Inference in regressions

Bernt Arne Ødegaard

24 November 2021

1 Inference, particularly in regression settings

Given a set of estimates, we need to judge our confidence in the estimates, and more generally, in the model we have specified.

This lecture does so by first going through some very simple examples, illustrating the typical inference we do, before going back, being very general, talking about the classical testing one does in regression frameworks:

- 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.

2 Inference in univariate regressions

- What affects accuracy of regressions?

- Confidence intervals

- Hypothesis tests

- Using R-squared

When we use the regression
$$y = a + bx + e$$
and find the OLS estimates, we have made a number of assumptions

- This is the true model.

- The numbers x are constants, and represent the possible set of x values.

- The uncertainty, or noise, in the relationship, as represented by e, the noise term, satisfy a number of assumptions.

    These assumptions rougly imply that each drawing of $\tilde{e}$ is independent of other drawings of $\tilde{e}$.

With enough independence, the distribution of e is "as if" drawn from a normal distribution.

This is what underlies testing in regression models.

The etimates of a and b are functions of normally distributed random variables, and therefore normal.

Recall our earlier univariate setting, where we constructed confidence intervals for the mean μ_0 of a normally distributed variable. From the sample we assume $\hat{\mu}$ is normally distributed with mean $\bar{x}$ and standard deviation s.

$$\bar{x}$$
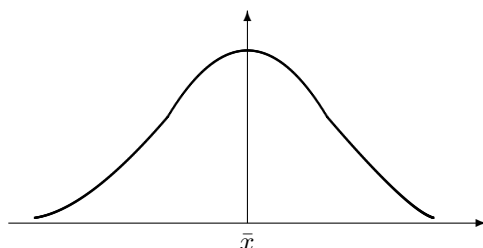
We constructed the variable $z = \frac{\bar{x}}{\frac{s}{\sqrt{n}}}$ which follows a normal distribution with unit variance.

In testing concerning parameters of regressions we do similar operations.

We construct the *probability distribution* of for example b in the regression. We then use this sample probability distribution to make probability statements.

$$\hat{b}$$
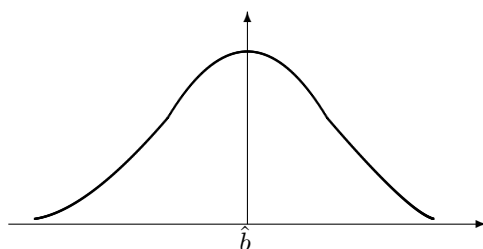
To do se need estimates of

- Parameters

- Parameter standard deviations / standard errors.

Note that the model is
$$\tilde{y} = a + bx + \tilde{e}$$

Here, the only random element is $\tilde{e}$.

The uncertainty of $\tilde{e}$ is its variance
$$\mathrm{var}(e) = E[(e - E[e])^2]$$

The error e has expectation zero, therefore
$$\mathrm{var}(e) = E[e^2]$$

We want to estimate the variance of e, and to do so use the sample equivalent

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

However this is not an unbiased estimate, we often divide by $n - 2$ instead of n to correct for this. (Think of this as "using up" two pieces of information when estimating a and b.)

To find the uncertainty in the estimate of a and b "scale" by the fact that $\hat{a}$ and $\hat{b}$ are functions of the uncertainty in $\tilde{e}$. We calculate the standard errors of the two parameter estimates as:

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

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

where

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n}\sum_{i=1}^n e_i^2}$$

We can move the estimate of S inside the formulas and also calculate them as

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

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

As soon as we have these standard errors we are ready to do inference.

We can use this standard error to do confidence interval estimation and hypothesis testing just like in the univariate case *as long as we only look at one parameter at the time*. This is a very important note to make. Even in the univariate regression there are two parameters, a and b. Potentially want to consider these two variables jointly. We will return to this issue later.
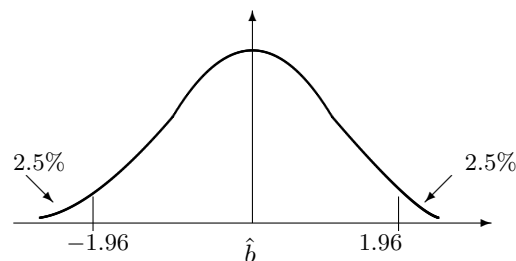
2.1 Confidence interval construction

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.

For the usual example, with $\alpha = 5\%$

if we assume a normal distribution we use $d = 1.96$.

2.2 Testing hypotheses about single parameters

Will test hypothesis of the types

$$a = a_0$$

and

$$b = b_0$$

To do so construct test statistics

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

and

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

If we use a normal assumption both these test statistics will *under the null* be normally distributed with mean zero and variance one, which we can use to test whether the null hypothesis is reasonable given the data.

2.3 The F test

Another well known test statistic that we will return to later is the F-statistic.

It tests whether the nonconstant terms of the regression is zero, i.e. it compares the two regression formulations

$$y = a + bx + e$$

to

$$y = a + e$$

This works by comparing the fit of these two cases. We already have the R^2 as a measure of fit, but that we can not say anything about the signifiance of. The F statistic lets us move to a situation where we can make probability statements.

In the univariate regression case the F statistic is calculated as

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

It has an F distribution $F(1, n)$, i.e. with degrees of freedom of 1 and n, the number of observations.

2.4 Finance example: CAPM

Exercise 1.

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.

Solution to Exercise 1.

The data is read in. Here the input file contains data aligned on observation date. But some thought necessary here. The interest rate is observed looking forward, while returns are looking backward from observation date. Also note that the risk free rate is in annualzed percentages, which need correcting

```
> 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]
```
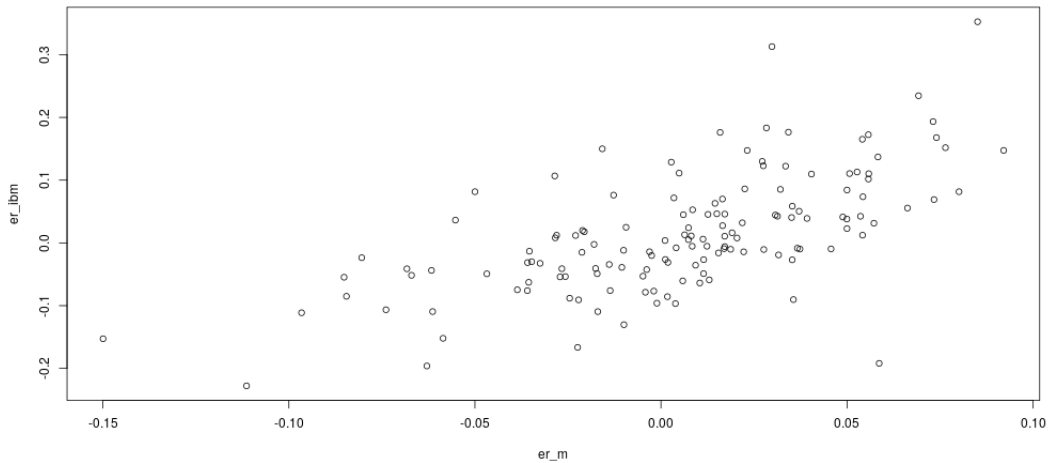
Calculating execess returns

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

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

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

Plotting one against the other

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
```
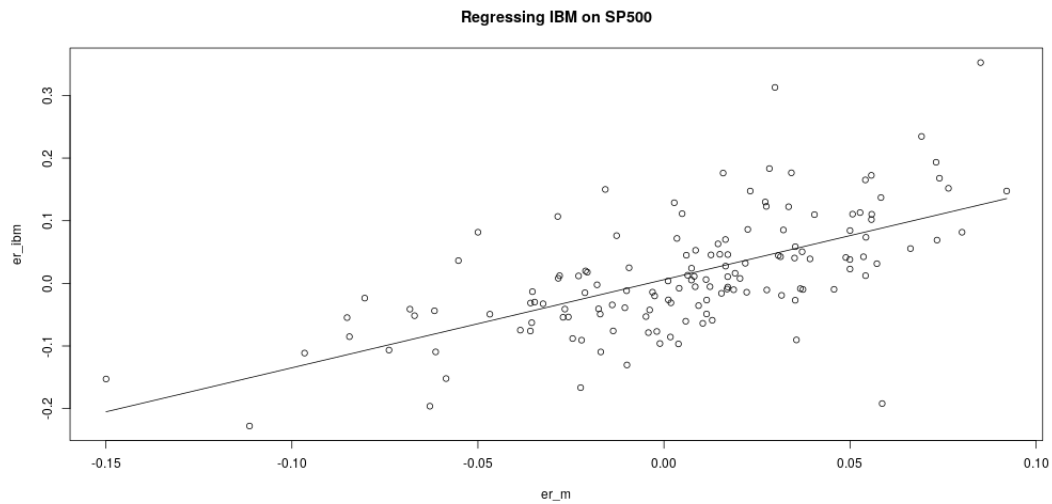
We estimate the parameters as

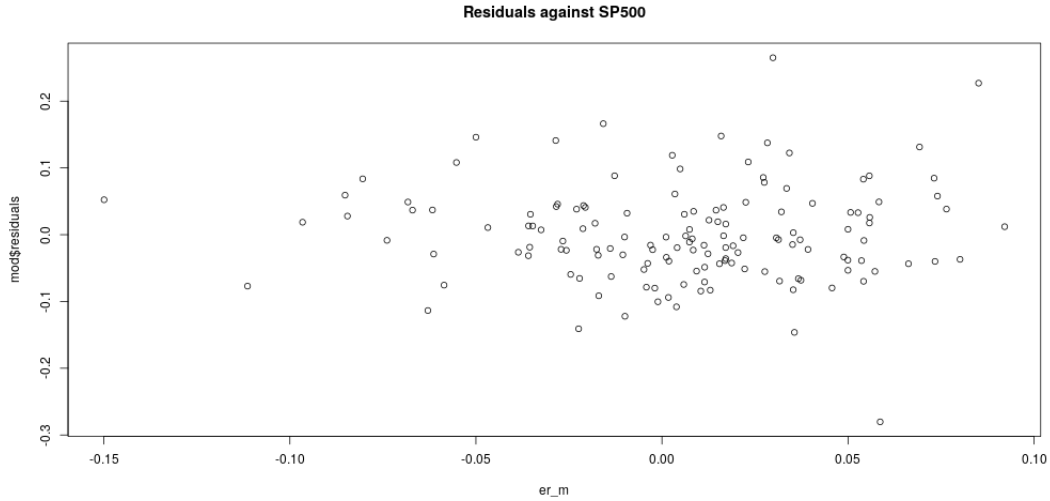$$\hat{a} = 0.005$$
$$\hat{b} = 1.407$$

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")
```

6

Residuals against SP500

Now, what restrictions do the CAPM impose on the parameters? Well, that in this regression $a = 0$.

Set up a hypotesis test:

$$H_0: \quad a = a_0 = 0$$
$$H_A: \quad a \neq 0$$

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.

3 Predicting

An important use of regressions is for prediction purposes. So far, in looking at

$$y = \hat{a} + \hat{b}x + e$$

we have asked about the significance of estimated $\hat{a}$ and $\hat{b}$, which is related to the uncertainty in $\hat{a}$ and $\hat{b}$.

A related issue is the uncertainty of the predicted value $\hat{y}$.

If you, for a given $x = x_0$, calculated the predicted value

$$\hat{y}_0 = \hat{a} + \hat{b}x_0$$

you are concerned about the confidence of your prediction, which you for example will express in a confidence interval

$$[y_0 - dS_y, y_0 + dS_y]$$

so, how do you calculate S_y?

You need the variance of the difference between $\hat{y}$ and its true value y.

Since $\hat{y} = \hat{a} + \hat{b}x_0 + e$ need to factor in the uncertainty of the estimators $\hat{a}$ and $\hat{b}$,

$$\mathrm{var}(\hat{y}_0 - y_0) = \mathrm{var}(\hat{a} - a) + \mathrm{var}(\hat{b} - b) + 2\mathrm{cov}(\hat{a} - a, \hat{b} - b) + \mathrm{var}(e)$$

It can be shown that this quantity is calculated as

$$\mathrm{var}(\hat{y}_0 - y_0) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$$

The important featur of this relationship is that it increases in the distance between x_0 and the mean x, $\bar{x}$.

7

Confidence bands for prediction of y
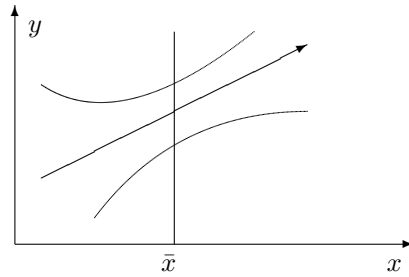
Lesson to take home: Predictions get progressively more uncertain the futher you go from the values of x used to generate the estimates $\hat{a}$ and $\hat{b}$.

4   Summarizing

Regression estimated variance

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2$$

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2}$$

Parameter Standard Errors

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

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

Confidence intervals

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

d: function of confidence level Hypothesis test (of the form)

$$a = a_0$$

Construct test statistics

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

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

Testing for whether x explains y

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

5   References

?

6 Inference in multivariate regressions

6.1 Finance Example

Exercise 2.

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

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

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 to Exercise 2.

Reading the data:

```
> 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
```

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 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

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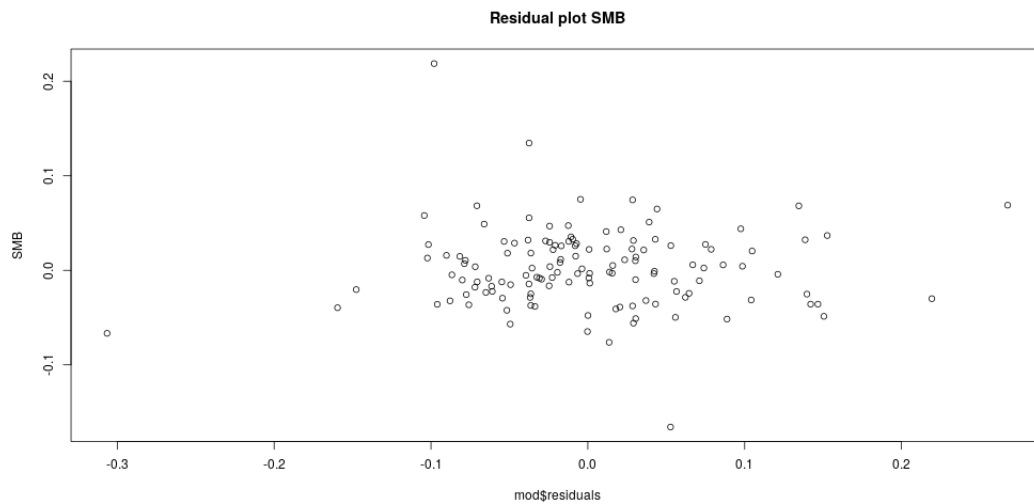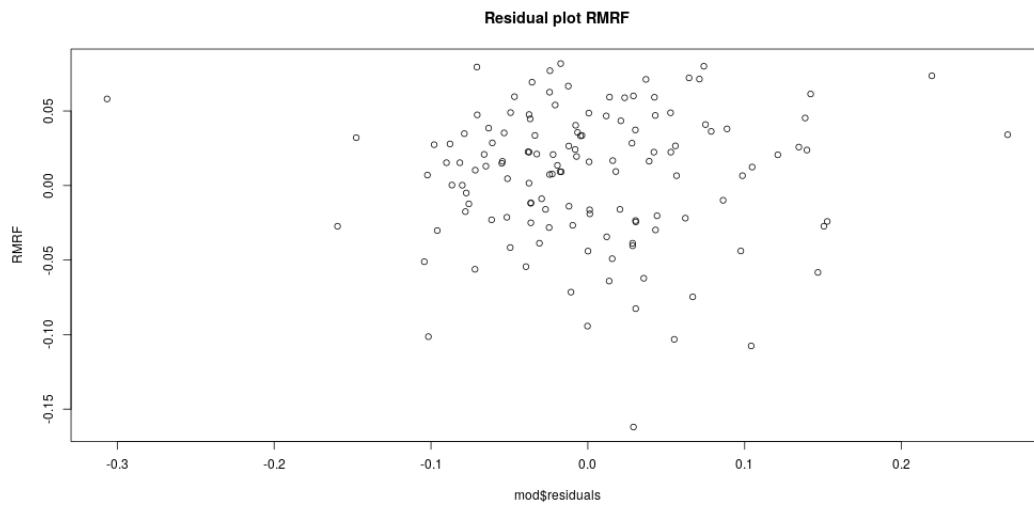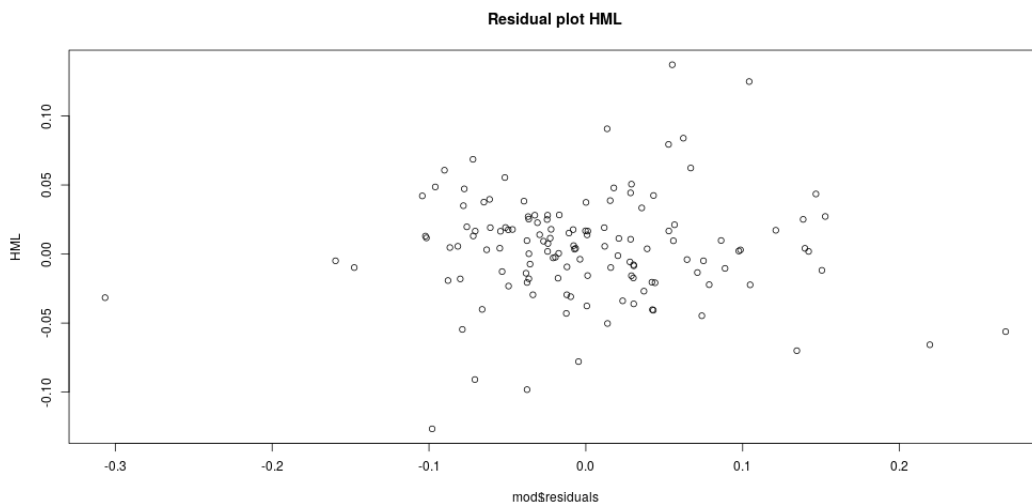
Observe the R^2 of the regression. $R^2 = 38.7\%$.

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%.

6.2 Output overview

The following important output features

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)

This is typically the situation you face. after having done your regression, you are faced with this output.

What do you do with it?

Let us go over each of the elements printed.

Coefficient estimates: $\hat{a}$, $\hat{b_1}$, $\hat{b_2}$, ... , $\hat{b_k}$.

These are variables of main interest, which you would use for prediction purposes

$$y_0 = \hat{a} + \hat{b}x_1^0 + \hat{b}_2 x_2^0 + \ldots + \hat{b}_k x_k^0$$

If you want to do inference, you need to look at the other calculated numbers.

11

Standard errors for each parameter s_i: standard error of estimate $\hat{b}_i$.

T values calculated as $t_i = \frac{\hat{b}_i}{s_i}$

The t-values are used for testing the hypothesis that $b_i = 0$.

The last two columns gives the actual probability values for this test, $p(z > |t_i| H_0)$ i.e. probability of observing that t value under the null hypothesis that $b_i = 0$.

Note that most computer packages test $H_0 : b_i = 0$ against the alternative $H_A : b_i \neq 0$, but some may do the one sided calculation for the test $H_0 : b_i > 0$ agains $H_A : b_i \leq 0$.

Note that the calculation of standard errors are conceptually different, because you also need to consider the fact that estimates of b_i and b_j are correlated

$\text{cov}(b_i, b_j) \neq 0$ execpt in very special circumstances.

In practice, done as follows

$$\sigma^2 = \text{var}(e)$$

$$\hat{\sigma}^2 = \frac{1}{n-k}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = SSR/(n-k)$$

$$V = \text{var}(\hat{\mathbf{b}}) = S^2(\mathbf{X}'\mathbf{X})^{-1}$$

6.3 Doing statistical inference on individual parameters

To do statistical inference on *one coefficient at the time* is exactly like the univariate regression case, you use the estimated $\hat{b}_i$ and its estimated standard error s_i in exactly the same manner.

To construct a 95% confidence interval you calculate

$$ci(\hat{b}_i) = [\hat{b}_i - d\hat{s}_i, \hat{b}_i + d\hat{s}_i]$$

where d follows from the usual considerations of confidence level and assumed distribution.

Under the normal distribution and $\alpha = 95\%$, $d = 1.96$, but with other confidence levels and other distributional assumptions d is something else.

Similarly , to test a hypothesis

$$H_0 : b_i = b_i^0$$

we again construct

$$z_i = \frac{\hat{b}_i - b_i^0}{\hat{s}_i}$$

and find the critical level from the size of the test.

With the usual 5% size and normality, test whether $|z| > 1.96$.

Will look at some examples of this, but first, look at the other reported numbers for the whole regression.

6.4 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

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

The analogous calculation is done 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 R^2 for multivariate regressions is thus calculated exactly as for the univariate case, and has the same interpretation.

It measures how much better the regression does in explaining the data than the naive estimate of setting y equal to its mean $\bar{y}$, which can also be written as

$$y = a + e$$

In the multivariate case R^2 has an unfortunate property, though. R^2 will never decrease when additional variables are added. Even if one potential explanatory variable is pure noise R^2 does not fall and typically rise when it is added to the regression.

The reason is that adding variables to minimize over will never hurt the $SSE(b)$. Therefore one often tries to correct for this problem by calculating

6.5 The Adjusted R^2

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

The adjusted $\bar{R}^2$ will fall if you add variables unrelated to y, and may become negative.

7 Hypothesis testing in regressions

7.1 The F test

Finally, see reported the F-value, and its p-value.

Recall from earlier, we looked at the univariate case, and compared

$$y = a + bx + e$$

to

$$y = a + e$$

The F test is a formal test for whether $b = 0$. It is better than the R^2 because it is possible to formally say something about significance.

For a multivariate regression

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k + e$$

The F test compares this to

$$y = a + e$$

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

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

Note that while in the univariate case the test is equivalent to a univariate test of whether $b = 0$ this is not the case in the multivariate case. Here we test several restrictions at the same time. This is therefore a meaningful extension of the t-values, which only test one parameter at the time.

The F statistic is calculated as

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

It has an F distribution $F(r, n-k)$, which can be used to make probability statements.

Exercise 3.

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 to Exercise 3.

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 $F(r, n-k)$.

This is the printout of the regression estimation in R.

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

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 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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.

7.2 Hypotheses involving selected parameters

Testing hypotheses involving more than one parameter.

We have seen how to do hypothesis tests involving one parameter in a regression of the form $b_i = b_i^0$, which can be tested by constructing

$$z_i = \frac{\hat{b}_i - b_i^0}{s_i}$$

where $\hat{b}_i$ is the estimated parameter, and s_i an estimate of the parameter standard error.

However, many hypotheses involve more than one parameter.

Suppose you have 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

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

i.e. you believe that the last two explanatory variables (x_3 and x_4) are irrelevant for explaining y. This is a joint hypothesis that $b_3 = b_4 = 0$.

You cannot test this hypothesis by the estimated standard errors for b_3 and b_4 alone

- What if one is significant?

- How do you deal with covariablity between $\hat{b}_3$ and $\hat{b}_4$?

There are several ways one can test hypotheses involving several parameters. Essentially we compare the fit of the two specifications, and ask wheterh the difference is significant.

As a practical matter, there are several different ways of asking that particular question in practice.

There are three general approaches.

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.

In theory, they should all give the same conclusions, at least asymptotically (as the sample gets larger)

The choice between them is made by practical considerations.

We will look at the likelihood ratio priciple, the one that requires estimation of two regressions, one restricted and one unrestricted. In the example we would estimate both (3) and (4) and then 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.

In the example

$$r = 2$$

$$k = 5$$

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

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

$LR \sim F(r, n - k)$ in small samples, and $LR \sim \chi^2(r)$ in large samples.

Exercise 4.

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.

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 to Exercise 4.

```
> 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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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 )
> summary(mod2)

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 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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:

```
> 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 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

7.3 Summarizing

Full regression specification test "The" F test

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

Comparison

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

and

$$y = a + e$$

Hypotheses involving selected parameters
   Restricted model: impose parameter restriction.
   Unrestricted model: do not impose parameter restriction.

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.

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

$LR \sim F(r, n-k)$ in small samples, and $LR \sim \chi^2(r)$ in large samples.

References   ?

8 Inference.

We now look at the basics of the typical inference done in econometric analysis.
   We have looked at some of the typical analysis that leads to distributional results about the estimated parameters.
   In classical OLS, where

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

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

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

We will look at this "classical case,"

Under the less stringent assumptions used when we discussed asymptotic theory we found that

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

where

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

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

or, $\mathbf{M}$ is some kind of average of the x values, and $\mathbf{V}$ some kind of average of the covariance matrix of $\sum_{t=1}^{T} x_t u_t$. Let me note that the classical OLS assumptions above is a special case of this, since in that case (iid normal)

$$E[x_t x_t] = \mathbf{X}'\mathbf{X}$$

$$
\begin{aligned}
\mathrm{var}\left(\sum_{t=1}^{T} x_t u_t\right) &= \sum_{t=1}^{T} \mathrm{var}(x_t u_t) \\
&= E[\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}] \\
&= \mathbf{X}'E[uu']\mathbf{X} \\
&= \sigma_0^2 \mathbf{X}'\mathbf{X}
\end{aligned}
$$

and hence

$$
\begin{aligned}
M^{-1}V_T M^{-1} &= (\mathbf{X}'\mathbf{X})^{-1}(\sigma_0^2 \mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma_0^2 (\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

thus the asymptotic result is a more general case that includes OLS as a special case.

While it may seem mathematically interesting to know that the limiting distribution is normal, so far it is not easy to see the practical relevance of these results.

But they are used when we want to assess how reliable our estimates are, and they are therefore the most important part of econometrics, because the distributions are used to test hypotheses that are relevant from the view of economic theory about the parameters of interest.

Thus, one of the important parts of the inference procedures in econometrics is to estimate covariance matrices of the general from

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

The procedures for doing this estimation is discussed elsewhere, but once this is done, the issue is how to answer the relevant questions, which means we have to figure out *what are* the relevant questions?

9 Confidence in parameter estimates.

Issue: How confident can we be about our estimates? For example, if we have estimated a linear model with two parameters,

$$y_t = b_0 + b_1 x_t + e_t,$$

and found parameter estimates

$$y_t = 2 + 0.5 x_t + e_t$$

Without an estimated covariance matrix for

$$\mathbf{b} = \left[\begin{array}{c} b_0 \\ b_1 \end{array}\right] = \left[\begin{array}{c} 2 \\ 0.5 \end{array}\right]$$

we know little about the relation between **y** and **x**.

If we knew that

$$\text{cov}(\mathbf{b}) = \left[\begin{array}{cc} 0.5 & 0 \\ 0 & 0.1 \end{array} \right]$$

we know that the estimates are much more precise than would be the case if e.g.

$$\text{cov}(\mathbf{b}) = \left[\begin{array}{cc} 2 & 0 \\ 0 & 1 \end{array} \right]$$

In the latter case we do not have any reason to think that there is a significant relation between **y** and **x**, whereas in the former case we could say with a reasonable degree of confidence that **y** is related to **x**.

The two main techniques we are concerned with to answer these questions is *confidence intervals/regions* and *hypothesis testing.*

10   Construction of confidence regions.

From basic statistics we know how to interpret confidence intervals for a parameter such as a sample mean, where we would be given a mean and a standard deviation, eg, we are given the report

$$\hat{\mu} = 0.5 + / - 0.25,$$

we construct a confidence interval

$$[a, b] = [0.5 - c \cdot 0.25, 0.5 + c \cdot 0.25],$$

where c is some constant which will depend on a parameter α(typically, α is 5 percent). The interpretation of α is that we assess the probability that the confidence interval covers the true parameters as $(1 - \alpha)$. In the typical (normal) case with $\alpha = 5\%$, c would be 1.96, and the confidence interval would be

$$[0.5 - 1.96 \cdot 0.25, 0.5 + 1.96 \cdot 0.25] = [0.01, 0.99]$$

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

CONFIDENCE REGIONS

$(\hat{b_0}, \hat{b_1}).$
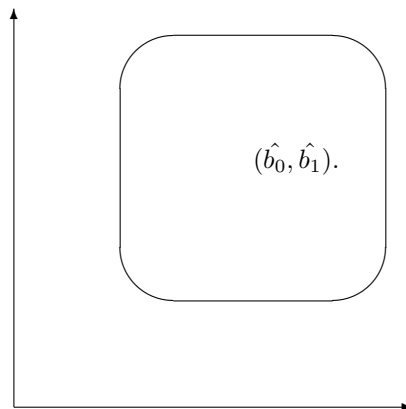
To construct a confidence interval for one parameter, let us take the standard OLS example, where we have shown

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

To find the standard deviation of one parameter, we would look at the diagonal of

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

For example, if we have two parameters b_1 and b_2, $\widehat{\mathbf{V}}$ would have the form:

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

and we would use the diagonal element to construct one parameter confidence intervals,

$$[b_1 - c\hat{\sigma}_1, b_1 + c\hat{\sigma}_1]$$

To be conservative, we could use the t-distribution instead of the normal to find the parameter c.

Confidence regions take somewhat more work to construct, since we need to 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 $\widehat{\mathbf{b}}$, either

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

or the asymptotic case

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

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 show this we use the following useful result

Exercise 5.

Let

$$x = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right]$$

have a normal distribution with mean zero and variance covariance matrix V.

$$x \sim N(0, V)$$

Claim:

$$x'V^{-1}x \sim \chi^2$$

Show this claim.

Solution to Exercise 5.

To see this, we use the result about $N(0, 1)$ variables:

If $x \sim N(0, 1)$, then $x'x \sim \chi_n^2$.

Since V is symmetric, positive definite, there is a matrix $(V^{\frac{1}{2}})$ (the square root matrix) such that

$$V^{\frac{1}{2}}V^{\frac{1}{2}} = V$$

Let
$$V^{-\frac{1}{2}} = (V^{\frac{1}{2}})^{-1}$$

Consider
$$V^{-\frac{1}{2}}x$$

Claim: The covariance matrix of $V^{-\frac{1}{2}}x$ is I:

$$E[(V^{-\frac{1}{2}}x)(V^{-\frac{1}{2}}x)']$$
$$= E[V^{-\frac{1}{2}}xx'V^{-\frac{1}{2}}]$$
$$= V^{-\frac{1}{2}}E[xx']V^{-\frac{1}{2}}$$
$$= V^{-\frac{1}{2}}VV^{-\frac{1}{2}}$$
$$= \left(V^{-\frac{1}{2}}V^{\frac{1}{2}}\right)\left(V^{-\frac{1}{2}}V^{-\frac{1}{2}}\right)$$
$$= II = I$$

but then
$$V^{-\frac{1}{2}}x \sim N(0, I)$$

and
$$(V^{-\frac{1}{2}}x)'(V^{-\frac{1}{2}}x) = x'V^{-\frac{1}{2}}V^{-\frac{1}{2}}x = x'V^{-1}x$$

has a chi-square distribution.

To construct the confidence region, would look at what values of **b** would make this chi-square statistic less than some critical value, which is found from a chi-square distribution.

Note that you can not take the one-dimensional critical values to construct a hypercube between those values. The true confidence region adjust for the cross-terms in the parameter covariance matrix. (See discussion in (**?**, pg 74-77).)

Confidence region, intervals.

$(\hat{b_0}, \hat{b_1})$.
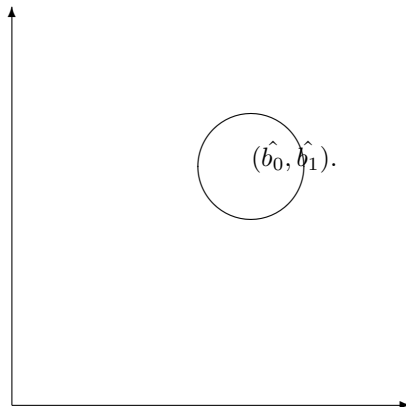
11   Test of hypotheses.

Once you move beyond two (or maybe three) dimensions in construction of confidence regions, it becomes extremely hard to visualize these regions.

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 6.

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 to Exercise 6.

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.

For further discussion about Granger Causality, see (**?**, pg 304), (**?**, Ch 18.2)

Once the hypothesis we want to investigate has been formulated, there is a number of ways to proceed, but most of them are based on three basic principles, which is what we investigate next.

12 The three classical test statistics.

Almost all test statistics we will look at in this course is based on three general principles: The Wald, Lagrange multiplier and likelihood ratio principles.

- **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.

12.1   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

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

or

$$b_1 = 0$$

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

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

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

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

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

$$\mathbf{R}\mathbf{b} - \mathbf{r} = b_1$$

Thus, in this case,

$$\operatorname{var}\left(\mathbf{R}\mathbf{b} - \mathbf{r}\right) = \operatorname{var}(b_1)$$

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

$$\mathbf{R}\mathbf{V}\mathbf{R}'$$

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
$$W = (\mathbf{R}\widehat{\mathbf{b}} - \mathbf{r})'(\mathbf{R}\mathbf{V}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\mathbf{b}} - \mathbf{r})$$
has a chi-square distribution with degrees of freedom equal to the number of restrictions, (or the dimensionality of the $\mathbf{R}$ transform.)

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

Let us look at some further examples of linear restrictions, and how they are constructed.

Exercise 7.

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 to Exercise 7.

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

is written

$$b_2 - b_3 = 0,$$

or in the form

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

where

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

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

Exercise 8.

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

Solution to Exercise 8.

$$Y_t = [y_t]$$

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

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

$$e_t = [\epsilon_t]$$

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

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

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

12.2 Lagrange multiplier test.

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

For example, if we consider

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

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

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

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

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

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:

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

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

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

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

implying

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

Consider

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

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,

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

will equal

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

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

or

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

Since

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

that means that

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

is equal to

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

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"

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

13 The likelihood ratio principle.

We now come to 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,

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

13.1 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

Readings. (**?**, 3.3,3.4,3.6)

Further Reading. Confidence intervals: (**?**, 7b)

The original articles on Granger causality are **?** and **?**.