Multiple Regression Analysis: Topics on functional forms

Chapter 5 (Ch. 6 of the textbook)

Wooldridge: Introductory Econometrics:
A Modern Approach, 5e

Multiple Regression Analysis: Topics on functional forms

- **<u>Using quadratic functional forms</u>**
- **Example: Wage equation**

Concave experience profile

$$\widehat{wage} = \underset{(.35)}{3.73} + \underset{(.041)}{.298}\ exper - \underset{(.0009)}{.0061}\ exper^2$$

$$n = 526, R^2 = .093$$

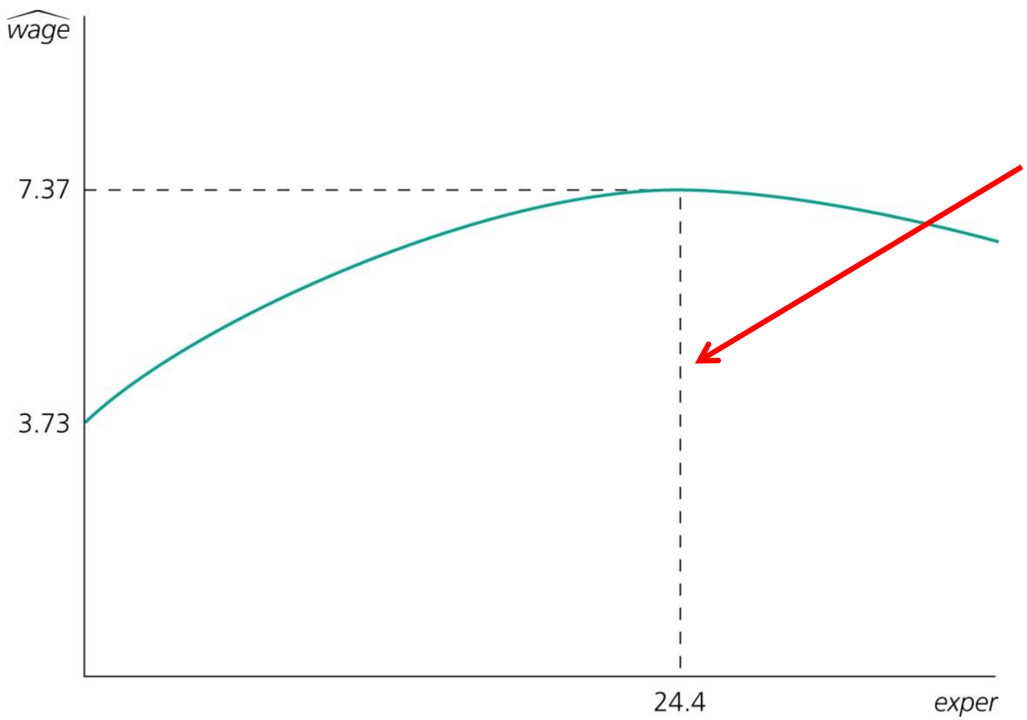
- **Marginal effect of experience**

The first year of experience increases the wage by some .30$, the second year by .298-2(.0061)(1) = .29$ etc.

$$\frac{\partial wage}{\partial exper} = .298 - 2(.0061) exper$$

Multiple Regression Analysis: Topics on functional forms

- **Wage maximum with respect to work experience**



$$x^* = \left|\frac{\widehat{\beta}_1}{2\widehat{\beta}_2}\right| = \left|\frac{.298}{2(.0061)}\right| \approx 24.4$$

Does this mean the return to experience becomes negative after 24.4 years?

Not necessarily. It depends on how many observations in the sample lie right of the turnaround point.

In the given example, these are about 28% of the observations. There may be a specification problem (e.g. omitted variables).

Multiple Regression Analysis: Topics on functional forms

Nitrogen oxide in aır, distance from employment centers, student/teacher ratio

- **Example: Effects of pollution on housing prices**

$$\widehat{\log}(price) = \underset{(.57)}{13.39} - \underset{(.115)}{.902} \log(nox) - \underset{(.043)}{.087} \log(dist)$$

$$- \underset{(.165)}{.545}\, rooms + \underset{(.013)}{.062}\, rooms^2 - \underset{(.006)}{.048}\, stratio$$

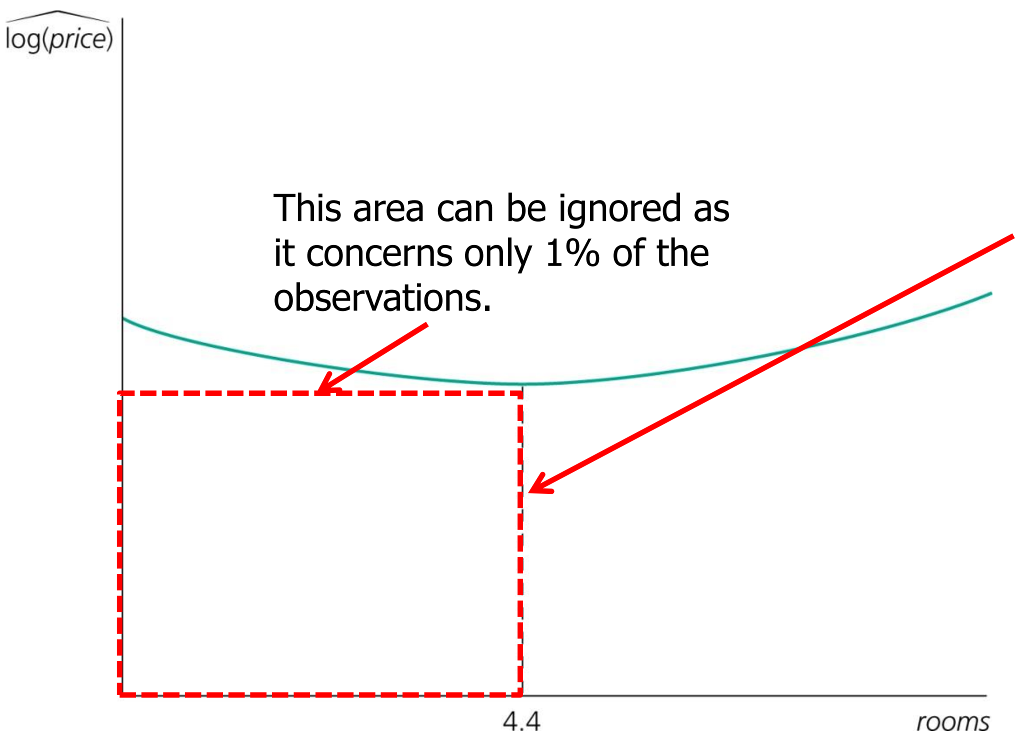
$$n = 506, R^2 = .603$$

Does this mean that, at a low number of rooms, more rooms are associated with lower prices?

$$\Rightarrow \quad \frac{\partial \log(price)}{\partial rooms} = \frac{\% \partial price}{\partial rooms} = -.545 + .124\, rooms$$

Multiple Regression Analysis: Topics on functional forms

- **Calculation of the turnaround point**



Turnaround point:

$$x^* = \left|\frac{-.545}{2(.062)}\right| \approx 4.4$$

Increase rooms from 5 to 6:

$$-.545 + .124(5) = +7.5\%\ price$$

Increase rooms from 6 to 7:

$$-.545 + .124(6) = +19.9\%\ price$$

Multiple Regression Analysis: Topics on functional forms

- **Other possibilities**

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 [\log(nox)]^2$$

$$+\beta_3 crime + \beta_4 rooms + \beta_5 rooms^2 + \beta_6 stratio + u$$

$$\Rightarrow \quad \frac{\partial \log(price)}{\partial \log(nox)} = \frac{\% \partial price}{\% \partial nox} = \beta_1 + 2\beta_2 [\log(nox)]$$

- **Higher polynomials**

$$cost = \beta_0 + \beta_1 quantity + \beta_2 quantity^2 + \beta_3 quantity^3 + u$$

Multiple Regression Analysis: Topics on functional forms

- **Models with interaction terms**

$$\log(price) = \beta_0 + \beta_1 sqrft + \beta_2 bdrms$$

$$+\beta_3 \boxed{sqrft \cdot bdrms} + \beta_4 bthrms + u$$

Interaction term

$$\Rightarrow \quad \frac{\partial \log(price)}{\partial bdrms} = \beta_2 + \beta_3 sqrft$$

The effect of the number of bedrooms depends on the level of square footage

- **Interaction effects complicate interpretation of parameters**

$\beta_2 =$ Effect of number of bedrooms, but for a square footage of zero

Multiple Regression Analysis: Topics on functional forms

- **Reparametrization of interaction effects**

Population means; may be replaced by sample means

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u$$

Effect of x_2 if all variables take on their mean values

- **Advantages of reparametrization**
  - Easy interpretation of all parameters
  - If necessary, interaction may be centered at other interesting values

Multiple Regression Analysis: Topics on functional forms

- **Example**

Dependent Variable: PRICE
Method: Least Squares
Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	165.4265	87.73015	1.885629	0.0628
LOTSIZE	0.001993	0.000628	3.173764	0.0021
SQRFT	0.033793	0.041462	0.815034	0.4174
BDRMS	-33.71534	22.82291	-1.477258	0.1434
BDRMS*SQRFT	0.021827	0.009663	2.258787	0.0265

Dependent Variable: PRICE
Method: Least Squares
Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	33.549	37.791	0.8878	0.377
LOTSIZE	0.002	0.0006	3.1738	0.002
SQRFT	0.099	0.0166	5.9825	0.000
BDRMS	10.244	8.9418	1.1456	0.255
(BDRMS-3)*(SQRFT-2014)	0.022	0.0097	2.2589	0.027

Multiple Regression Analysis: Topics on functional forms

- **<u>Tests for functional form misspecification</u>**
  - One can always test whether explanatory should appear as squares or higher order terms by testing whether such terms can be excluded
  - Otherwise, one can use general specification tests such as RESET
- **Regression specification error test (RESET)**
  - If the model is correctly specified then it includes all the relevant variables. If we add variables that are statistically significant then the model is misspecified.
  - The idea of RESET is to include squares and possibly higher order fitted values in the regression.

Multiple Regression Analysis: Topics on functional forms

- **Test procedure**

1. OLS of the initial model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + u \xrightarrow{\text{OLS}} \hat{y}$$

2. OLS of the auxiliary regression test:

i. **Option 1:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + \gamma \hat{y}^2 + v_1$

Test for the exclusion of this term, that is test $\gamma = 0$ against $\gamma \neq 0$

ii. **Option 2:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + \gamma_1 \hat{y}^2 + \gamma_2 \hat{y}^3 + v_2$

Test for the exclusion of these terms, that is test jointly $\gamma_1 = 0, \gamma_2 = 0$

2. Conclusion

If the terms cannot be excluded, (H_0 is rejected) this is evidence for omitted higher order terms and interactions, i.e., for **misspecification of functional form**.

Multiple Regression Analysis: Specification and Data Issues

- **Example: Housing price equation**

Dependent Variable: PRICE
Method: Least Squares
Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	221.5857	77.85786	2.846029	0.0056
LOTSIZE	-0.000947	0.001226	-0.772512	0.4420
SQRFT	-0.051335	0.064752	-0.792795	0.4301
FIT^2	0.001975	0.000683	2.893919	0.0048

Evidence for misspecification

With FIT the estimated PRICE in the regression on LOTSIZE and SQRFT