Econometrics

Natalia Nehrebecka

Part 2

Agenda

- Function form (part 2)
- Classical linear regression model

Units of Measurement and Functional Form

- **Partial Effect:** $\frac{\Delta E(y)}{\Delta x_k}$
  - For $\Delta x_k = 1$ partial effect is equal to β_k
  - β_k describes the impact of a unit change in explanatory variable x_k in expected value of dependent variable holding all other regressors fixed

- **Elasticity:** $\frac{\partial lnE(y)}{\partial lnx_k}$
  - percentage change in expected value of dependent variable in response to 1% change in x_k

- **Semi-Elasticity**
  - multiplied by 100% - percentage change in expected value of dependent variable in response to 1 unit change in x_k

Ex 2.1 Data set

```
-----------------------------------------------------------------------------
              storage  display    value
variable name   type   format     label      variable label
-----------------------------------------------------------------------------
salary          int    %9.0g                 1990 salary, thousands $
pcsalary        int    %9.0g                 % change salary, 89-90
sales           float  %9.0g                 1990 firm sales, millions $
roe             float  %9.0g                 return on equity, 88-90 avg
pcroe           float  %9.0g                 % change roe, 88-90
ros             int    %9.0g                 return on firm's stock, 88-90
indus           byte   %9.0g                 =1 if industrial firm
finance         byte   %9.0g                 =1 if financial firm
consprod        byte   %9.0g                 =1 if consumer product firm
utility         byte   %9.0g                 =1 if transport. or utilties
lsalary         float  %9.0g                 natural log of salary
lsales          float  %9.0g                 natural log of sales
-------------------------------------------------------------------------------
```

Ex 1. Partial Effect

```
. reg salary roe

      Source |       SS           df       MS      Number of obs   =       209
-------------+----------------------------------   F(1, 207)       =      2.77
       Model |  5166419.04         1  5166419.04   Prob > F        =    0.0978
    Residual |   386566563       207  1867471.32   R-squared       =    0.0132
-------------+----------------------------------   Adj R-squared   =    0.0084
       Total |   391732982       208  1883331.64   Root MSE        =    1366.6

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         roe |   18.50119   11.12325     1.66   0.098    -3.428196    40.43057
       _cons |   963.1913   213.2403     4.52   0.000     542.7902    1383.592
------------------------------------------------------------------------------
```

1. Interpret the value of parameter in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Ex 2. Elasticity

```
. reg lsalary lsales

      Source |       SS       df       MS              Number of obs =     209
-------------+------------------------------           F(1, 207)     =   55.30
       Model |  14.0661688     1  14.0661688           Prob > F      =  0.0000
    Residual |  52.6559944   207  .254376785           R-squared     =  0.2108
-------------+------------------------------           Adj R-squared =  0.2070
       Total |  66.7221632   208  .320779631           Root MSE      =  .50436

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lsales |   .2566717   .0345167     7.44   0.000     .1886224    .3247209
       _cons |   4.821997   .2883396    16.72   0.000     4.253538    5.390455
------------------------------------------------------------------------------
```

1. Interpret the value of parameter in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Ex 3. Data set

```
------------------------------------------------------------------------
              storage  display    value
variable name   type   format     label      variable label
------------------------------------------------------------------------
wage            float  %8.2g                 average hourly earnings
educ            byte   %8.0g                 years of education
exper           byte   %8.0g                 years potential experience
tenure          byte   %8.0g                 years with current employer
nonwhite        byte   %8.0g                 =1 if nonwhite
female          byte   %8.0g                 =1 if female
married         byte   %8.0g                 =1 if married
numdep          byte   %8.0g                 number of dependents
smsa            byte   %8.0g                 =1 if live in SMSA
northcen        byte   %8.0g                 =1 if live in north central U.S
south           byte   %8.0g                 =1 if live in southern region
west            byte   %8.0g                 =1 if live in western region
construc        byte   %8.0g                 =1 if work in construc. indus.
ndurman         byte   %8.0g                 =1 if in nondur. manuf. indus.
trcommpu        byte   %8.0g                 =1 if in trans, commun, pub ut
trade           byte   %8.0g                 =1 if in wholesale or retail
services        byte   %8.0g                 =1 if in services indus.
profserv        byte   %8.0g                 =1 if in prof. serv. indus.
profocc         byte   %8.0g                 =1 if in profess. occupation
clerocc         byte   %8.0g                 =1 if in clerical occupation
servocc         byte   %8.0g                 =1 if in service occupation
lwage           float  %9.0g                 log(wage)
expersq         int    %9.0g                 exper^2
tenursq         int    %9.0g                 tenure^2
-----------------------------------------------------------------------------
```

Ex 3. Semi-Elasticity

```
. reg lwage educ

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(1, 524)       =    119.58
       Model |  27.5606288         1  27.5606288   Prob > F        =    0.0000
    Residual |  120.769123       524  .230475425   R-squared       =    0.1858
-------------+----------------------------------   Adj R-squared   =    0.1843
       Total |  148.329751       525   .28253286   Root MSE        =    .48008

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0827444   .0075667    10.94   0.000     .0678796    .0976091
       _cons |   .5837727   .0973358     6.00   0.000     .3925563    .7749891
------------------------------------------------------------------------------
```

1. Interpret the value of parameter in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Discrete variables

$$sex = \begin{cases} 0 \ - Male \\ 1 - Female \end{cases}$$

$$education = \begin{cases} 1 \ - primary\ education \\ 2 \ - secondary\ education \\ 3 \ - higher\ education \end{cases}$$

Ex 4. Data set

variable name	storage type	display format	value label	variable label
wage	float	%8.2g		average hourly earnings
educ	byte	%8.0g		years of education
exper	byte	%8.0g		years potential experience
tenure	byte	%8.0g		years with current employer
nonwhite	byte	%8.0g		=1 if nonwhite
female	byte	%8.0g		=1 if female
married	byte	%8.0g		=1 if married
numdep	byte	%8.0g		number of dependents
smsa	byte	%8.0g		=1 if live in SMSA
northcen	byte	%8.0g		=1 if live in north central U.S
south	byte	%8.0g		=1 if live in southern region
west	byte	%8.0g		=1 if live in western region
construc	byte	%8.0g		=1 if work in construc. indus.
ndurman	byte	%8.0g		=1 if in nondur. manuf. indus.
trcommpu	byte	%8.0g		=1 if in trans, commun, pub ut
trade	byte	%8.0g		=1 if in wholesale or retail
services	byte	%8.0g		=1 if in services indus.
profserv	byte	%8.0g		=1 if in prof. serv. indus.
profocc	byte	%8.0g		=1 if in profess. occupation
clerocc	byte	%8.0g		=1 if in clerical occupation
servocc	byte	%8.0g		=1 if in service occupation
lwage	float	%9.0g		log(wage)
expersq	int	%9.0g		exper^2
tenursq	int	%9.0g		tenure^2

Ex 4. Binary variables

Dependence of wage on gender

```
. reg wage female

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(1, 524)     =   68.54
       Model |  828.220467     1  828.220467           Prob > F      =  0.0000
    Residual |  6332.19382   524  12.0843394           R-squared     =  0.1157
-------------+------------------------------           Adj R-squared =  0.1140
       Total |  7160.41429   525  13.6388844           Root MSE      =  3.4763

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -2.51183   .3034092    -8.28   0.000    -3.107878   -1.915782
       _cons |   7.099489   .2100082    33.81   0.000     6.686928     7.51205
------------------------------------------------------------------------------
```

1. Interpret the value of parameter in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Ex 5. Binary variables

Logarithm of wage and gender

```
. reg lwage female

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(1, 524)       =     85.04
       Model |  20.7120004         1  20.7120004   Prob > F        =    0.0000
    Residual |  127.617751       524  .243545326   R-squared       =    0.1396
-------------+----------------------------------   Adj R-squared   =    0.1380
       Total |  148.329751       525   .28253286   Root MSE        =     .4935

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.3972175   .0430732    -9.22   0.000    -.4818349   -.3126001
       _cons |    1.81357   .0298136    60.83   0.000     1.755001    1.872139
------------------------------------------------------------------------------
```

1. Interpret the value of parameter in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Discrete variables

$$education = \begin{cases} 1 \ - primary\ education \\ 2 \ - secondary\ education \\ 3 \ - higher\ education \end{cases}$$

reg salary educ - mistaken regression!!!

Source	SS	df	MS
Model	32138756.3	1	32138756.3
Residual	715502891	1,085	659449.669
Total	747641647	1,086	688436.139

Number of obs	=	1,087
F(1, 1085)	=	48.74
Prob > F	=	0.0000
R-squared	=	0.0430
Adj R-squared	=	0.0421
Root MSE	=	812.07

salary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	**353.7361**	50.67056	6.98	0.000	254.3127	453.1595
_cons	99.2835	104.8803	0.95	0.344	-106.5077	305.0747

Ex 6. Discrete variables

$$education = \begin{cases} 1 & -\ primary\ education \\ 2 & -\ secondary\ education \\ 3 & -\ higher\ education \end{cases}$$

Wage and education – base level: primary education

regress salary secondary_edu higher_edu

```
      Source |       SS           df       MS      Number of obs   =     1,087
-------------+----------------------------------   F(2, 1084)      =     28.09
       Model |  36838530.3         2  18419265.2   Prob > F        =    0.0000
    Residual |   710803116     1,084  655722.432   R-squared       =    0.0493
-------------+----------------------------------   Adj R-squared   =    0.0475
       Total |   747641647     1,086  688436.139   Root MSE        =    809.77

------------------------------------------------------------------------------
       salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
secondary_edu |   192.8418   78.51625     2.46   0.014     38.78072    346.9028
   higher_edu |   695.4986   101.1532     6.88   0.000     497.0203    893.9769
        _cons |   577.3607   73.31285     7.88   0.000     433.5095    721.2118
------------------------------------------------------------------------------
```

1. Interpret the values of parameters in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Ex 7. Discrete variables

$$education = \begin{cases} 1 \ - primary\ education \\ 2 \ - secondary\ education \\ 3 \ - higher\ education \end{cases}$$

Wage and education – base level: primary education

```
regress lsalary secondary_edu higher_edu
```

lsalary	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]
secondary_edu	0.19284				
higher_edu	0.69549				
_cons	0.57736				

1. Interpret the values of parameters in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Function form - Interactions

- An important technique that allows for non-linearities in an econometric model is the use of interaction terms – the product of explanatory variables
- An interaction may arise when considering the relationship among three or more variables, and describes a situation in which the simultaneous influence of two variables on a third is not additive.
- If two variables of interest interact, the relationship between each of the interacting variables and a third ”dependent variable” depends on the value of the other interacting variable.

Ex 8. Function form - Interactions

$$sex = \begin{cases} 0 - Male \\ 1 - Female \end{cases}; \qquad education = \begin{cases} 1 - primary\ education \\ 2 - secondary\ education \\ 3 - higher\ education \end{cases}$$

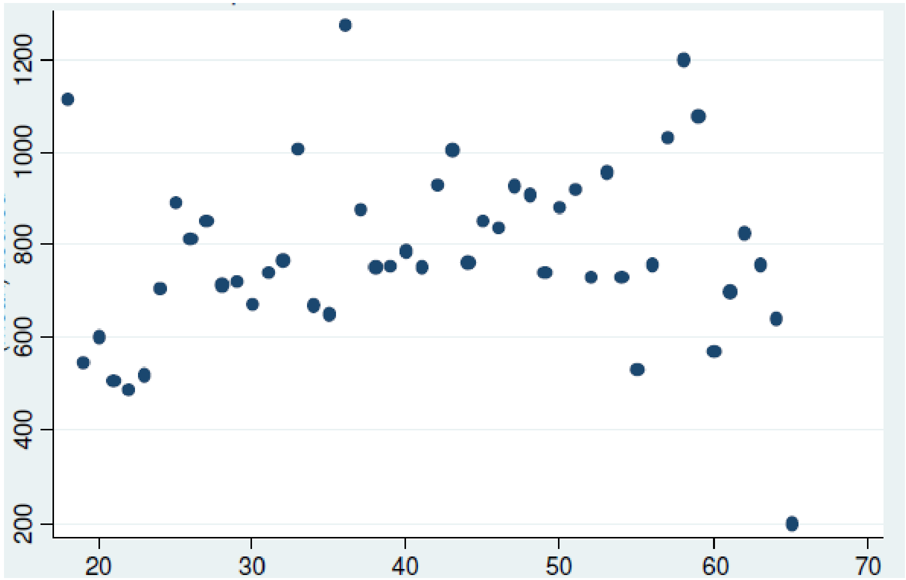
Wage: interactions between gender and education

```
xi: regress salary age i.sex*i.education
      Source |       SS       df       MS              Number of obs =    1083
-------------+------------------------------           F(  7,  1075) =   18.83
       Model |  81648217.6     7  11664031.1           Prob > F      =  0.0000
    Residual |   665832918  1075  619379.458           R-squared     =  0.1092
-------------+------------------------------           Adj R-squared =  0.1034
       Total |   747481135  1082  690832.842           Root MSE      =  787.01
------------------------------------------------------------------------------
      Salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   36.38318
     _Isex_1 |  -144.4044
_Ieducatio~2 |   274.2703
_Ieducatio~3 |   1040.998
_IsexXedu_~2 |  -143.4455
_IsexXedu_~3 |   -682.341
       _cons |  -121.1625

------------------------------------------------------------------------------------------------
```

1. Interpret the value of parameter in this model.
2. Interpret the fit of the model to the data (**R-squared**).

Function form



Ex 9. Function form

```
regress salary age age_2 sex secondary_edu higher_edu

      Source |       SS       df       MS              Number of obs =    1083
-------------+------------------------------           F(  5,  1077) =   22.98
       Model |  72048793.8     5  14409758.8           Prob > F      =  0.0000
    Residual |   675432341  1077  627142.378           R-squared     =  0.0964
-------------+------------------------------           Adj R-squared =  0.0922
       Total |   747481135  1082  690832.842           Root MSE      =  791.92

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   36.06131   15.48328     2.33   0.020     5.680494    66.44212
       age_2 |  -.3998842   .1973767    -2.03   0.043    -.7871707   -.0125977
         sex |  -338.0671   48.25867    -7.01   0.000    -432.7588   -243.3755
secondary_edu|   208.5538   77.72619     2.68   0.007     56.04182    361.0657
  higher_edu |   708.2862   99.55596     7.11   0.000     512.9406    903.6318
       _cons |  -26.64989   298.3288    -0.09   0.929    -612.0215    558.7217
------------------------------------------------------------------------------
```

1. Interpret the values of parameters in this model.
2. Interpret the fit of the model to the data (**R-squared**).

$$x = -\frac{\beta_{age}}{2\beta_{age_2}} = -\frac{36.06131}{2 * (-0.3998842)} \approx 45.09$$

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 $i = 1,2, \ldots, N$

June, 1 from 1.15 PM to 4.45 PM

1. Why do we decode discrete variable into dummy variables?
2. Why we may not include a constant and all dummies for given discrete variables in a model?
3. What do we mean by interactions in the model?
4. How can we approximate non-linear relation by linear model?
5. Describe Classical Linear Regression Model assumptions.