The multiple regression model (III)

Ragnar Nymoen

University of Oslo

26 February 2013

This lecture (no 11):

Based on the references and the *model specification* in Lecture 9 and 10:

- Remarks on "goodness of fit"
- The most common approaches to hypothesis testing in the multivariate model

Adjusted R squared I

```
. reg sales price advert

      Source |       SS       df       MS              Number of obs =      75
-------------+------------------------------           F(  2,    72) =   29.25
       Model |  1396.53921     2  698.269603           Prob > F      =  0.0000
    Residual |  1718.94281    72  23.8742057           R-squared     =  0.4483
-------------+------------------------------           Adj R-squared =  0.4329
       Total |  3115.48202    74  42.1011083           Root MSE      =  4.8861

------------------------------------------------------------------------------
       sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -7.907856   1.095993    -7.22   0.000    -10.09268   -5.723034
      advert |   1.862584   .6831955     2.73   0.008     .5006587    3.224509
       _cons |   118.9136   6.351638    18.72   0.000     106.2519    131.5754
------------------------------------------------------------------------------
```

- R-squared $= 1396.53921/3115.48202 = 0.44826$
- R^2 is non-decrasing in the number of regressors included. Adj R^2 corrects for that:

Adjusted R squared II

- Adj R squared $= 1 - \frac{1718.94281}{3115.48202} \cdot (\frac{(74-1)}{(74-2-1)}) = 0.43272$

$$\overline{R}^2 = 1 - \frac{\frac{1}{n-k-1}}{\frac{1}{n-1}} \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}. \qquad (1)$$

k = the number of explanatory variables including the intercept, equal to $K - 1$ in HGL notation

- Both R^2 and Adj R^2 are *descriptive measures of goodness-of-fit*. They are not test statistics.

Adjusted R squared III

- Along with other information criteria, they can nevertheless be used as "tie breakers" between models that are equal in all other relevant aspects
- So we will remark briefly on that issue, (see HGL section 6.3.4 in particular; BN, kap 7.6.3-7.6.5)

Non-invariance of R-squared I

- Assume that we estimate

$$sala_i = \beta_0 + \beta_1 price_i + \beta_2 advert_i + \varepsilon_i$$

  where *sala* is a new lhs variable defined as $sala_i = sales_i - advert_i$
- We then know that OLS gives $\hat{\beta}_0 = 118.9136$, $\hat{\beta}_1 = -7.907856$, $\hat{\beta}_2 = 1.86 - 1 = 0.86258$
- All three standard errors are unchanged from the first regression
- Moreover, we know that $RSS = 1718.94294$ as in the original formulation
- But $R^2 = 0.424968$ which is different. What has happened?

Non-invariance of R-squared II

- R^2 is not invariant to *re-parameterizations* of the model (changes that do no affect the disturbance)

Measures of fit that are more invariant than R-sq I

```
. reg sala price advert
```

Source	SS	df	MS
Model	1270.35665	2	635.178327
Residual	1718.94309	72	23.8742096
Total	2989.29974	74	40.3959425

Number of obs	=	75
F(2, 72)	=	26.61
Prob > F	=	0.0000
R-squared	=	0.4250
Adj R-squared	=	0.4090
Root MSE	=	4.8861

sala	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
price	-7.907856	1.095993	-7.22	0.000	-10.09268	-5.723033
advert	.8625836	.6831955	1.26	0.211	-.4993417	2.224509
_cons	118.9136	6.351638	18.72	0.000	106.2519	131.5754

- Root MSE is unchanged. It is $\sqrt{\hat{\sigma}^2} = \sqrt{1718.94309/72} = \sqrt{23.874} = 4.8861$
- Hence, our estimate of σ^2 is a more invariant measure of fit than both R^2 and R^2-adj

Measures of fit that are more invariant than R-sq II

- $\hat{\sigma}$ is not invariant to how the data is scaled. The *coefficient of variation*

$$\frac{\hat{\sigma}}{\overline{Y}}100$$

  is often reported. It is the *residual standard deviation* as a percent of the level of the dependent variable (Y)

- If the data have been log-transformed, $\hat{\sigma} \cdot 100$ has a similar interpretation, since

$$\hat{\varepsilon}_i = \ln(Y_i / \hat{Y}_i) = \ln(\frac{Y_i - \hat{Y}_i}{\hat{Y}_i} + 1) \approx \frac{Y_t - \hat{Y}_i}{\hat{Y}_i},$$

  and $\hat{\varepsilon}_i 100$ becomes approximately equal to the percentage deviation between actual and fitted Y.

- See section 2 of the Lecture note: "2 points about the use of logs in econometric models"

Information criteria I

In modern econometrics two information criteria are often cited alongside, or instead of Adj R^2:

- *AIC*: Akaike information criterion

$$AIC = \ln\left(\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n}\right) + \frac{2(k-1)}{n}$$

- *SC*: Schwarz criterion

$$SC = \ln\left(\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n}\right) + \frac{(k-1)\ln(n)}{n}$$

- Like Adj R^2 they penalize extra regressors
- For $n \geq 8$ (HGL p 238) *SC* is stricter than *AIC*

Comparing the fit of linear and log-linear specifications I

Suppose we want to compare the linear model:

$$sales_i = \beta_0 + \beta_1 price_i + \beta_2 advert_i + \varepsilon_i \tag{2}$$

against the log-linear (log-log) model

$$\log(sales)_i = \beta_0 + \beta_1 \log(price_i) + \beta_2 \log(advert_i) + \varepsilon_i \tag{3}$$

Memo: The parameters β_1 and β_2 are partial elasticities in (3) and partial derivatives in (2).
This gives

$$\widehat{sales}_i = 118.9 - 7.91 price_i + 1.86 advert_i$$

$$\widehat{\ln(sales)}_i = 5.31 - 0.5 \ln(price_i) + 0.0454404 \ln(advert_i)$$

Comparing the fit of linear and log-linear specifications II

	lin	log-lin
R^2	0.448258	0.469105
Adj-R^2	0.432932	0.454358
$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$	4.88612	0.0623737
$\frac{\hat{\sigma}}{\overline{Y}}100$	$\frac{4.88612}{77.3747} \cdot 100 = 6.31$	$\frac{0.0623737}{4.34516} \cdot 100 = 1.44$
AIC	3.21198	-5.51004
SC	3.30468	-5.41734

t-tests I

- The t-test in the regression output is for the test situation H_0: $\beta_j = 0$ against $H_1 : \beta_j \neq 0$
- The only difference from the simple regression case is the formula for $\widehat{se}(\hat{\beta}_j)$ (see Lecture 10) and the degrees of freedom for the *t-distribution* which is $n - k - 1$ in general.
- If the questions is about including a regressor or not, these test can be used instead of the information criteria (It can be shown that $|t| > 1$ is enough to increase $Adj\ R^2$)
- Often, the economic problem that we work with leads to other test,situations that also can be tackled by t-tests
- Example. Log-linear model for $sales_i$. Could be interesting to test $H_0 : \beta_2 = 1$ against $H_1 : \beta_2 < 0$.

t-tests II

- If H_0 is rejected would then have formal evidence that, for a given price level, advertisement expenditure is taking a bigger share of sale revenues.

$$\widehat{\ln(sales)}_i = 5.31 - \underset{(0.079)}{0.5} \ln(price_i) + \underset{(0.0137)}{0.0454} \ln(advert_i)$$

- The relevant statistic, which is $T(72)$ distributed under H_0 is

$$t = \frac{0.0454 - 1}{0.0137}$$

- Calculate the one-side p-value and conclude!

F-tests I

- Often the test situation implies two or more linear restrictions on the parameters $\beta_1, \beta_2, \ldots, \beta_k$
- An F-test is used for such joint hypotheses
- Let d denote the number or linear restrictions
- We can make two regressions:
  - One **unrestricted** regression where the k variable none of the d restrictions are imposed. Call the sum of squared residuals RSS_U
  - One **restricted** regression where all the d restrictions are imposed. Collect RSS_R
- Heuristically we reject the joint H_0 that the d restrictions hold if RSS_R is significantly larger than RSS_U

F-tests II

- Specifically:

$$F = \frac{RSS_R - RSS_U}{RSS_U} \frac{n-k-1}{d} \sim F(d, n-k-1) \qquad (4)$$

under the joint H_0.

If we choose a 5 % significance level the joint H_0 is rejected if.

$$\frac{RSS_R - RSS_U}{RSS_U} \frac{n-k-1}{d} > f_{0.95,d,n-k-1}$$

Testing the existence of a relationship I

$H_0 : \beta_j = 0$ for $j = 1, 2, ..., k$ against $\beta_j \neq 0$ for at least one j
Under H_0,

$$E(Y_i \mid X_{1i}, X_{2i}, \dots X_{ki}) = \beta_0$$

so Y is linearly independent of the set of k explanatory variables.

$$\begin{aligned} \frac{RSS_R - RSS_U}{RSS_U}\frac{d}{n-k-1} &= \frac{TSS - RSS}{RSS}\frac{n-k-1}{k} \\ &= \frac{ESS}{TSS - ESS}\frac{n-k-1}{k} \\ &= \frac{R^2}{1-R^2}\frac{n-k-1}{k} \sim F(d, n-k-1) \end{aligned}$$

This is not a test of the "significance of R^2"

Subset F-test I

- In general $d < k$ (a subset of parameters are restricted)
- or the coefficients are not restricted to zero under H_0.
- In these cases the general formula (4) applies
- Example

$$\log(sales)_i = \beta_0 + \beta_1 \log(price_i) + \beta_2 \log(advert_i) + \varepsilon_i$$

and the test

$H_0 : \beta_1 = -1$ and $\beta_2 = 1$ against $H_1 : \beta_1 \neq -1$ and/or $\beta_2 \neq 1$

$RSS_R = 19.3975258.$

- $F(2, 72) = \frac{19.3975258 - 0.280114762}{0.280114762} \cdot \left(\frac{72}{2}\right) = 2456.9[0.00000]$

Testing with the use of the delta method I

- Lecture 3: A non-linear function of two random variables X and Y:

$$g(X,Y) = \frac{X}{Y}$$

- Since E and Var are linear operators, we must first find a linear approximation to $g(X,Y)$.
- This is done by Taylor expansion (Sydsæter 2003, Kap 7).
- HGL use the name *delta method*, see p. Ch 5.6.3 and A 5B.5.

Testing with the use of the delta method II

- BN page 72-73 it is show that the following holds

$$E\left(\frac{X}{Y}\right) \approx \frac{\mu_X}{\mu_Y}, \tag{5}$$

$$Var\left(\frac{X}{Y}\right) \approx \left(\frac{1}{\mu_Y}\right)^2 \left[\sigma_X^2 + \left(\frac{\mu_X}{\mu_Y}\right)^2 \sigma_Y^2 - 2\left(\frac{\mu_X}{\mu_Y}\right)\sigma_{X,Y}\right] \tag{6}$$

- Before leaving Andy's we can apply the delta method
- We then consider the linear model

$$sales_i = \beta_0 + \beta_1 price_i + \beta_2 advert_i + \beta_3 advert_i^2 + \varepsilon_i$$

- Let $advert_0$ be the optimal level of advertisement defined by the 1oc
- From HGL p 193 we have:

$$\beta_2 + 2\beta_3 advert_o = 1$$

$advert_0$ is a derived parameter that is a non-linear function of the regression parameter β_2 and β_3.

$$advert_o = \frac{1}{2}\frac{1-\beta_2}{\beta_3}$$

We also consider

$$\widehat{advert_o} = \frac{1}{2}\frac{1-\hat{\beta}_2}{\hat{\beta}_3}$$

as an estimator of the parameter $advert_o$.
If we want to test an hypothesis like

$$H_0 : advert_o = 0$$

we need to approximate $Var(\widehat{advert_o})$ by the *delta method*:

$$
\begin{aligned}
Var(\widehat{advert_o}) &= \left(\frac{1}{2}\right)^2 Var\left(\frac{1-\hat{\beta}_2}{\hat{\beta}_3}\right) \\
&\approx \left(\frac{1}{2}\right)^2 \left(\frac{1}{2.768}\right)^2 \times \\
&\left[(3.556)^2 + \left(\frac{1-12.151}{-2.768}\right)^2 \cdot (0.941)^2 - 2 \cdot \left(\frac{(1-12.151)}{-2.768}\right) \cdot 3.2887\right] \\
&= \frac{1}{4} * 0.130\,52 * 0.518\,41 = 0.016916.
\end{aligned}
$$

HGL finds almost the same in page 194 (rounding off?).

- An approximate t-value for $H_0 : advert_o = 0$ is therefore:

$$t = \frac{\frac{1}{2}\frac{1-12.1512}{-2.768} - 0}{\sqrt{0.016916}} = \frac{2.014}{\sqrt{0.016916}} = 15.0$$

- Clearly significant.
- Will start with other examples, from macro economics, on Thursday