange multiplier principle**: Estimate the model *imposing* the restrictions, and then test how much relaxing the restrictions *improves* the fit of the model.
- ▶ **Likelihood Ratio principle**: Estimate *both* the restricted and the unrestricted forms of the model, and then *compare the fit* of the two estimates.

Comments

- ▶ Asymptotically, all these test statistics are equivalent, i.e. as the sample size increases, they will all give the same answers.
- ▶ In finite samples, the test statistics will have different values.
- ▶ It is advised to calculate all three of these statistics in applications, if it is feasible. (The Likelihood Ratio is often "better," but that is also the most expensive in terms of computing)
- ▶ If you have to choose between them, this choice is usually made from practical considerations. It is often easier to estimate either the restricted or the unrestricted model, and the choice of test statistic will reflect this.
    - ▶ If the restricted model is easier to estimate, use the LM test.
    - ▶ If the unrestricted model is easier, use the Wald test.

Construction of the Wald test

Remember the result we used in constructing the confidence region, that a quadratic form in unit normal variables has a chi-square distribution, so if $v \sim \mathcal{N}(0, \Omega)$ then $v'\Omega^{-1}v \sim \chi^2$. Hence, if

$$\widehat{\mathbf{b}} \sim \mathcal{N}(\mathbf{b}, \mathbf{V}),$$

then

$$(\widehat{\mathbf{b}} - \mathbf{b}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}),$$

and

$$(\widehat{\mathbf{b}} - \mathbf{b})'\mathbf{V}^{-1}(\widehat{\mathbf{b}} - \mathbf{b}) \sim \chi^2(k)$$

if $\mathbf{b}$ is a k-vector.

When we are testing a hypothesis, we impose restrictions on the parameters that will reduce the dimensionality of the test statistic.

We start by writing the restrictions on the form

$$R(\mathbf{b}) = 0,$$

where $R(\cdot)$ is some (maybe nonlinear) function of the parameters $\mathbf{b}$.

The simplest case is a linear restriction, which is written as

$$\mathbf{Rb} = \mathbf{r}$$

where $\mathbf{R}$ is a matrix and $\mathbf{r}$ is a vector. For example, let

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

Consider testing

$$\mathcal{H}_0 : \; b_1 = 0$$

In this case we let the other parameters $b_2 \cdots b_4$ vary freely, and only test whether b_1 equals zero. To write this linear restriction of the above form

$$\mathbf{Rb} - \mathbf{r} = \mathbf{0}$$

Let

$$\mathbf{R} = [1, 0, 0, 0]$$

$$\mathbf{r} = [0]$$

Then

$$\mathbf{Rb} - \mathbf{r} = \mathbf{0}$$

reduces to

$$[1, 0, 0, 0] \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} - [0] = [0] \quad \text{or} \quad b_1 = 0$$

Claim: Under the null,

$$E[\mathbf{R}\widehat{\mathbf{b}} - \mathbf{r}] = \mathbf{0}$$

$$
\begin{aligned}
E[\mathbf{R}\widehat{\mathbf{b}} - \mathbf{r}] &= E[\mathbf{R}\widehat{\mathbf{b}}] - \mathbf{R}\mathbf{b}_0 + \mathbf{R}\mathbf{b}_0 - \mathbf{r}] \\
&= E[\mathbf{R}(\widehat{\mathbf{b}} - \mathbf{b}_0)] + (\mathbf{R}\mathbf{b}_0 - \mathbf{r}) \\
&= \mathbf{R}E[\widehat{\mathbf{b}} - \mathbf{b}_0] + (\mathbf{R}\mathbf{b}_0 - \mathbf{r}) \\
&= 0
\end{aligned}
$$

since

$$E[\widehat{\mathbf{b}} - \mathbf{b}_0] = 0$$

which is the case if $\widehat{\mathbf{b}}$ is unbiased, and

$$\mathbf{R}\mathbf{b}_0 - \mathbf{r} = \mathbf{0}$$

by hypothesis.

Hence

$$(\mathbf{R}\widehat{\mathbf{b}} - \mathbf{r})'\mathbf{A}(\mathbf{R}\widehat{\mathbf{b}} - \mathbf{r})$$

will have a chi-square distribution if we can find $\mathbf{A}$ such that

$$\mathbf{A}^{-1} = \text{var}\left(\mathbf{R}\widehat{\mathbf{b}} - \mathbf{r}\right)$$

To see how we find this covariance matrix, let us look at how we find it in the simple example above.

$$\mathbf{Rb} - \mathbf{r} = \mathbf{0}$$

where

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

$$\mathbf{b} \sim \mathcal{N}(\mathbf{b}_0, \mathbf{V})$$

$$\mathbf{R} = [1, 0, 0, 0]$$

$$\mathbf{r} = [0]$$

Performing the multiplication, we found that

$$\mathbf{Rb} - \mathbf{r} = b_1$$

Thus, in this case,

$$\text{var}(\mathbf{Rb} - \mathbf{r}) = \text{var}(b_1)$$

Since

$$\mathbf{b} \sim \mathcal{N}(\mathbf{b}_0, \mathbf{V}),$$

where $\mathbf{V}$ has the form:

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{12} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{43} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}$$

Since we want var($\mathbf{Rb} - \mathbf{r}$), which equals b_1 in this case, it should be clear that we are after σ_1^2 as the variance of $\mathbf{Rb} - \mathbf{r}$. Note that a simple way to achieve this is to perform the transform

$$\mathbf{RVR}'$$

It turns out that this is true in general, that taking this transform of the covariance matrix of the parameters will give the covariance matrix for the restrictions, which is what we need to get a chi-square distribution.

Thus, the test statistic

$$W = (\mathbf{R\hat{b}} - \mathbf{r})'(\mathbf{RVR}')^{-1}(\mathbf{R\hat{b}} - \mathbf{r})$$

has a chi-square distribution with degrees of freedom equal to the number of restrictions, (or the dimensionality of the $\mathbf{R}$ transform.)

$$W \sim \chi^2(r)$$

Exercise 2.

Restriction Suppose you are estimating a model

$$y = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

How would you test the null hypothesis

$$\mathcal{H}_0: \; b_2 = b_3$$

Solution

$$\mathcal{H}_0: \ b_2 = b_3$$

is written

$$b_2 - b_3 = 0,$$

or in the form

$$\mathbf{Rb} - \mathbf{r} = 0,$$

where

$$\mathbf{R} = [0, 1, -1, 0]$$

$$\mathbf{r} = \left[\begin{array}{c} 0 \\ 0 \end{array} \right]$$

Exercise 3.
Null in a Granger causality test
Let us look at the test for Granger Causality we looked at before.
We had

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \epsilon_t$$

Rewrite this in vector form and specify how one would test the hypothesis that x does not Granger

$$Y_t = [y_t]$$

$$X_t = \begin{bmatrix} 1 \\ y_{t-1} \\ y_{t-2} \\ x_{t-1} \\ x_{t-2} \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} c \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$e_t = [\epsilon_t]$$

Then we can write the system as

$$Y_t = X_t' \mathbf{b} + e_t$$

Recall that we would like to test the hypothesis that $\beta_1 = \beta_2 = 0$ to check for Granger causality.
We would write this restriction as

$$\mathbf{Rb} - \mathbf{r}$$

where

$$\mathbf{R} = \left[\begin{array}{ccccc} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

$$\mathbf{r} = \left[\begin{array}{c} 0 \\ 0 \end{array} \right]$$

Lagrange multiplier test.

The LM statistic is calculated by imposing the hypothesis we want to test.

For example, if we consider

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

$$\mathbf{b} = \left[\begin{array}{c} b_1 \\ b_2 \\ b_3 \\ b_4 \end{array} \right]$$

and we want to test $b_1 = b_2 = 0$, we would only estimate the parameters b_3 and b_4.

If we are doing LS, would estimate

$$\tilde{\mathbf{b}} = \underset{b=(0,0,b_3,b_4)}{\arg \min} \ \mathbf{SSE}(b)$$

the estimate $\tilde{\mathbf{b}}$ is thus a restricted estimate.

In any estimation problems it may be easier to find the restricted estimate, since it has less parameters to minimize over.

If we let $\mathbf{V}(\mathbf{b})$ be the parameter covariance matrix, under the null $\mathbf{V}(\tilde{\mathbf{b}})$ will consistently estimate $\mathbf{V}(\mathbf{b})$. We will stick to the OLS case, where

$$\mathbf{V}(\tilde{\mathbf{b}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

$$\widehat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}})$$

Just like before, we will want to calculate a chi-square distributed variable by creating a quadratic form.

Like the Wald statistic, define the restrictions on the form

$$R(\mathbf{b}) = 0,$$

or in the linear case

$$\mathbf{Rb} - \mathbf{r} = \mathbf{0}$$

Consider the estimation of

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

subject to

$$\mathbf{Rb} - \mathbf{r} = \mathbf{0}$$

To estimate this with least squares without the restrictions, would minimize

$$\widehat{\mathbf{b}}^{ols} = \arg\min_{b}(\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})$$

If we consider the restricted optimimization problem we are minimizing this subject to a constraint, which we solve by use of a Lagrangian:

$$\min_{b,\lambda} L(b, \lambda, \mathbf{y}, \mathbf{X}) = (\mathbf{y} - \mathbf{X}b)'(\mathbf{y} - \mathbf{X}b) + (\mathbf{R}b - \mathbf{r})'\lambda$$

The first order conditions (Kuhn-Tucker conditions) are:

$$-\mathbf{X}'(\mathbf{y} - \mathbf{X}b) + \mathbf{R}\lambda = 0$$

$$\mathbf{R}b - \mathbf{r} = 0$$

implying

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}b) = \mathbf{R}\lambda$$

Consider

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\tilde{b})$$

Under the null hypothesis $(\mathbf{y} - \mathbf{Xb})$ has the same distribution as the error term $\mathbf{e}$, which is normal:

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

Thus,

$$\text{var}\left(\mathbf{X}'(\mathbf{y} - \mathbf{X\tilde{b}})\right)$$

will equal

$$E\left[\left(\mathbf{X}'(\mathbf{y} - \mathbf{X\tilde{b}})\right)\left(\mathbf{X}'(\mathbf{y} - \mathbf{X\tilde{b}})\right)'\right]$$

$$= E\left[\mathbf{X}'(\mathbf{y} - \mathbf{X\tilde{b}})(\mathbf{y} - \mathbf{X\tilde{b}})'\mathbf{X}\right]$$

or

$$E\left[(\mathbf{X}'\mathbf{ee}'\mathbf{X})\right] = \mathbf{X}'E\left[\mathbf{ee}'\right]\mathbf{X} = \mathbf{X}'\sigma^2 \mathbf{I}\mathbf{X} = \sigma^2 \mathbf{X}'\mathbf{X}$$

Since

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{R}\lambda$$

that means that

$$\text{var}(\mathbf{R}\lambda)$$

is equal to

$$\text{var}(\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b})),$$

which means that $\mathbf{R}\lambda$ has a normal distribution

$$\mathbf{R}\lambda \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{X}'\mathbf{X})$$

By taking the the usual quadratic form in the inverse of the covariance matrix, we find that

$$LM = (\mathbf{R}\lambda)'(\sigma^2 \mathbf{X}'\mathbf{X})^{-1}(\mathbf{R}\lambda)$$

has an asymptotic chi-square distributions, with degrees of freedom equal to the dimension of $\mathbf{R}$ (the number of restrictions).

Instead of using the Lagrange multipliers in the LM statistic, we can alternatively write it in its "score form"

$$
\begin{aligned}
LM &= (\mathbf{X}'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}}))'(\sigma^2\mathbf{X}'\mathbf{X})^{-1}\left(\mathbf{X}'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}})\right) \\
&= (\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}})'\mathbf{X}(\sigma^2\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}})
\end{aligned}
$$

$$
LM \rightarrow \chi^2(r)
$$

The likelihood ratio principle.

The final of the three classical test statistics, the Likelihood Ratio-like (LR) test.

The principle behind this type of test statistic is to measure the difference in fit for two estimates, one restricted and one unrestricted.

Suppose we have r restrictions, n observations and k variables to estimate. If we let $\widehat{\mathbf{b}}$ be the unrestricted estimate(which could be used in a Wald type test), and $\tilde{\mathbf{b}}$ the restricted estimate(which could be used in an LM test), we construct

$$LR = \frac{\frac{1}{r}(\mathbf{SSE}(\tilde{\mathbf{b}}) - \mathbf{SSE}(\widehat{\mathbf{b}}))}{\frac{1}{n-k}\mathbf{SSE}(\widehat{\mathbf{b}})}$$

Note that this is always positive since the objective function for a restricted estimate will always be larger than the one for an unrestricted estimate.

In large samples,

$$LR \to \chi^2(r)$$

Finite samples

For most of the arguments here we have relied on asymptotic theory, in which the normal (and chi-square) distributions are the distributions of interest. The use of these distributions are based on the properties of an estimator as the size of the sample increases to infinity.

In many statistical texts one will introduce a correction for the fact that we do not really have an infinite sample. The relevant distributions are the t and the F distributions, which are the finite sample counterparts of the normal and chi-square distributions.

If you have normality, and a finite sample, you could alternatively use the F distribution instead of the chi-square distribution. If you want to be conservative, this is usually also a good idea even if you are using asymptotic theory to justify the chi-square distribution. The error one makes in using the chi-square is usaully not large