More on functional forms

EC 339

Marcio Santetti
Fall 2023

Motivation

New functional forms

There is more to OLS than **linear-in-variables** or **log-transformed** models.

But do these models **preserve** OLS *Classical Assumptions*?

- They do!
- But under what conditions?

As long as the model remains **linear in parameters**, everything is fine.

New functional forms

1. Regression through the **origin**

2. Regression with **quadratic** terms

3. **Inverse** forms

4. **Interaction** terms

5. **Binary** (*dummy*) variables

Regression through the origin

Regression through the origin

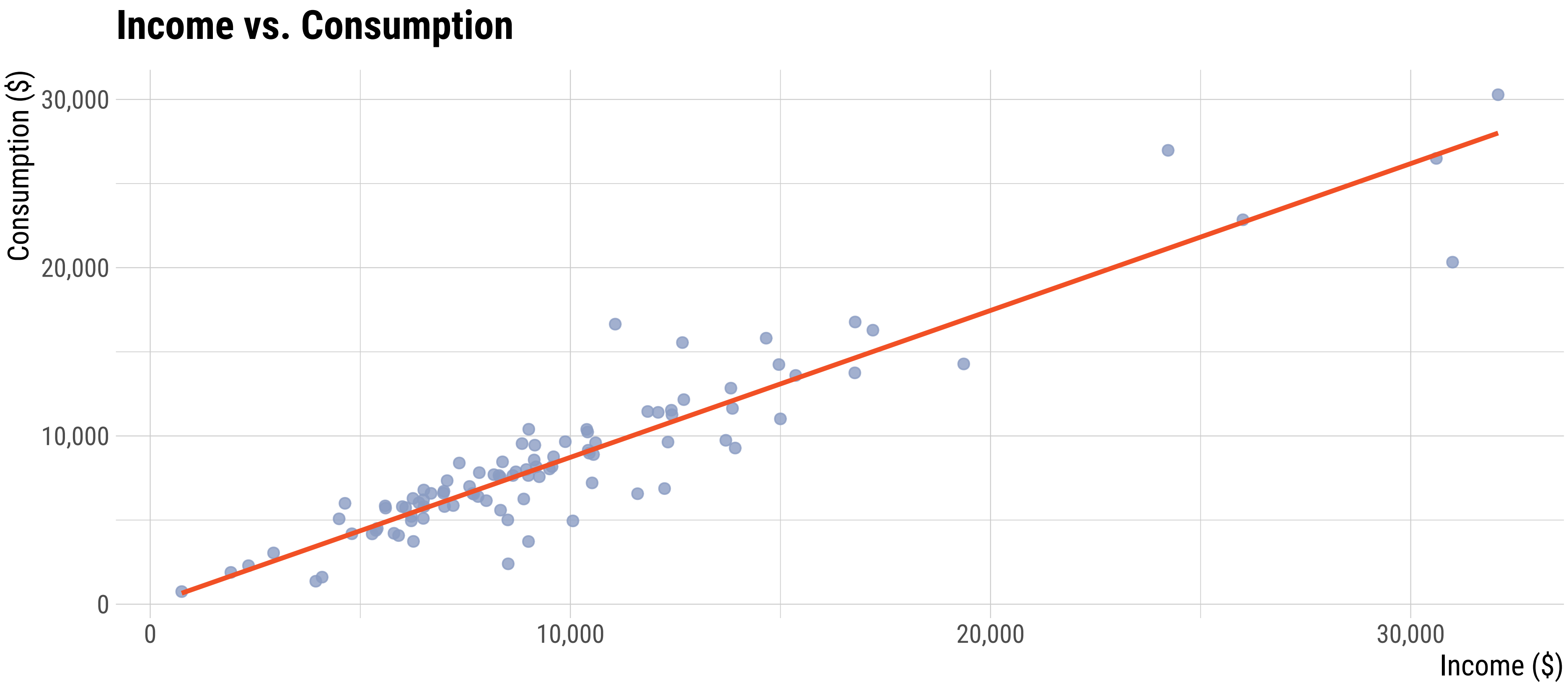
It is used whenever we need to impose the **restriction** that, when $x = 0$, the expected value of y is also zero.

It should be applied **only** when theory recommends to do so.

$$y_i = \beta_1 x_{1i} + u_i$$

Regression through the origin

$$Cons_i = \beta_1 Inc_i + u_i$$



Using quadratic terms

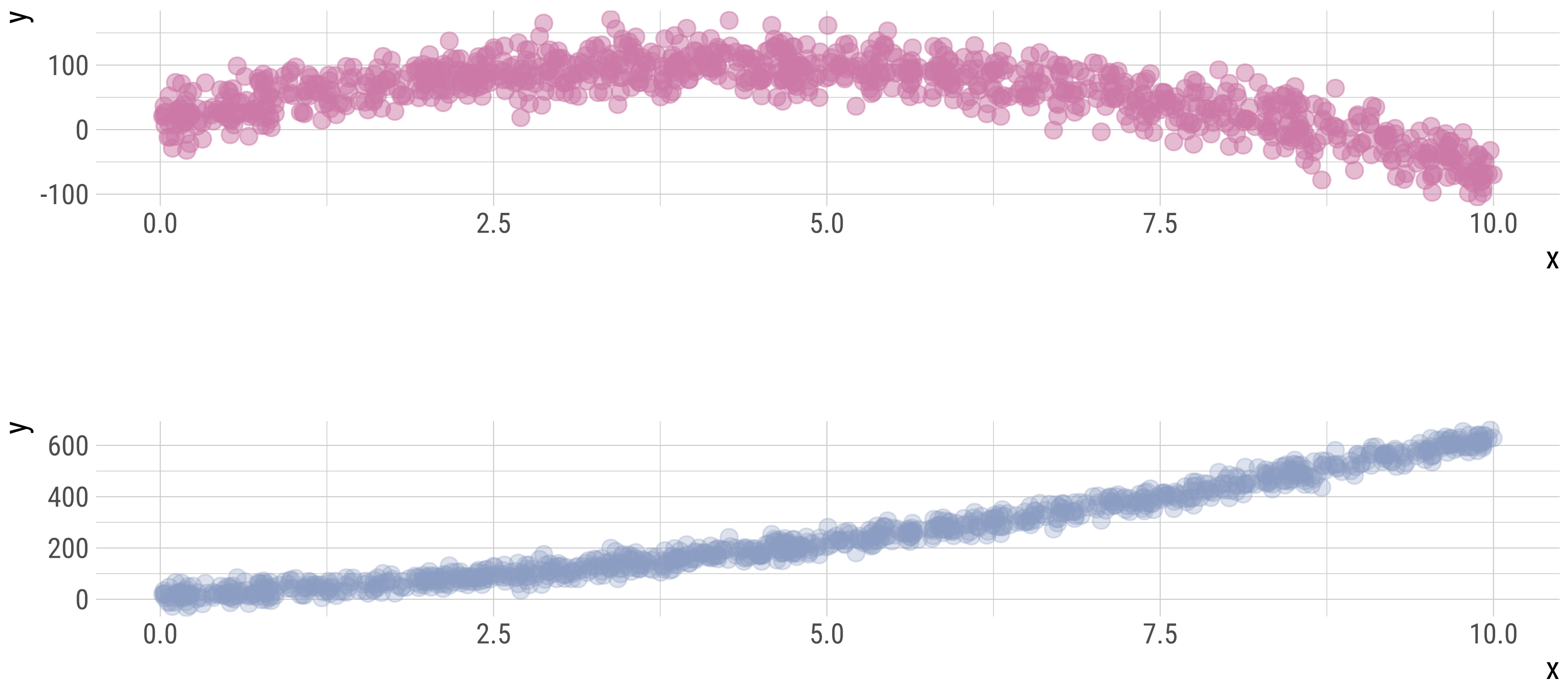
Using quadratic terms

Many times, the effect of a variable x_i on y also depends on the **level** of that independent variable.

We can also apply quadratic terms when the effect of x_i on y **changes** after a given threshold.

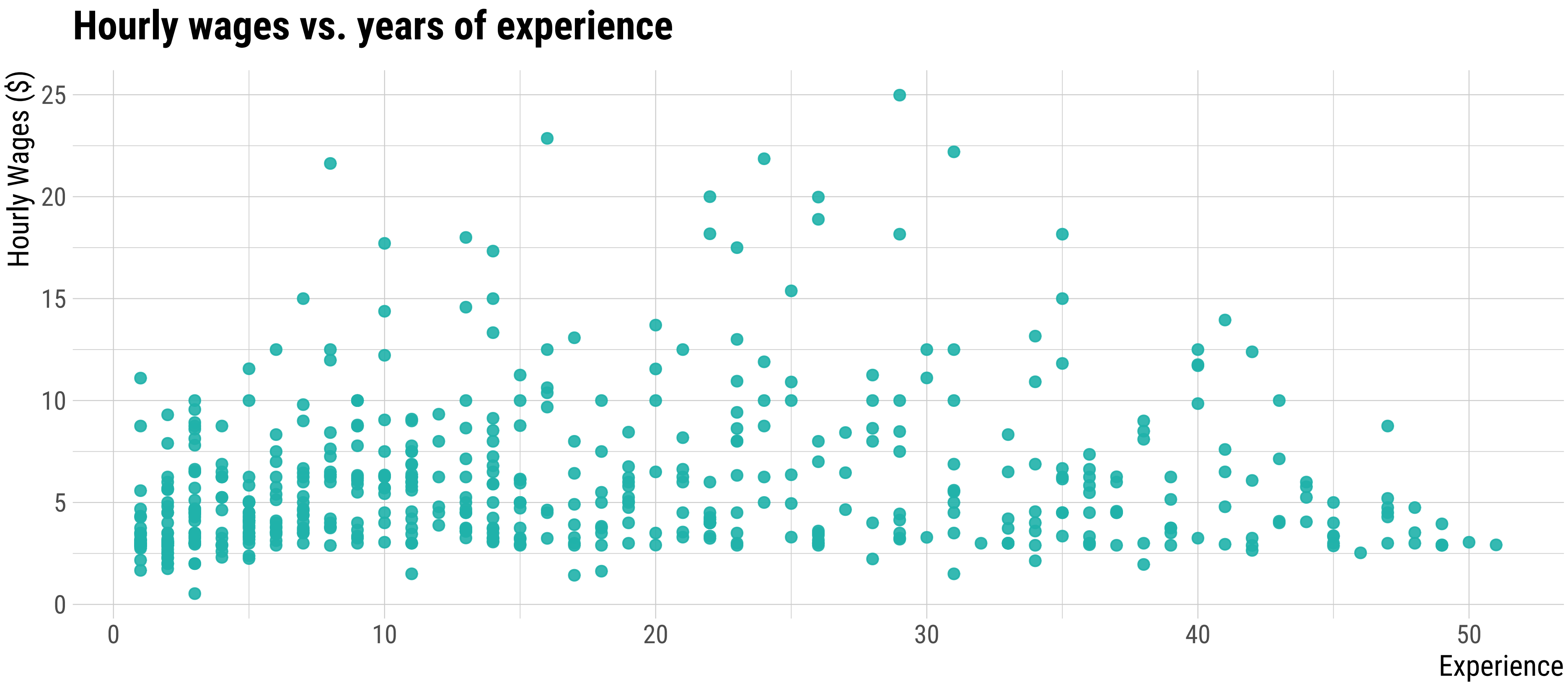
$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 (x_{1i})^2 + \cdots + \beta_k x_{ki} + u_i$$

Using quadratic terms



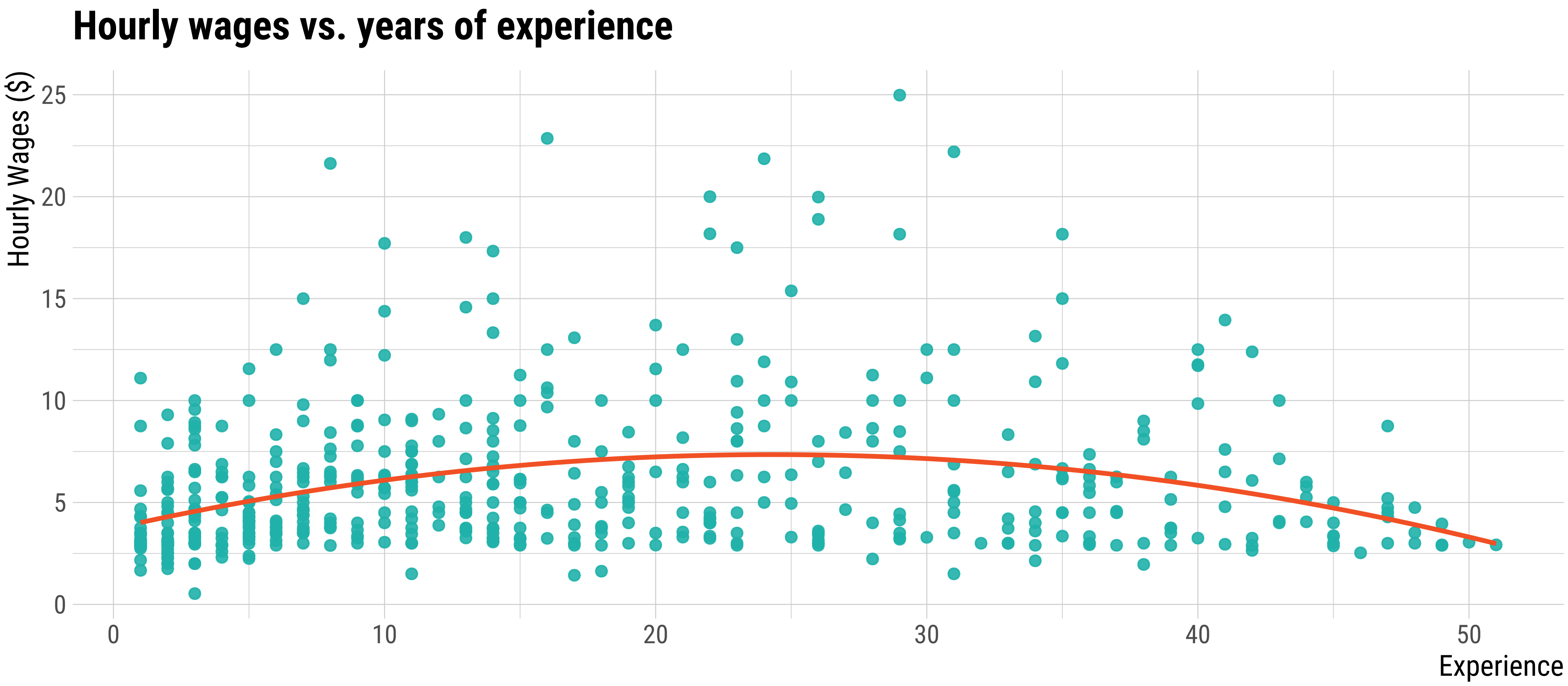
Using quadratic terms

$$wage_i = \beta_0 + \beta_1 exper_i + \beta_2 exper_i^2 + u_i$$



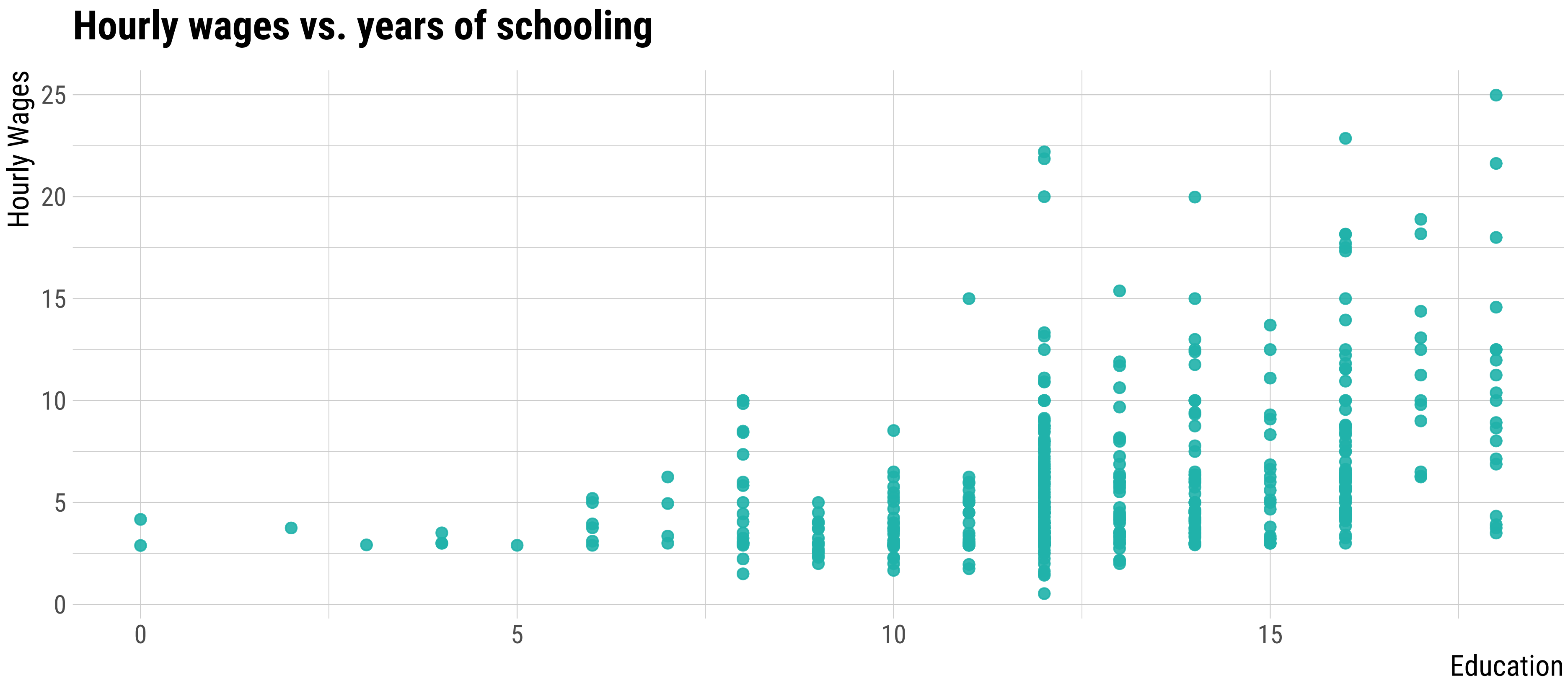
Using quadratic terms

$$wage_i = \beta_0 + \beta_1 exper_i + \beta_2 exper_i^2 + u_i$$



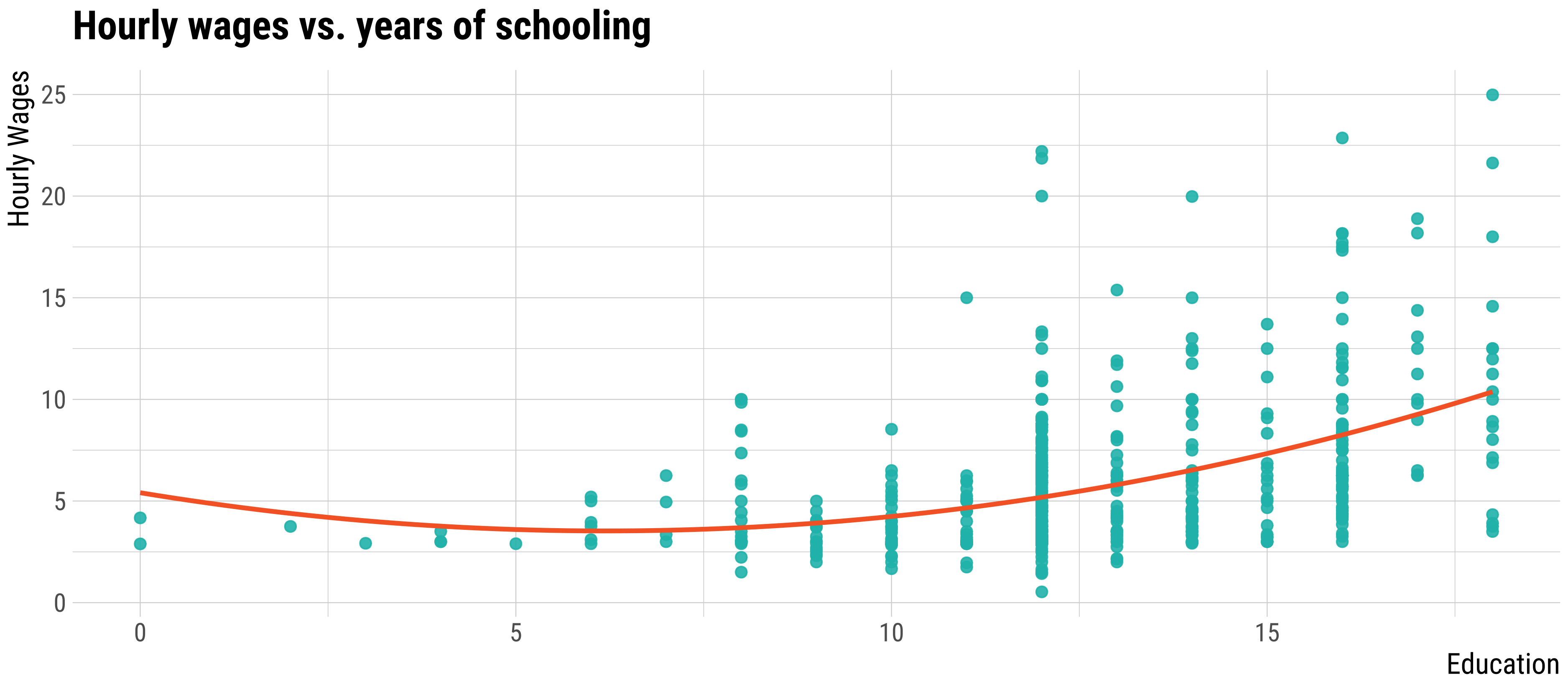
Using quadratic terms

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 educ_i^2 + u_i$$



Using quadratic terms

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 educ_i^2 + u_i$$



Using quadratic terms

Interpretation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + u_i$$

$$\frac{\partial\, y}{\partial\, x_1} = \beta_1 + 2 \ \cdot \ \beta_2 \ \cdot \ x_1$$

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 educ_i^2 + u_i$$

$$\frac{\partial\, wage}{\partial\, educ} = \beta_1 + 2 \ \cdot \ \beta_2 \ \cdot \ educ$$

Inverse forms

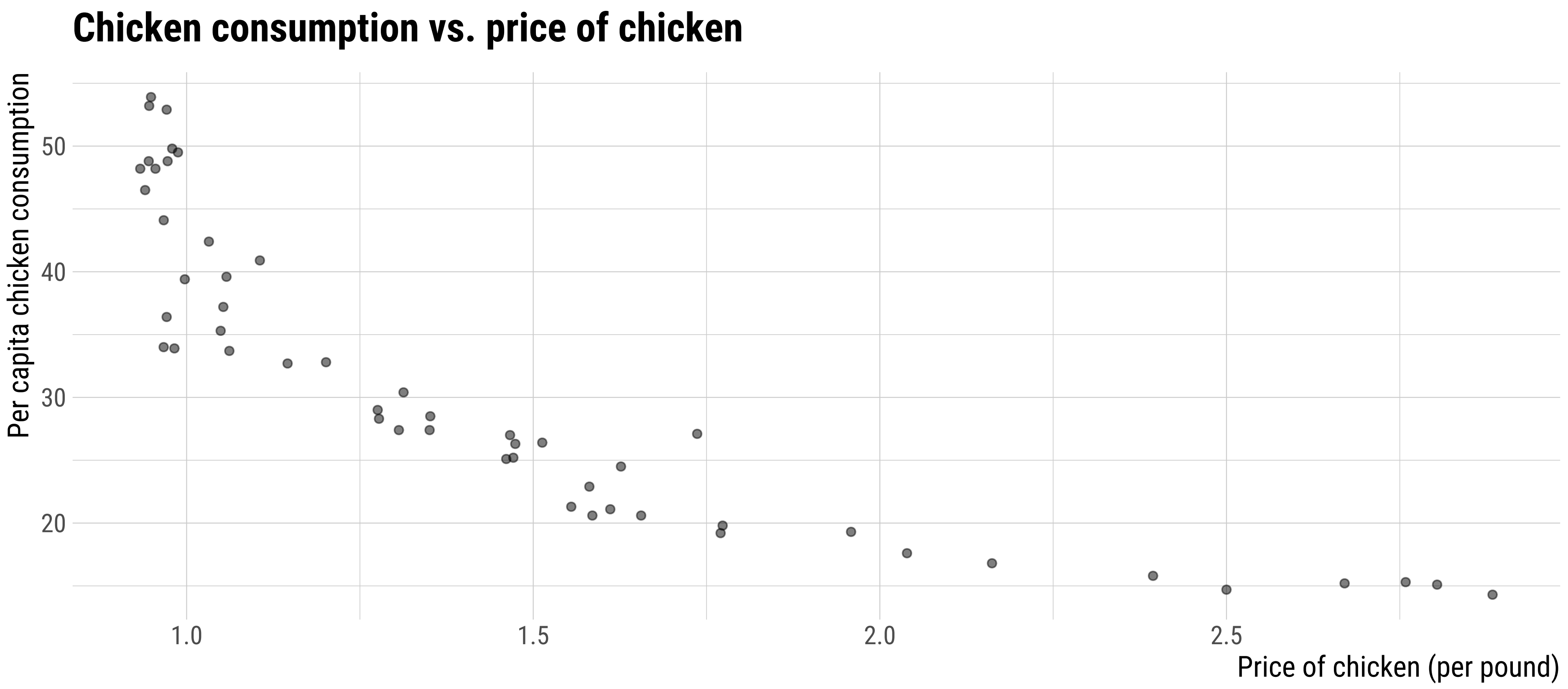
Inverse forms

Inverse forms are used whenever the effect of an independent variable on y_i is expected to approach **zero** as its value approaches **infinity**.

As always, but especially important to this category, **economic theory** should *strongly recommend* the use of such functional form.

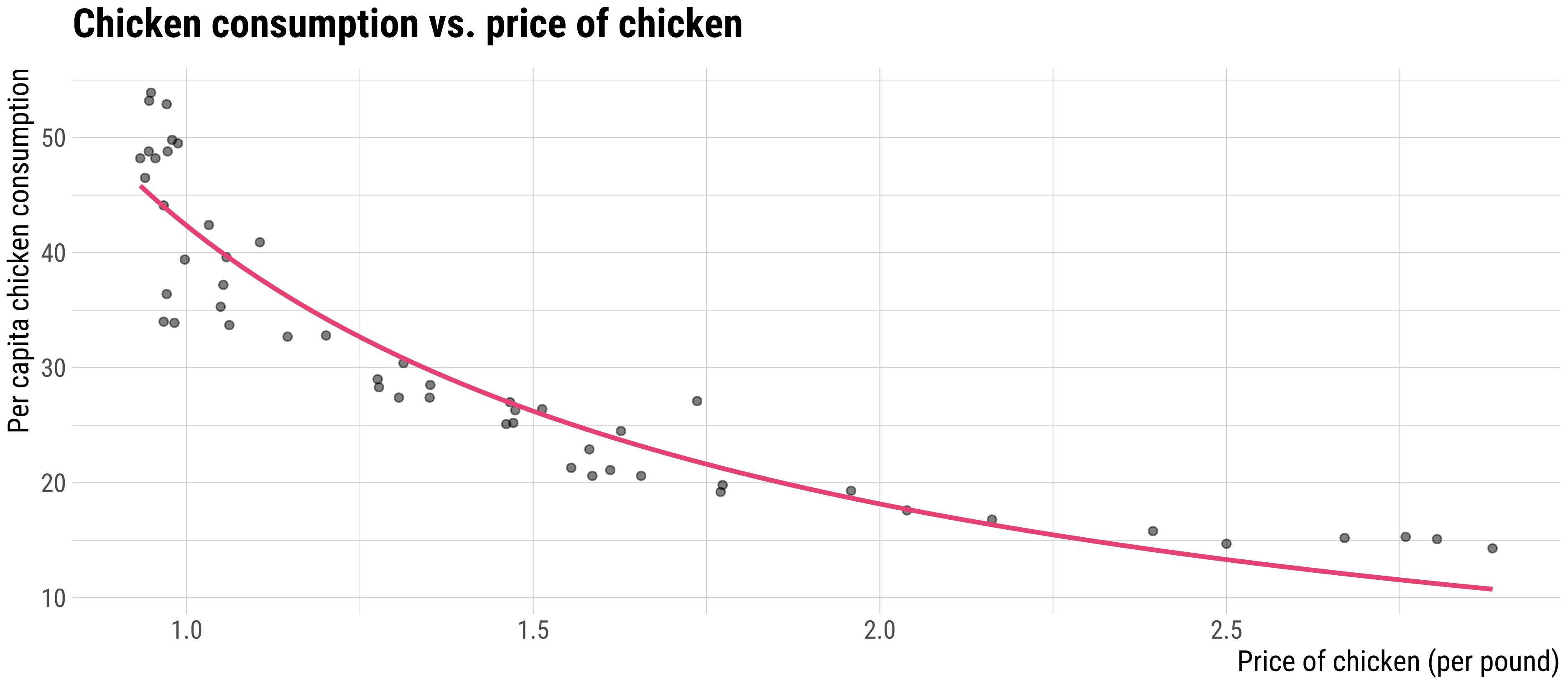
Inverse forms

$$qchicken_i = \beta_0 + \beta_1 \frac{1}{pchicken_i} + u_i$$



Inverse forms

$$qchicken_i = \beta_0 + \beta_1 \frac{1}{pchicken_i} + u_i$$



Inverse forms

Interpretation

$$y_i = \beta_0 + \beta_1 \frac{1}{x_{1i}} + u_i$$

$$\frac{\partial\, y}{\partial\, x_1} = \frac{-\beta_1}{x_1^2}$$

$$qchicken_i = \beta_0 + \beta_1 \frac{1}{pchicken_i} + u_i$$

$$\frac{\partial\, qchicken}{\partial\, pchicken} = \frac{-\beta_1}{pchicken^2}$$

Interaction terms

Interaction terms

Whenever the effect of one variable on y depends on the **level of another variable**, the best **modeling strategy** is to use *interaction terms*.

For example, do we believe that an individual's **wage** depends on their **education**?

- If so, is this effect the **same** or **different** for two individuals with, e.g., a *college* degree, but with different years of experience on the job market?

- Then, we represent a model by

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 educ_i \cdot exper_i + u_i$$

Interaction terms

In more general terms, regression estimates ($\hat{\beta}_i$) describe **average effects**.

Some of these average effects may "hide" **heterogeneous effects** that differ by **group** or by the **level of another variable**.

Interaction terms help us in modeling such **heterogeneous** effects.

- For instance, it is plausible to consider that returns on education will differ by *gender*, *race*, *region*, etc.

Interaction terms

Interpretation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + u_i$$

$$\frac{\partial\, y}{\partial\, x_1} = \beta_1 + \beta_3 \ \cdot \ x_2$$

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 educ_i \cdot exper_i + u_i$$

$$\frac{\partial\, wage}{\partial\, exper} = \beta_2 + \beta_3 \ \cdot \ educ$$

Binary variables

Binary variables

Categorical variables are used to translate **qualitative information** into **numbers**.

- For instance, *race*, *gender*, *being employed or not*, *enrolled in EC 339 or not*, etc.

The **easiest** way to work with qualitative information is by using **binary (*dummy*)** variables.

For example,

$$y_i = \beta_0 + \beta_1 D_i + u_i$$

where $D_i = 1$ if the criterion is fulfilled, and $D_i = 0$ otherwise.

Binary variables

When **interpreting** regression coefficients associated with *dummy* variables, the *intercept*'s interpretation changes slightly.

Moreover, the **slope** coefficient on D_i is not interpreted in the same way we are used to.

Consider:

$$interviews_i = \beta_0 + \beta_1 graduate_i + u_i$$

where

- $interviews_i$ is the number of interviews a candidate is called for in a given period;
- $graduate_i$ equals 1 if she has graduated from college, and 0 otherwise.

Binary variables

$$interviews_i = \beta_0 + \beta_1 graduate_i + u_i$$

For this model,

- β_0 is the expected number of interviews when $graduate_i = 0$ (non-graduates);
- β_1 is the expected **difference** in interview calls between graduates and non-graduates;
- And $\beta_0 + \beta_1$ is the expected number of interviews for graduates (when $graduate_i = 1$).

- In this case, *non-graduates* are the **reference group**.

Binary variables

$$interviews_i = \beta_0 + \beta_1 graduate_i + u_i$$

The model above is an example of an **intercept** *dummy* variable.

- We only have different **intercepts** when comparing two groups, but **slopes** are the same.

In order to allow for different **slopes**, we appeal to *interaction terms* involving categorical variables.

- i.e., **slope** *dummy* variables.

Log-Level Model

Important! If you have a **log-linear** model with a *binary* variable, the interpretation of the coefficient on that variable **changes**.

$$\log(y_i) = \beta_0 + \beta_1 D_i + u_i$$

with D being a *dummy* variable.

Interpretation of β_1:

- When $D = 1$, y will increase by $100 \times \left(e^{\beta_1} - 1\right)$ percent.
- When $D = 0$, y will decrease by $100 \times \left(e^{-\beta_1} - 1\right)$ percent.

Log-Level Example

Binary explanatory variable: `inlf`

- `inlf = 1` if the i^{th} individual is in the labor force.
- `inlf = 0` if the i^{th} individual is not in the labor force.

$$\widehat{log(sleep_i)} = 8.08 - 0.00365\, inlf_i$$

- How do we interpret the coefficient on `inlf`?
  - Labor force participants sleep `0.3665%` less than non-participants.
  - Individuals that are not in the labor force sleep `0.3692%`% more than participants.

Intercept *dummy* variables

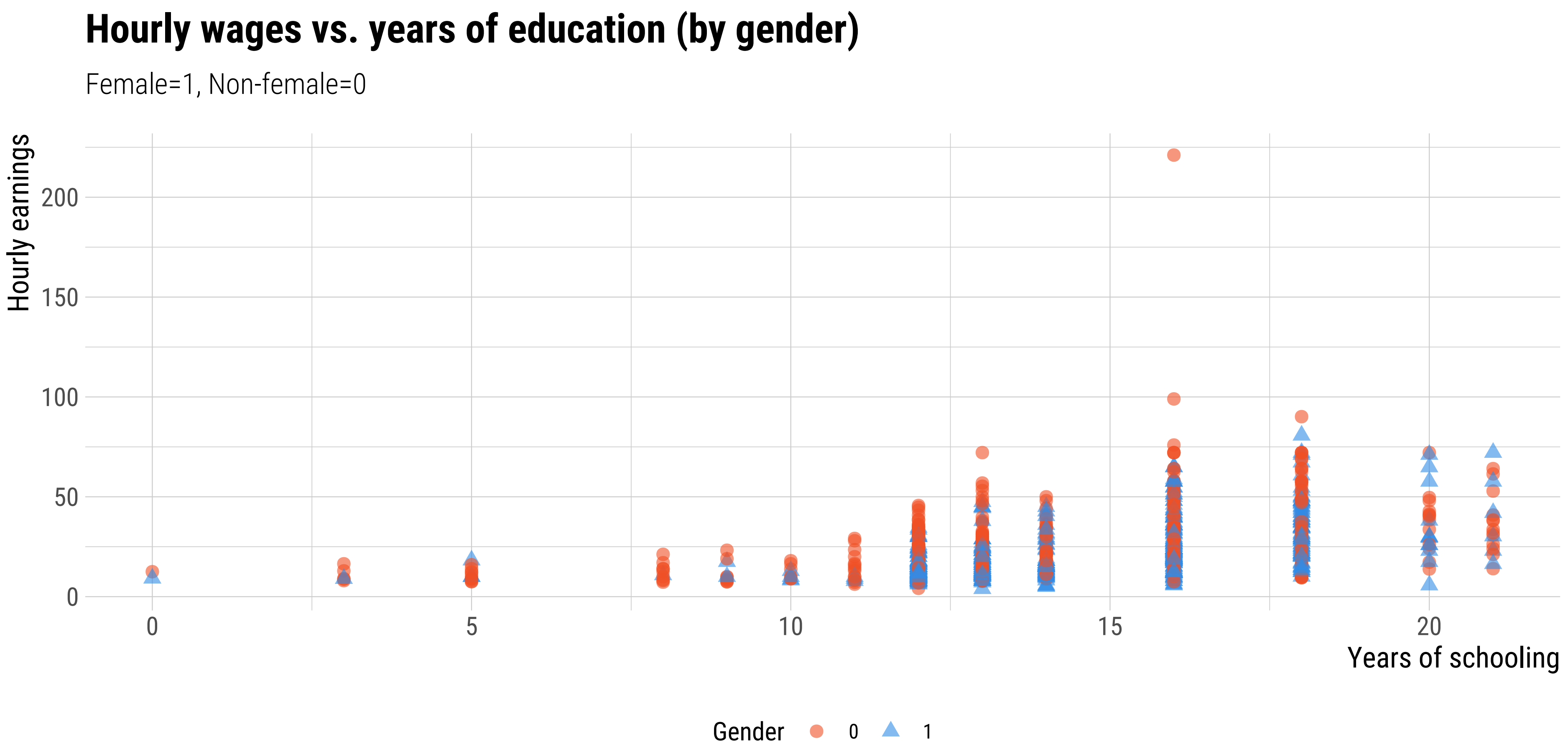
```
. reg lwage female educ exper tenure educ_tenure
```

Source	SS	df	MS
Model	58.8964115	5	11.7792823
Residual	89.4333399	520	.171987192
Total	148.329751	525	.28253286

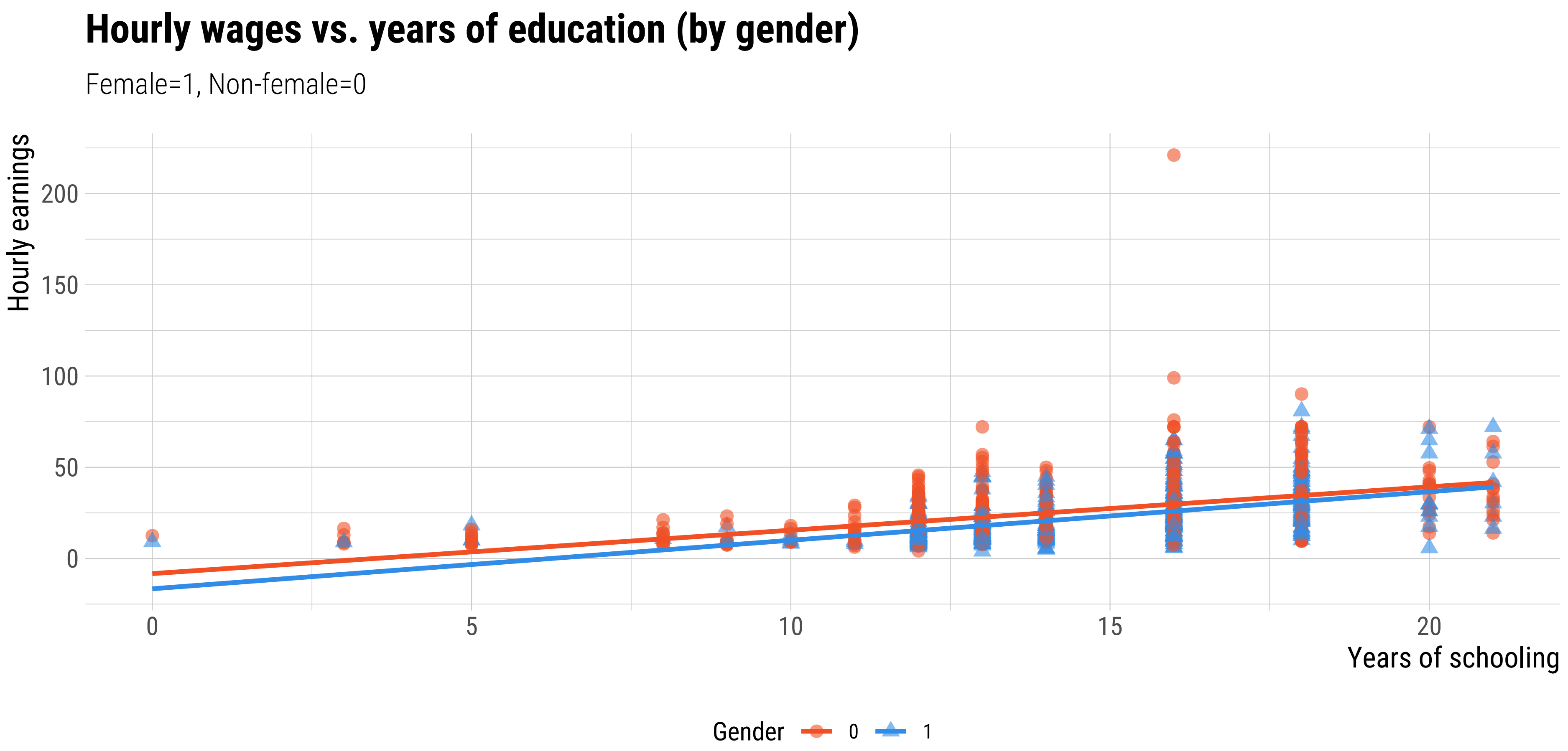
Number of obs	=	526
F(5, 520)	=	68.49
Prob > F	=	0.0000
R-squared	=	0.3971
Adj R-squared	=	0.3913
Root MSE	=	.41471

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
female	-.3045109	.0371709	-8.19	0.000	-.3775344	-.2314873
educ	.0769119	.0086477	8.89	0.000	.0599233	.0939005
exper	.0045324	.001623	2.79	0.005	.001344	.0077208
tenure	-.0023517	.0101413	-0.23	0.817	-.0222746	.0175712
educ_tenure	.001636	.0008046	2.03	0.043	.0000554	.0032166
_cons	.6347293	.120933	5.25	0.000	.3971521	.8723065

Slope *dummy* variables



Slope *dummy* variables



Slope *dummy* variables

Interpretation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 D_i + \beta_3 D_i x_{1i} + u_i$$

$$\frac{\partial\, y}{\partial\, x_1} = \beta_1 + \beta_3 \ \cdot \ D$$

$$\frac{\partial\, y}{\partial\, D} = \beta_2 + \beta_3 \ \cdot \ x_1$$

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 female_i + \beta_3 educ_i \cdot female_i + u_i$$

$$\frac{\partial\, wage}{\partial\, educ} = \beta_1 + \beta_3 \ \cdot \ female$$

$$\frac{\partial\, wage}{\partial\, female} = \beta_2 + \beta_3 \ \cdot \ educ$$

Slope *dummy* variables

```
. reg lwage female educ female_educ exper expersq tenure tenursq

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(7, 518)       =     58.37
       Model |  65.4081534         7  9.34402192   Prob > F        =    0.0000
    Residual |   82.921598       518  .160080305   R-squared       =    0.4410
-------------+----------------------------------   Adj R-squared   =    0.4334
       Total |  148.329751       525   .28253286   Root MSE        =     .4001

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      female |  -.2267886   .1675394    -1.35   0.176    -.5559289    .1023517
        educ |   .0823692   .0084699     9.72   0.000     .0657296    .0990088
 female_educ |  -.0055645   .0130618    -0.43   0.670    -.0312252    .0200962
       exper |   .0293366   .0049842     5.89   0.000      .019545    .0391283
     expersq |  -.0005804   .0001075    -5.40   0.000    -.0007916   -.0003691
      tenure |   .0318967    .006864     4.65   0.000      .018412    .0453814
     tenursq |    -.00059   .0002352    -2.51   0.012    -.001052    -.000128
       _cons |    .388806   .1186871     3.28   0.001     .1556388    .6219732
------------------------------------------------------------------------------
```

Next time: Functional forms in practice