Econometrics

Natalia Nehrebecka

Part 3

Agenda

- Classical linear regression model
- Hypothesis testing

Classical Linear Regression Model Assumptions

1. **Linear in Parameters**
   - *The model in the population can be written as*

$$y_i = \beta_1 + \beta_2 X_{2i} + \cdots \beta_K X_{Ki} + \varepsilon_i$$

     - *where* $\beta_1, \ldots, \beta_K$ *are unknown parameters of interest and* ε_i *is an unobservable random error of disturbance term*

2. **Non-random Explanatory Variables**
   - *Explanatory variables* $X_{2i}, \ldots X_{Ki}$ *are non-random for* $i = 1, 2, \ldots, N$
3. **Zero Mean of Error Term**
   - *The error term* ε_i *has an expected value of zero* $E(\varepsilon_i)=0$
4. **No autocorrelation**
   - $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad for\ i \neq j$
5. **Homoscedasticity**
   - $Var(\varepsilon_i) = \sigma^2$ for $i = 1,2, \ldots, N$

Classical Linear Regression Model Assumptions

4. **No autocorrelation**
   - $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad for\ i \neq j$
5. **Homoscedasticity**
   - $Var(\varepsilon_i) = \sigma^2$ for $i = 1,2, \ldots, N$

$$Var(\varepsilon) = \sigma^2 I = \begin{bmatrix} Var(\varepsilon_1) & Cov(\varepsilon_2, \varepsilon_1) & \cdots & Cov(\varepsilon_N, \varepsilon_1) \\ Cov(\varepsilon_1, \varepsilon_2) & Var(\varepsilon_2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\varepsilon_1, \varepsilon_N) & Cov(\varepsilon_2, \varepsilon_N) & \cdots & Var(\varepsilon_N) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}$$

Classical Linear Regression Model Assumptions

- Additional assumption:

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathrm{I})$$

Theorem: (Unbiasedness of OLS)

- Under **CLRM** assumptions the OLS estimators are the unbiased estimators of the population parameters

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- Vector $\boldsymbol{b}$ is a random vector, because it's a function of random vector

$$\mathrm{E}(\boldsymbol{b}) = \boldsymbol{\beta}$$

Variance of OLS

$$Var(\mathbf{b}) = Var((X'X)^{-1}X'y) = Var((X'X)^{-1}X'(X\beta + \varepsilon)$$

$$= Var(\beta + (X'X)^{-1}X'\varepsilon) = (X'X)^{-1}X' \underbrace{Var(\varepsilon)}_{\sigma^2 I} X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1} = \Sigma$$

$$\Sigma = \begin{bmatrix} Var(b_1) & \cdots & cov(b_p, b_q) \\ \vdots & \ddots & \vdots \\ cov(b_q, b_p) & \cdots & Var(b_k) \end{bmatrix}$$

Gauss-Markov Theorem

- Under CLRM assumptions (in the linear regression model in which the errors have expectation zero and are uncorrelated and have equal variances), **the OLS estimator** is **the best linear unbiased estimator** of the coefficients.

- It is often stated in shorthand as "**OLS is BLUE**" (best linear unbiased estimator)

  - **Best** means giving **the lowest variance of the estimate**, as compared to other unbiased, linear estimators.
  - The errors do not need to be normal, nor do they need to be independent and identically distributed (only uncorrelated with mean zero and homoscedastic with finite variance).
  - The requirement that **the estimator be unbiased** cannot be dropped, since there exist biased estimators with lower variance.

Estimator of error term variance

$$Var(\varepsilon_i) = \sigma^2 \text{ is unknown}$$

- Estimator of error term variance:

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{N - K}$$

Ex 1.

```
reg lwage educ exper

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     86.86
       Model |  36.9850396         2  18.4925198   Prob > F        =    0.0000
    Residual |  111.344712       523  .212896199   R-squared       =    0.2493
-------------+----------------------------------   Adj R-squared   =    0.2465
       Total |  148.329751       525   .28253286   Root MSE        =    .46141

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0979356   .0076224    12.85   0.000     .0829613    .1129099
       exper |   .0103469   .0015551     6.65   0.000     .0072919     .013402
       _cons |   .2168544    .108595     2.00   0.046     .0035183    .4301904
------------------------------------------------------------------------------
```

1. What is the standard deviation of error term in this model?

Estimator of variance-covariance matrix of b

$$Var(b) = \boldsymbol{\sigma^2 (X'X)^{-1} = \Sigma}$$

- $\hat{\boldsymbol{\Sigma}}$ is unbiased estimator of $Var(b)$

$$\hat{\Sigma} = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$\widehat{se}(b_k) = \sqrt{\left[\hat{\Sigma}_b\right]_{kk}}$$

Ex 2.

$$\widehat{se}(b_k) = \sqrt{\left[\hat{\Sigma}_b\right]_{kk}}$$

```
reg lwage educ exper
      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     86.86
       Model |  36.9850396         2  18.4925198   Prob > F        =    0.0000
    Residual |  111.344712       523  .212896199   R-squared       =    0.2493
-------------+----------------------------------   Adj R-squared   =    0.2465
       Total |  148.329751       525   .28253286   Root MSE        =    .46141
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0979356   .0076224    12.85   0.000     .0829613    .1129099
       exper |   .0103469   .0015551     6.65   0.000     .0072919     .013402
       _cons |   .2168544    .108595     2.00   0.046     .0035183    .4301904
------------------------------------------------------------------------------
matrix list e(V)
```

	educ	exper	_cons
educ	.0000581		
exper	3.551e-06	2.418e-06	
_cons	-.00079033	-.00008576	.01179288

$$\hat{\Sigma} = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

Distribution of b estimator

$$\varepsilon \sim N(0, \sigma^2 I)$$

$$b = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon$$

$$E(b) = \beta$$

$$Var(b) = \sigma^2(X'X)^{-1}$$

β and X are nonrandom

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

Normal sampling distributions

Statistical Hypothesis Testing

- **Testing hypotheses about a single population parameter**
- **Null hypothesis:**

$$H_0 : \beta_j = 0$$

The population parameter is equal to zero, i.e. after controlling for the other independent variables, there is no effect of x_j on y

$$\frac{b_j - \beta_j^*}{se(b_j)} \sim t(N-K)$$

- **Test statistic:**
  - t-statistic (or t-ratio)

$$\boldsymbol{t} = \frac{b_j}{se(b_j)}$$

- **Critical statistic**:

$$\boldsymbol{t}^* = t(N-K)$$

Sourse: Introductory Econometrics - A Modern Approach, J.M. Wooldridge

Statistical Hypothesis Testing

- **Testing against two-sided alternatives**

$$H_0: \beta_j = 0$$
$$H_0: \beta_j \neq 0$$

- **Rejection rule :**

$$| t | \geq t^* - \text{Reject the null hypothesis}$$

Sourse: Introductory Econometrics - A Modern Approach, J.M. Wooldridge

Ex 3.

```
reg lwage educ exper

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     86.86
       Model |  36.9850396         2  18.4925198   Prob > F        =    0.0000
    Residual |  111.344712       523  .212896199   R-squared       =    0.2493
-------------+----------------------------------   Adj R-squared   =    0.2465
       Total |  148.329751       525   .28253286   Root MSE        =    .46141

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0979356   .0076224    12.85   0.000     .0829613    .1129099
       exper |   .0103469   .0015551     6.65   0.000     .0072919     .013402
       _cons |   .2168544    .108595     2.00   0.046     .0035183    .4301904
------------------------------------------------------------------------------
```

1. Which variables are significant **(at the 5% significance level)**?
2. What is the interpretation of parameters in the model?

Confidence interval

Critical value of two-sided test

$$P\left(b_j - t^*_{1-\frac{\alpha}{2}} \cdot se(b_j) \leq \beta_j \leq b_j + t^*_{1-\frac{\alpha}{2}} \cdot se(b_j)\right) = 1 - \alpha$$

Lower bound of the Confidence interval

Upper bound of the Confidence interval

Confidence level

Sourse: Introductory Econometrics - A Modern Approach, J.M. Wooldridge

Ex 4.

```
reg lwage educ exper

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     86.86
       Model |  36.9850396         2  18.4925198   Prob > F        =    0.0000
    Residual |  111.344712       523  .212896199   R-squared       =    0.2493
-------------+----------------------------------   Adj R-squared   =    0.2465
       Total |  148.329751       525   .28253286   Root MSE        =    .46141

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0979356   .0076224    12.85   0.000     .0829613    .1129099
       exper |   .0103469   .0015551     6.65   0.000     .0072919     .013402
       _cons |   .2168544    .108595     2.00   0.046     .0035183    .4301904
------------------------------------------------------------------------------
```

1. Compute 95% confidence interval for parameters β.

Joint hypothesis tests

- Joint test with F-statistic

$$y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + \varepsilon_i$$

- Testing exclusion restrictions (eg.:

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1 : H_0 \text{ is not true}$$

- Test of overall significance of a regression:

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_K = 0$$

$$H_1 : H_0 \text{ is not true}$$

Joint hypothesis tests

- **Testing exclusion restrictions**

- **Null hypothesis:**

$$H_0: \beta_3 = \beta_4 = \beta_5 = 0$$

$$\frac{\frac{RSS_R - RSS}{g}}{\frac{RSS}{(N-K)}} \sim F(g, N-K)$$

- **Test statistic:**

$$\boldsymbol{F} = \frac{\frac{RSS_R - RSS}{g}}{\frac{RSS}{(N-K)}} = \frac{\frac{R_R^2 - R^2}{g}}{\frac{R^2}{(N-K)}}$$

where: g – number of restrictions

- **Critical statistic**:

$$\boldsymbol{F}^* = F(g, N-K)$$

Joint hypothesis tests

- **Rejection rule**

$$\mathrm{F} > F^{*} - \text{Reject the null hypothesis}$$

Sourse: Introductory Econometrics - A Modern Approach, J.M. Wooldridge

Ex 5.

```
. des wage games avgmin points rebounds assists

variable name   type    format     label      variable label
------------------------------------------------------------------------------
wage            float   %9.0g                 annual salary, thousands $
games           byte    %9.0g                 average games per year
avgmin          float   %9.0g                 minutes per game
points          float   %9.0g                 points per game
rebounds        float   %9.0g                 rebounds per game
assists         float   %9.0g                 assists per game
```

Ex 5.

```
. reg lwage games avgmin points rebounds assists
      Source |       SS           df       MS      Number of obs   =       269
-------------+----------------------------------   F(5, 263)       =     40.27
       Model |  90.2698185         5  18.0539637   Prob > F        =    0.0000
    Residual |  117.918945       263  .448361006   R-squared       =    0.4336
-------------+----------------------------------   Adj R-squared   =    0.4228
       Total |  208.188763       268  .776823743   Root MSE        =     .6696

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       games |   .0004132    .002682     0.15   0.878    -.0048679    .0056942
      avgmin |   .0302278   .0130868     2.31   0.022     .0044597     .055996
      points |   .0363734   .0150945     2.41   0.017     .0066519    .0660949
    rebounds |   .0406795   .0229455     1.77   0.077    -.0045007    .0858597
     assists |   .0003665   .0314393     0.01   0.991    -.0615382    .0622712
       _cons |   5.648996   .1559075    36.23   0.000      5.34201    5.955982
------------------------------------------------------------------------------
```

1. Are variaables **(games, rebounds, assists)** jointly significant?

Ex 5.

```
. reg lwage avgmin points

      Source |       SS           df       MS      Number of obs   =       269
-------------+----------------------------------   F(2, 266)       =     97.59
       Model |   88.109836         2   44.054918   Prob > F        =    0.0000
    Residual |  120.078927       266  .451424538   R-squared       =    0.4232
-------------+----------------------------------   Adj R-squared   =    0.4189
       Total |  208.188763       268  .776823743   Root MSE        =    .67188

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      avgmin |   .0388686   .0091319     4.26   0.000     .0208886    .0568486
      points |   .0356983     .01506     2.37   0.018     .0060463    .0653503
       _cons |   5.655762   .1163731    48.60   0.000     5.426632    5.884891
------------------------------------------------------------------------------
```

Ex 5.

```
. test games assists rebounds

 ( 1)  games = 0
 ( 2)  assists = 0
 ( 3)  rebounds = 0

       F(  3,   263) =    1.61
            Prob > F =    0.1884
```

Statistical Hypothesis Testing

- Test of overall significance of a regression
- Null hypothesis: $H_0: \beta_2 = \beta_3 = \cdots = \beta_K = 0$

$$\frac{\frac{TSS - RSS}{K-1}}{\frac{RSS}{(N-K)}} \sim F(K-1, N-K)$$

- Test statistic:

$$\boldsymbol{F} = \frac{\frac{TSS - RSS}{K-1}}{\frac{RSS}{(N-K)}} = \frac{\frac{ESS}{K-1}}{\frac{RSS}{(N-K)}} = \frac{\frac{R^2}{K-1}}{\frac{(1-R^2)}{(N-K)}}$$

- Critical statistic:

$$\boldsymbol{F}^* = F(K-1, N-K)$$

Ex 6.

```
reg lwage educ exper
      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(2, 932)       =     70.16
       Model |  21.6776613         2  10.8388306   Prob > F        =    0.0000
    Residual |  143.978622       932    .1544835   R-squared       =    0.1309
-------------+----------------------------------   Adj R-squared   =    0.1290
       Total |  165.656283       934  .177362188   Root MSE        =    .39304
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .077782   .0065769    11.83   0.000     .0648748    .0906892
       exper |   .0197768   .0033025     5.99   0.000     .0132956     .026258
       _cons |    5.50271    .112037    49.12   0.000     5.282836    5.722584
------------------------------------------------------------------------------
reg lwage
      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(0, 934)       =      0.00
       Model |           0         0           .   Prob > F        =         .
    Residual |  165.656283       934  .177362188   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =    0.0000
       Total |  165.656283       934  .177362188   Root MSE        =    .42114
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   6.779004   .0137729   492.20   0.000     6.751974    6.806033
------------------------------------------------------------------------------
```

1. Are all variables jointly significant?

June, 7 from 1.15 PM to 4.45 PM

1. Proof, that in **CLRM** estimator b is unbiased.
2. Derive the variance-covariance matrix of b. Interpret elements of this matrix.
3. Give Gauss-Markov theorem.
4. Prove that σ^2 is unbiased estimator of s^2.
5. Prove that $s^2(X'X)^{-1}$ unbiased estimator of $Var(b)$.
6. Derive the small-sample distribution of OLS estimator. What is to be assumed, except for **CLRM** assumptions?
7. Give the form of statistics to test the following hypothesis: $\beta_j = \beta_j^*$.
8. We have estimator b_k and estimator of its standard deviation se_{b_k}. How should we build the confidence interval for β_k? N – number of observations, K – number of estimated parameters, $(1 - \alpha)$ – confidence level.
9. How do we test the **joint hypothesis**, using residual sum of squares from the model with and without restrictions?
10. What are the benefits and dangers of imposing restrictions on the model?