ECON 3150/4150. Lecture 8.

The multiple regression model (II)

Ragnar Nymoen

University of Oslo

6 February 2014

This lecture:

- References are the same as noted in slide set to Lecture 7.
- *t-ratios* for the multivariate case (although other tests for the multivariate regression come in Lecture 9 and 10).
- Measures of degree of fit
- Interpretation of the model when all the regressors are indicator variables (dummies)
  —and when one or more dummys are regressors together with continuous variables.

Estimated standard errors and t-values I

- Just like in simple regression we need to replace $\sqrt{Var(\hat{\beta}_j)}$ by

$$\widehat{se}(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{n\hat{\sigma}^2_{X_j}\left[1 - r^2_{X_1,X_2}\right]}}$$

  where $\hat{\sigma}^2$ is an estimator.
- Also, by the same logic as before we choose the unbiased estimator

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n-3}. \tag{1}$$

  for σ^2, where $\hat{\varepsilon}_i$ are the OLS residuals from the bivariate regression model.

Estimated standard errors and t-values II

- Note $n-3$ instead of $n-2$ since we have now 3 exact relationships between the n residuals.
- Again, in direct parallel to the model with a single regressor, we now have

$$t = \frac{\hat{\beta}_j - E(\hat{\beta}_j)}{\widehat{se}(\hat{\beta}_j)} \; j = 1, 2. \tag{2}$$

- which is used in hypotheses testing an in the different forms of interval estimation.
- Some examples of null hypotheses that can be tested with the aid of the t-ratios:
  - H_0: $\beta_1 = \beta_1^0$
  - H_0: $\beta_2 = \beta_2^0$
  - H_0: $\beta_1 + \beta_2 = a^0$

Estimated standard errors and t-values III

- Use $N(0,1)$ or $t(n-3)$ for determination of critical values, confidence interval limits, and *p-values*

Frisch-Waugh theorem I

- We have several times stated that the β_j $(j = 1, 2, \ldots k)$ in the classical multiple regression model

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + \varepsilon_i$$

shall be interpreted as *partial effects*, since

$$\frac{\partial E(Y \mid X_{1i}, \ldots X_{ki})}{\partial X_{ji}} = \beta_j \ \forall \ j$$

- Two caveats:
  - The partial derivative is not relevant if $E(Y \mid X_{1i}, X_{2i}) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$ for example

Frisch-Waugh theorem II

- Cannot take derivative with respect to an X_j which is an indicator variable (see below)

- We can give an alternative derivation of the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ that shows that they are indeed (BLUE) estimators of the *partial derivatives*.
- Assume first that we have observations of Y_i and X_{1i} that have been "cleaned" of the influence of X_{2i}.
- Call these observations $\left\{e_{Y|X_2\,i}, e_{X_1|X_2\,i}\right\} \quad i = 1, 2, \ldots, n$.
- Based on this data set we could estimate the partial effect of X_1 on Y from the simple regression model

$$e_{Y|X_2\,i} = \beta_0^* + \beta_1^* e_{X_1|X_2\,i} + \varepsilon_i^* \tag{3}$$

The question is: *how to obtain the data set* $\left\{e_{Y|X_2\,i}, e_{X_1|X_2\,i}\right\}$ $i = 1, 2, \ldots, n$?

Frisch-Waugh theorem III

- Rest, in class and in a note
- *Conclusion:* We obtain the same estimate of $\hat{\beta}_1$in two ways:

1. Estimate the $k = 2$ regression model by OLS
2. "Regress out" the effect that X_2 has on Y and X_1, and use these residuals to estimate the partial effect of X_1 on Y.

This result is a special case of the general Frisch-Waugh theorem. This theorem, dating back to an 1933 journal article is central in modern econometrics, and you will encounter it in more advanced textbook from the last decade and in later courses.

Adjusted R squared I

- Example: We have a data set with observations ($n = 75$) of sales income (in USD) per outlet, price per burger and USD spent on advertisement.
-

```
. 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 $= 1396.53921/3115.48202 = 0.44826$

Adjusted R squared II

- R^2 is non-decrasing in the number of regressors included. Adj R^2 corrects for that:
- Adj R squared $= 1 - \frac{1718.94281}{3115.48202} \cdot (\frac{(74-1)}{(74-2-1)}) = 0.43272$

$$\overline{R}^2 = 1 - \frac{\frac{1}{n-k-1}}{\frac{1}{n-1}} \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}. \quad (4)$$

$k =$ the number of explanatory variables including the intercept.

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

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 estimated 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 $\sqrt{\hat{\sigma}^2} = \sqrt{1718.94309/72} = \sqrt{23.874} = 4.8861$
- This is SER in eq. (6.13) in SW
- 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)
- Although this is jumping ahead a little: We can note that 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.

Representing qualitative explanatory factors I

- Qualitative explanatory variables are important in econometric models:
  - Discrete levels of qualifications;
  - policy on/off;
  - seasonal effects on consumption, temporary or permanent structural breaks etc
- We represent qualitative factors by one or more *indicator variables* or *dummies*.
- We treat them as ordinary regressors, they represent no new problems for estimation and inference
- The difference from continuous regressors lie in the interpretation of the coefficients of the dummies

Indicator variables as the only explanatory variable I

- In the simplest case we have (as we have seen)

$$Y_i = \beta_0 + \beta_1 D_{1i} + \varepsilon_i \tag{5}$$

where D_i is an indicator variable:

$$D_{1i} = \begin{cases} 1 & \text{if individual } i \text{ belongs to category 1} \\ 0 & \text{else} \end{cases}$$

- As we have seen, the OLS estimators are

$$\begin{aligned} \hat{\beta}_0 &= \bar{Y}_0 \\ \hat{\beta}_1 &= \bar{Y}_1 - \bar{Y}_0 \end{aligned} \tag{6}$$

Indicator variables as the only explanatory variable II

- In modern terminology (6) is called the *difference estimator* . $D_{1i} = 1$ is then typically representing "individual in treatment group" and $D_{1i} = 0$ "no treatment" (control group)
- The difference estimator can be extended to data sets where we observe the individual Y 's before and after a *treatment period*, and where we can define a second qualitative variable

$$D_{2t} = \begin{cases} 1 & \text{if the period is after treatment} \\ 0 & \text{else} \end{cases}$$

- This leads to the *difference-in-difference estimator* in which is the OLS estimator of β_3 in the multivariate regression model:

$$Y_{it} = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2t} + \beta_3 D_{1i} D_{2t} + \varepsilon_{it} \qquad (7)$$

- Graph in Class

The combination rule for dummy variables I

- We can use several dummy variables for several qualitative factors in the same model providing we observe the following rule:
  If the intercept is included in the equation, then no-sub group of additive dummy variable should sum to a constant values.
- The purpose of this rule is to avoid creating *perfect multicollinearity* in the form of the "*dummy variable trap*".
- Operationalization: Assume that the qualitative factor is made up of m categories: it is represented in the model by $m-1$ dummy variables. The left-out category is called the **reference value**
- In the simple model we had category 0 and 1. That factor is represented by the single variable D_{1i}.

Dummys together with continuous variables I

- A common case is that $k-$variable regression model contains both continuous variables and dummies as regressors
- Example: log-linear consumption function for quarterly data:

$$\ln(C_t) = \beta_0 + \beta_1 \ln(INC_t) + \beta_2 D_{1t} + \beta_3 D_{2t} + \beta_4 D_{3t} + \varepsilon_t \tag{8}$$

where C is private consumption (in real terms), INC: household disposable income and

$$D_{ji} = \begin{cases} 1, \text{ if } j \text{ quarter} \\ 0, \text{ else} \end{cases}, \; j = 1, 2, 3.$$

4th quarter is the reference value of the qualitative variable "seasonality".

Dummys together with continuous variables II

- β_1 is the "marginal propensity to consume" (in elasticity form!)
- β_2, β_3 and β_4 represent quarterly **shifts in the intercept relative to the reference quarter**: They are NOT derivatives!
- Example in class.

Interaction variables I

- Dummies can be uses to model changes in the slope coefficients.
- An alternative model to (8) might be

$$\ln(C_t) = \beta_0 + \beta_1 \ln(INC_t) + \beta_2 \ln(INC_t) \cdot D_{4t} \\ + \beta_3 D_{1t} + \beta_4 D_{2t} + \beta_5 D_{3t} + \varepsilon_t$$

where

$$D_{4t} = \begin{cases} 1 & \text{if } t \text{ after financial deregulation} \\ 0 & \text{else} \end{cases}$$

- The hypothesis is that the elasticity $\partial \ln(C_t)/\partial \ln(INC_t)$ was permanently affected by easier access to credit etc.

Interaction variables II

- If $H_0 : \beta_2 = 0$ is rejected, we have evidence of a *structural break*: One single regression function is not representative of the whole sample.