Economics 326
Methods of Empirical Research in Economics

Lecture 13: Hypothesis testing in the multiple regression model, Part 1

Vadim Marmer
University of British Columbia

May 4, 2009

The model

- We consider the classical normal linear regression model:
  1. $Y_i = \beta_0 + \beta_1 X_{1,i} + \ldots + \beta_k X_{k,i} + U_i$.
  2. Conditional on X's, $E\left(U_i\right) = 0$ for all i's.
  3. Conditional on X's, $E\left(U_i^2\right) = \sigma^2$ for all i's.
  4. Conditional on X's, $E\left(U_i U_j\right) = 0$ for all $i \neq j$.
  5. Conditional on X's, U_i's are jointly normally distributed.
- We also continue to assume no perfect multicolinearity: The k regressors and constant do not form a perfect linear combination, i.e. we cannot find constants $c_1, \ldots, c_k, c_{k+1}$ (not all equal to zero) such that for all i's:

$$c_1 X_{1,i} + \ldots + c_k X_{k,i} + c_{k+1} = 0.$$

Testing a hypothesis about a single coefficient

- Take the j-th coefficient β_j, $j \in \{0, 1, \ldots, k\}$.
- Under our assumptions, its OLS estimator $\hat{\beta}_j$ satisfies that conditional on X's: $\hat{\beta}_j \sim N\left(\beta_j, Var\left(\hat{\beta}_j\right)\right)$, where $Var\left(\hat{\beta}_j\right) = \sigma^2 / \sum_{i=1}^{n} \tilde{X}_{j,i}^2$ (see Lecture 12).
- Therefore, $\left(\hat{\beta}_j - \beta_j\right) / \sqrt{Var\left(\hat{\beta}_j\right)} \sim N(0, 1)$.
- The conditional variance $Var\left(\hat{\beta}_j\right)$ is unknown because σ^2 is unknown.The estimator for $Var\left(\hat{\beta}_j\right)$ is

$$\widehat{Var}\left(\hat{\beta}_j\right) = \frac{s^2}{\sum_{i=1}^{n} \tilde{X}_{j,i}^2},$$

where $s^2 = \sum_{i=1}^{n} \hat{U}_i^2 / (n - k - 1)$ (see Lecture 11).

Testing a hypothesis about a single coefficient

- We have that conditional on X's,

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{Var}\left(\hat{\beta}_j\right)}} \sim t_{n-k-1}.$$

- Standard error: $SE\left(\hat{\beta}_j\right) = \sqrt{\widehat{Var}\left(\hat{\beta}_j\right)} = \sqrt{s^2 / \sum_{i=1}^{n} \tilde{X}_{j,i}^2}$.

Testing a hypothesis about a single coefficient: Two-sided alternatives

- Consider testing $H_0 : \beta_j = \beta_{j,0}$ against $H_1 : \beta_j \neq \beta_{j,0}$.
- Under H_0, we have that

$$T = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{\widehat{Var}\left(\hat{\beta}_j\right)}} \sim t_{n-k-1}.$$

- Let $t_{df,\tau}$ be the τ-th quantile of the t_{df} distribution.
- **Test:** Reject H_0 when $|T| > t_{n-k-1,1-\alpha/2}$.
- **P-value:** Find $t_{n-k-1,1-\tau}$ such that $|T| = t_{n-k-1,1-\tau}$. The p-value$=\tau \times 2$.

Testing a hypothesis about a single coefficient: One-sided alternatives

- Consider testing $H_0 : \beta_j \leq \beta_{j,0}$ against $H_1 : \beta_j > \beta_{j,0}$.
- When $\beta_j = \beta_{j,0}$ we have that

$$T = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{\widehat{Var}\left(\hat{\beta}_j\right)}} \sim t_{n-k-1}.$$

- Let $t_{df,\tau}$ be the τ-th quantile of the t_{df} distribution.
- **Test:** Reject H_0 when $T > t_{n-k-1,1-\alpha}$.
- **P-value:** Find $t_{n-k-1,1-\tau}$ such that $T = t_{n-k-1,1-\tau}$. The p-value$=\tau$.

Testing a hypothesis about a single linear combination of the coefficients

- Let $c_0, c_1, \ldots, c_k, r$ be some constants. Consider testing

$$H_0 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k = r \text{ against}$$
$$H_1 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k \neq r.$$

- **Example 1:** Consider the model

$$\ln Y_i = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + U_i.$$

  - We want to test for constant returns to scale $H_0 : \beta_1 + \beta_2 = 1$
  - In this case: $c_0 = 0, c_1 = 1, c_2 = 1, r = 1$.

Testing a hypothesis about a single linear combination of the coefficients

- Let $r, c_0, c_1, \ldots, c_k$ are some constants. Consider testing

$$H_0 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k = r \text{ against}$$
$$H_1 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k \neq r.$$

- **Example 2:** Consider the model

$$\ln(Wage_i) = \beta_0 + \beta_1 Experience_i + \beta_2 PrevExperience_i + \beta_3 X_{3,i} + \ldots \beta_k X_{k,i} + U_i.$$

  - We want to test that *Experience* and *PrevExperience* have the same effect on wage: $H_0 : \beta_1 = \beta_2$ or $H_0 : \beta_1 - \beta_2 = 0$.
  - In this case: $c_0 = 0, c_1 = 1, c_2 = -1, c_3 = \ldots = c_k = 0, r = 0$.

Testing a hypothesis about a single linear combination of the coefficients

- We have that under $H_0 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k = r$

$$\frac{c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \ldots + c_k\hat{\beta}_k - r}{\sqrt{Var\left(c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \ldots + c_k\hat{\beta}_k\right)}} =$$
$$= \frac{c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \ldots + c_k\hat{\beta}_k - (c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k)}{\sqrt{Var\left(c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \ldots + c_k\hat{\beta}_k\right)}}$$
$$\sim N(0,1).$$

- Note that

$$Var\left(c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \ldots + c_k\hat{\beta}_k\right) =$$
$$= \sum_{j=1}^{k} c_j^2 Var\left(\hat{\beta}_j\right) + \sum_{j=1}^{k}\sum_{l\neq j} c_j c_l Cov\left(\hat{\beta}_j, \hat{\beta}_l\right).$$

Testing a hypothesis about a single linear combination of the coefficients

- Consider

$$T = \frac{c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \ldots + c_k\hat{\beta}_k - r}{\sqrt{\widehat{Var}\left(c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \ldots + c_k\hat{\beta}_k\right)}}.$$

- Under $H_0 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k = r$,

$$T \sim t_{n-k-1}.$$

- **Two-sided Test:** Reject H_0 when $|T| > t_{n-k-1,1-\alpha/2}$.
- **One-sided:** When testing $H_0 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k \leq r$ against $H_1 : c_0\beta_0 + c_1\beta_1 + \ldots + c_k\beta_k > r$, reject H_0 when $T > t_{n-k-1,1-\alpha}$.

Testing a hypothesis about a single linear combination of the coefficients

- Consider the model $\ln Y_i = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + U_i$.
- We want to test for constant returns to scale:
  $H_0 : \beta_1 + \beta_2 = 1$.
- The test statistic: $T = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{\sqrt{\widehat{Var}\left(\hat{\beta}_1 + \hat{\beta}_2\right)}}$.
- $\widehat{Var}\left(\hat{\beta}_1 + \hat{\beta}_2\right) = \widehat{Var}\left(\hat{\beta}_1\right) + \widehat{Var}\left(\hat{\beta}_2\right) + 2\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)$.
  - $\widehat{Var}\left(\hat{\beta}_1\right)$ and $\widehat{Var}\left(\hat{\beta}_2\right)$ can be computed from the corresponding standard errors reported by Stata.
  - In Stata, $\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)$ can be obtained (together with the variances) by using the command "matrix list e(V)" after running a regression.
- Reject $H_0 : \beta_1 + \beta_2 = 1$ if $|T| > t_{n-3,1-\alpha/2}$.

Example

- 1000 observations were generated using the following model:

$$\left.\begin{array}{l} L_i = e^{l_i} \\ K_i = e^{k_i} \end{array}\right\} \text{ where } l_i, k_i \text{ are iid } N(0,1), Cov(l_i, k_I) = 0.5,$$

$$U_i \sim \text{iid } N(0,1) \text{ is independent of } l_i, k_i,$$

$$Y_i = L_i^{0.35} K_i^{0.52} e^{U_i}.$$

- The following equation was estimated:

$$\ln Y_i = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + U_i.$$

- We test $H_0 : \beta_1 + \beta_2 = 1$ against $H_1 : \beta_1 + \beta_2 \neq 1$ at 5% significance level.

Example

```
. regress lnY lnL lnK

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  2,   997) =  321.51
       Model |  630.003101     2  315.00155            Prob > F      =  0.0000
    Residual |  976.803234   997  .979742461           R-squared     =  0.3921
-------------+------------------------------           Adj R-squared =  0.3909
       Total |  1606.80633   999  1.60841475           Root MSE      =  .98982

------------------------------------------------------------------------------
         lnY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnL |   .4484374   .0356212    12.59   0.000     .3785364    .5183385
         lnK |    .466826   .0350918    13.30   0.000     .3979636    .5356883
       _cons |  -.0195782   .0313531    -0.62   0.532    -.0811039    .0419476
------------------------------------------------------------------------------

. matrix list e(V)

symmetric e(V)[3,3]
              lnL          lnK        _cons
  lnL   .00126887
  lnK  -.00059823    .00123144
_cons   5.066e-06     -.000058    .00098302

. display invttail(997 ,0.025)
1.9623462
```

Example

- We obtained:
  - $\hat{\beta}_1 = 0.4484374$,
  - $\hat{\beta}_2 = 0.466826$.
  - $\widehat{Var}\left(\hat{\beta}_1\right) = 0.00126887 = 0.0356212^2$
  - $\widehat{Var}\left(\hat{\beta}_2\right) = 0.00123144 = 0.0350918^2$.
  - $\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) = -0.00059823$.
  - $t_{997,0.975}$ =1.9623462.
- $\sqrt{\widehat{Var}\left(\hat{\beta}_1 + \hat{\beta}_2\right)} =$ $\sqrt{0.00126887 + 0.00123144 - 2 \times 0.00059823}$ =0.036108863.
- $T = \left(0.4484374 + 0.466826 - 1\right)/0.036108863 \approx -2.35$,
- $|T| = 2.35 > 1.962 = t_{997,0.975} \Longrightarrow$ We reject H_0.
- Note that ignoring the covariance leads to an incorrect result: $\left(0.4484374 + 0.466826 - 1\right)/\sqrt{0.0356212^2 + 0.0350918^2} \approx -1.69$.

An alternative approach

- We want to test $\beta_1 + \beta_2 = 1$ in $\ln Y_i = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + U_i$.
- Define $\delta = \beta_1 + \beta_2$ or $\beta_2 = \delta - \beta_1$ so that

$$\begin{aligned} \ln Y_i &= \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + U_i \\ &= \beta_0 + \beta_1 \ln L_i + (\delta - \beta_1) \ln K_i + U_i \\ &= \beta_0 + \beta_1 (\ln L_i - \ln K_i) + \delta \ln K_i + U_i. \end{aligned}$$

- Generate a new variable $D_i = \ln L_i - \ln K_i$.
- Estimate $\ln Y_i = \beta_0 + \beta_1 D_i + \delta \ln K_i + U_i$.
- Test $H_0 : \delta = 1$ against $H_1 : \delta \neq 1$.

Example

```
. gen D=lnL-lnK

. regress lnY D lnK

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  2,   997) =  321.51
       Model |  630.003101     2  315.001551           Prob > F      =  0.0000
    Residual |  976.803233   997  .979742461           R-squared     =  0.3921
-------------+------------------------------           Adj R-squared =  0.3909
       Total |  1606.80633   999  1.60841475           Root MSE      =  .98982

------------------------------------------------------------------------------
         lnY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   .4484374   .0356212    12.59   0.000     .3785364    .5183385
         lnK |   .9152634   .0361088    25.35   0.000     .8444054    .9861213
       _cons |  -.0195782   .0313531    -0.62   0.532    -.0811039    .0419476
------------------------------------------------------------------------------
```

- The 95% CI for the coefficient on $\ln K$ in the transformed mode does not include 1 $\Longrightarrow$ We reject H_0.
- Note that in the original equation $\hat{\beta}_1 + \hat{\beta}_2 = 0.9152634$ and $\sqrt{\widehat{Var}\left(\hat{\beta}_1 + \hat{\beta}_2\right)} = 0.0361088$.