Topics on Functional Form, Wooldridge (2013), Chapter 6 (section 6.2) and Chapter 9 (section 9.1)

- Functional Form - The meaning of the term linear
- Quadratic Models
- Interaction Terms
- Tests of functional form
  - Ramsey's RESET Test
  - Nonnested Tests

A function $f(z_1, ..., z_J)$ is linear in $z_1, ..., z_J$ if it can be written in the following form

$$f(z_1, ..., z_J) = m_1 z_1 + m_2 z_2 + ... + m_J z_J + b$$

for some constants b and $m_1, ...m_J$.
That is, a function is linear if it can be written as a weighted sum of the arguments plus a constant.

Linearity in the Variables
The meaning of linearity in the variables is that the conditional expectation of y is a linear function of x, that is the regression curve in this case is a straight line. *Examples:*

$$E(y|x) = \beta_0 + \beta_1 x.$$

is linear in variables,but

$$E(y|x) = \beta_0 + \beta_1 x^2.$$

is not a linear function of x.

Further examples

1-

$$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

This function is linear in variables.

2-

$$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$

This function is non-linear in variables.

Linearity in the Parameters

The second interpretation of linearity is that the conditional expectation of y, $E(y|x)$, is a linear function of the parameters, the β's; it may or may not be linear in the variable x. **Examples:**

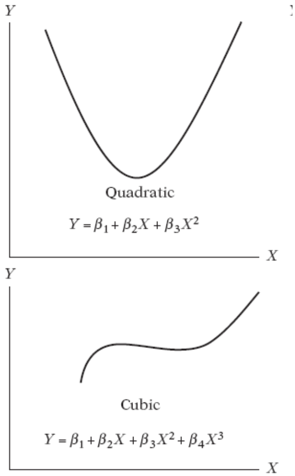
1. $$E(y|x) = \beta_0 + \beta_1 x^2.$$

   is a linear (in the parameters) regression model as it is a straight line (where the arguments now are β_0 and β_1).

2. $$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2,$$

   is a linear in the parameters

All the models shown in the figure below are linear regression models, that is, they are models linear in the parameters.



Now consider the model:

$$E(y|x) = \beta_0 + \beta_1^2 x.$$

- The preceding model is an example of a nonlinear (in the parameter) regression model.Why? Because it is a quadratic function in the parameters.
- The parameters of such model cannot be estimated using the ordinary least squares estimator.
- We have to use the non-linear least squares estimator

$$S^* \left(b_0, b_1\right) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - b_0 - b_1^2 x_i\right)^2$$

The term *"linear"* regression refers to a regression that is *linear in the parameters*; the β's (that is, the parameters are raised to the first power only).

LINEAR REGRESSION MODELS

Model linear in parameters?	Model linear in variables?	
	Yes	No
Yes	LRM	LRM
No	NLRM	NLRM

Note: LRM = linear regression model
NLRM = nonlinear regression model

- The ordinary least squares estimator can be used to study relationships that are not strictly linear in x and y by using nonlinear functions of x and y.
- An example considered before was the case that the dependent variable and/or regressors were in natural logs.
- Other popular nonlinear functions considered in empirical work are:
  - Quadratic forms of the regressors
  - Forms that include interactions of the regressors (cross-products).

Quadratic Models

For a model of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

we can't interpret β_1 alone as measuring the change in y with respect to x, we need to take into account β_2.
The estimated regression equation is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2.$$

Therefore

$$\frac{\partial \hat{y}}{\partial x} = \hat{\beta}_1 + 2\hat{\beta}_2 x.$$

Hence if x increases by 1, $\hat{y}$ increases by $\hat{\beta}_1 + 2\hat{\beta}_2 x$.

Example: We would like to study how wages are related with years of experience.
We have information on wages and experience for 526 people from the 1976 Current Population Survey (USA).
Running the regression of wages on experience and experience squared we obtain

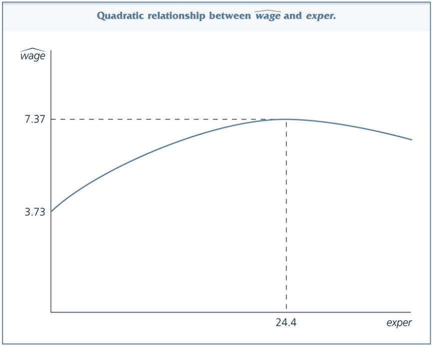
$$\begin{aligned}\widehat{wage} &= \underset{(0.35)}{3.73} + \underset{(0.041)}{0.298}\, exper - \underset{(0.0009)}{0.0061} exper^2, \\ R^2 &= 0.093,\end{aligned}$$

where the values in parentheses are the estimated standard errors. In this model

$$\frac{\partial\, \widehat{wage}}{\partial\, exper} = 0.298 - 2(0.0061) exper$$

Example:
Experience has a diminishing effect on wage:

$exper$	1	10	24.4	28
$\frac{\partial \widehat{wage}}{\partial exper}$	0.286	0.176	0.000	-0.047



Quadratic Models

Example:

- 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.

Example: (Effects of pollution on housing prices) Consider the model

$$\begin{aligned} \log(price) &= \beta_0 + \beta_1 \log(nox) + \beta_3 \log(dist) + \beta_4 rooms \\ &\quad + \beta_5 rooms^2 + \beta_6 stratio + u \end{aligned}$$

where

- *price*=median housing price of a community.
- *nox*=Nitrogen oxide air.
- *dist*=distance from from employment centres.
- *rooms*=average number of rooms
- *stratio*=student/teacher ratio.
- $n = 506$ communities in the Boston area

Quadratic Models

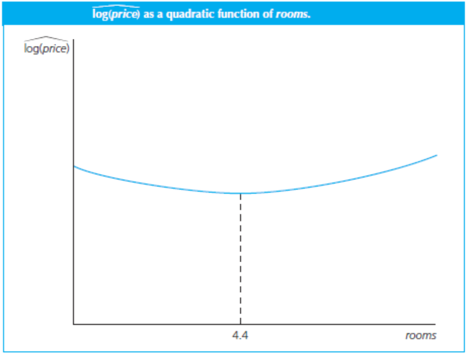
Estimating the model we obtain

$$\begin{aligned}
\widehat{\log(price)} &= \underset{(0.57)}{13.39} - \underset{(0.115)}{0.902}\log(nox) - \underset{(0.043)}{0.087}\log(dist) - \underset{(0.165)}{0.545}rooms \\
&\quad + \underset{(0.013)}{0.062}rooms^2 - \underset{(0.006)}{0.048}stratio, \\
R^2 &= 0.603
\end{aligned}$$

Hence

$$\frac{\partial \widehat{\log(price)}}{\partial rooms} = -0.545 + 2 \times 0.062 rooms$$

Quadratic Models



Turnaround point $rooms^* = \frac{0.545}{2 \times 0.062} = 4.4.$

Example:

- $\frac{\partial \widehat{\log(price)}}{\partial rooms} = -0.545 + 2 \times 0.062 \times 2 = -0.297$ if $rooms = 2$ →This is an odd result.
- Only 1% of the sample have houses averaging 4.4 rooms or less→We can ignore observations with $rooms \leq 4.4$
- $\frac{\partial \widehat{\log(price)}}{\partial rooms} = -0.545 + 2 \times 0.062 \times 5 = 0.075\,(7.5\%)$ if $rooms = 5$
- $\frac{\partial \widehat{\log(price)}}{\partial rooms} = -0.545 + 2 \times 0.062 \times 6 = 0.199\ (19.9\%)$ if $rooms = 6$

Remark: We can consider higher order polynomials

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + u$$

Interaction Terms

Sometimes we may want to allow the marginal effect of a regressor to vary with the level os some other regressor. In this case, the model of the form

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

We can't interpret β_1 alone as measuring the change in y with respect to x_1, we need to take into account β_3 as well.
The estimated equation is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1 x_2$$

Therefore

$$\frac{\partial \hat{y}}{\partial x_1} = \hat{\beta}_1 + \hat{\beta}_3 x_2.$$

Hence the interpretation is difficult. We have to evaluate it at particular values of x_2. For example, at the sample mean of $\bar{x}_2$.

- Original model

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

- New model

$$y = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u.$$

- μ_1 and μ_2 are population means. In practice they are replaced by sample means $\bar{x}_1$ and $\bar{x}_2$.
- We can show that $\delta_1 = \beta_1 + \beta_3 \mu_2$

- Notice that

$$\begin{aligned} y &= \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u. \\ &= \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 x_1 x_2 - \beta_3 x_1 \mu_2 - \beta_3 \mu_1 x_2 \\ &\quad + \beta_3 \mu_1 \mu_2 + u \\ &= \underbrace{\delta_0 + \beta_3 \mu_1 \mu_2}_{\beta_0} + \underbrace{(\delta_1 - \beta_3 \mu_2)}_{\beta_1} x_1 + \underbrace{(\delta_2 - \beta_3 \mu_1)}_{\beta_2} x_2 + \beta_3 x_1 x_2 + u \end{aligned}$$

- Therefore

$$\begin{aligned} \beta_1 &= \delta_1 - \beta_3 \mu_2, \\ \delta_1 &= \beta_1 + \beta_3 \mu_2 \end{aligned}$$

- Its estimate is $\bar{\delta}_1 = \hat{\beta}_1 + \hat{\beta}_3 \bar{x}_2$
- Advantages of reparametrization
- Easy interpretation of all parameters
- Standard errors for partial effects at the mean values available
- If necessary, interaction may be centered at other interesting values

Example

$$\log(price) = \beta_0 + \beta_1 bdrms + \beta_3 lotsize + \beta_4 sqrft + \beta_5 sqrft \times bdrms + u$$

where
price = house price, \$1000s
bdrms = number of bedrooms
lotsize = size of lot in square feet
sqrft = size of house in square feet
Sample: 88 observations collected from the real estate pages of the Boston Globe during 1990. These are homes that sold in the Boston, MA area.

Dependent variable $\log(price)$
$n = 88$

	Estimate	Std. Err.	t-Ratio
Intercept	5.0151932	0.2852878	17.5794155
bdrms	−0.0397661	0.0742173	−0.5358054
lotsize	0.0000055	0.0000020	2.6934041
sqrft	0.0002425	0.0001348	1.7986795
$sqrft \times bdrms$	0.0000298	0.0000314	0.9492329

$R^2 = 0.626333874$

Dependent variable $\log\left(price\right)$
$n = 88$

	Estimate	Std. Err.	t-Ratio
$Intercept$	4.8008736	0.1032982	46.4758498
$bdrms$	0.0202980	0.0290793	0.6980240
$lotsize$	0.0000055	0.0000020	2.6934041
$sqrft$	0.0003489	0.0000450	7.7595579
$\left(sqrft - \overline{sqrft}\right) \times \left(bdrms - \overline{bdrms}\right)$	0.0000298	0.0000314	0.9492329

$R^2 = 0.626333874$

$\overline{sqrft}$ -sample average of $sqrft$
$\overline{bdrms}$ -sample average of $bdrms$

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + u.$$

We've seen that a linear regression can really fit nonlinear relationships

- Can use logs on right hand side, left hand side or both.
- Can use quadratic forms of x's.
- Can use interactions of x's.
- How do we know if we've got the right functional form for our model?

- First, use economic theory to guide you.
- Think about the interpretation.
- Does it make more sense for x to affect y in percentage (use logs)?
- Does it make more sense for the derivative of y with respect to x_1 to vary with x_1 (quadratic) or with x_2 (interactions) or to be fixed?
- We already know how to test joint exclusion restrictions to see if higher order terms or interactions belong in the model.
- It can be tedious to add and test extra terms, plus may find a square term matters when really using logs would be even better.
- A test of functional form is Ramsey's regression specification error test (*RESET*)

The idea of RESET is to include squares and possibly higher order powers of the fitted values in the regression.
We can estimate:

- $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \delta_1 \hat{y}^2 + error$ and test $H_0 : \delta_1 = 0$ using the t statistic.
- Why should we use $\hat{y}^2$?
- Because $\hat{y}^2$ is a function of the regressors, their squares and the cross-products of the regressors.
- To see this notice that if $k = 2$, for instance

$$\begin{aligned} \hat{y}_i^2 &= \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}\right)^2 \\ &= \hat{\beta}_0^2 + \hat{\beta}_1^2 x_{i1}^2 + \hat{\beta}_2^2 x_{i2}^2 + 2\hat{\beta}_0\hat{\beta}_1 x_{i1} + 2\hat{\beta}_0\hat{\beta}_2 x_{i2} \\ &\quad + 2\hat{\beta}_1\hat{\beta}_2 x_{i1} x_{i2} \end{aligned}$$

- We can also use the cube of $\hat{y}$: We estimate

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + error$$

and test $H_0 : \delta_1 = 0, \delta_2 = 0$ using the F or LM statistic.

Example: Housing Price Equation
Consider the following two models for housing prices:
1- $price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u.$
$n = 88$

- Running the regression of *price* on *lotsize* ,*sqrft* and *bdrms* we obtain $R^2 = 0.67236$
- Running the regression of *price* on *lotsize* ,*sqrft* and *bdrms*, $\widehat{price}^2$ and $\widehat{price}^3$ we obtain $R^2 = 0.70585$.

2- $\log(price) = \beta_0 + \beta_1 \log(lotsize) + \beta_2 \log(sqrft) + \beta_3 \log(bdrms) + u.$

- Running the regression of $\log(price)$ on $\log(lotsize)$,$\log(sqrft)$ and $\log(bdrms)$ we obtain $R^2 = 0.63937$.
- Running the regression of of $\log(price)$ on $\log(lotsize)$,$\log(sqrft)$, $\log(bdrms)$, $\widehat{\log(price)}^2$ and $\widehat{\log(price)}^3$ we obtain $R^2 = 0.66248$.

Which is the preferred model?

Example: H_0 : Model 1 is correctly specified vs H_1 : Model 1 is misspecified

- We need to use the F-statistic:

$$F = \frac{\left(R^2_{ur} - R^2_r\right)/q}{\left(1 - R^2_{ur}\right)/\left(n-k-1\right)} \sim F(q, n-k-1)$$

  where R^2_r is the R^2 of the restricted model and R^2_{ur} is the R^2 of the unrestricted model.
- $R^2_{ur} = 0.70585$
- $R^2_r = 0.67236$
- $q = 2$
- $k = 5$
- $n - k - 1 = 88 - 5 - 1 = 82$
- $F^{act} = \frac{(0.70585-0.67236)/2}{(1-0.70585)/(88-5-1)} = 4.668$
- $f_{0.05} = 3.107$ (based on $F \sim F(2, 82)$
- $f_{0.05} = 3.1$ (based on $F \sim F(2, 90)$ - closest df in the book
- $4.668 > 3.107$, therefore we reject H_0 in favour of H_1 at 5% level.

Example: H_0 : Model 2 is correctly specified vs H_1 : Model 2 is misspecified

- We need to use the F-statistic:

$$F = \frac{\left(R^2_{ur} - R^2_r\right)/q}{\left(1 - R^2_{ur}\right)/\left(n-k-1\right)} \sim F(q, n-k-1)$$

  where R^2_r is the R^2 of the restricted model and R^2_{ur} is the R^2 of the unrestricted model.
- $R^2_{ur} = 0.66248$
- $R^2_r = 0.63937$
- $q = 2$
- $k = 5$
- $n - k - 1 = 88 - 5 - 1 = 82$
- $F^{act} = \frac{(0.66248 - 0.63937)/2}{(1-0.66248)/(88-5-1)} = 2.8073$
- $f_{0.05} = 3.107$ (based on $F \sim F(2,82)$)
- $f_{0.05} = 3.1$ (based on $F \sim F(2,90)$) - closest df in the book
- $2.8073 < 3.107$, therefore we do not reject H_0 in favour of H_1 at 5% level.
- Therefore we prefer model 2

- If the models have the same dependent variables, but nonnested x's could still just make a giant model with the x's from both and test joint exclusion restrictions that lead to one model or the other, approach suggested by Mizon and Richard (1986)
- We have two competing models:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$

against

$$y = \beta_0^* + \beta_1^* f(x_1) + \beta_2^* f(x_2) + u \tag{2}$$

- Estimate by OLS a comprehensive model

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 f(x_1) + \gamma_4 f(x_2) + u$$

- Use F test to test $H_0 : \gamma_3 = \gamma_4 = 0$ as a test of model 1, or
- Use F test to test $H_0 : \gamma_1 = \gamma_2 = 0$ as a test of model 2.

- The problem with the comprehensive approach: when we have many regressors, the power of the test is low.
- An alternative, the *Davidson-MacKinnon (1981) test*, uses the fitted values $\hat{y}$ from one model as regressor in the second model and tests for significance.
- In any case, Davidson-MacKinnon test may reject neither or both models rather than clearly preferring one specification.

Davidson-MacKinnon (1981) test against nonnested alternatives:

- We have two competing models:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + u \tag{3}$$

  against

$$y = \beta_0^* + \sum_{i=1}^{k} \beta_i^* f(x_i) + u \tag{4}$$

- To test model 3 against model 4, first estimate model 4 by OLS to obtain the fitted values $\widehat{\widehat{y}}$
- Estimate by OLS the model

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \theta \widehat{\widehat{y}} + u$$

- The rejection of $H_0 : \theta = 0$ (against a two-sided alternative) leads to the rejection of model 3.

Example: Let us consider the a sample taken from the 1976 US Current Population Survey ($n = 526$).Consider the models

1- $\log(wage) = \beta_0 + \beta_1 exper + u$

2- $\log(wage) = \beta_0^* + \beta_1^* \log(exper) + v$

Which is the most appropriate model?

Example: We run the regression of $\log(wage)$ on *exper* and compute the fitted values ($\hat{y}$). Running the regression of $\log(wage)$ on $\log(exper)$ and $\hat{y}$ we obtain

$$\log(wage) = \underset{(1.27826)}{8.36802} + \underset{(0.04719)}{0.35034} \log(exper) - \underset{(0.84937)}{4.67182}\hat{y}$$

Do you reject model 2 in favour of model 1?
We run the regression of $\log(wage)$ on $\log(exper)$ and compute the fitted values ($\widehat{\hat{y}}$). Running the regression of $\log(wage)$ on *exper* and $\widehat{\hat{y}}$ we obtain

$$\log(wage) = \underset{(0.59986)}{-2.89653} - \underset{(0.0037)}{0.02038} exper + \underset{(0.40384)}{2.998}\ \widehat{\hat{y}}$$

Do you reject model 1 in favour of model 2?

Example:

- We need to use the statistic

$$t = \frac{\hat{\theta}}{se(\hat{\theta})} \sim t(n-k-1)$$

  where $\hat{\theta}$ is the OLS estimator of θ and $se(\hat{\theta})$ is the corresponding standard error.
- H_0 : Model 2 is correct vs H_1 : Model 1 is correct
- $t^{act} = \frac{-4.67182}{0.84937} = -5.5003.$
- $n - k - 1 = 526 - 2 - 1 = 523$
- $t_{0.025} = 1.96.$
- $|-5.5003| = 5.5003 > 1.96$, therefore we reject H_0 in favour of H_1 at 5% level.

Example:

- We need to use the statistic

$$t = \frac{\hat{\theta}}{se(\hat{\theta})} \sim t(n-k-1)$$

  where $\hat{\theta}$ is the OLS estimator of θ and $se(\hat{\theta})$ is the corresponding standard error.
- H_0 : Model 1 is correct vs H_1 : Model 2 is correct
- $t^{act} = \frac{2.998}{0.40384} = 7.4237$
- $n - k - 1 = 526 - 2 - 1 = 523$
- $t_{0.025} = 1.96$.
- $|7.4237| = 7.4237 > 1.96$, therefore we reject H_0 in favour of H_1 at 5% level.