Kurt Schmidheiny
Universität Basel

Functional Form in the Linear Model

matrix-free

1 Introduction

Despite its name, the classical *linear* regression model, is not limited to a linear relationship between the dependent and the independent variables.

Consider K variables $x_{i1}\ x_{i2}\ ...\ x_{iL}$ for each observation i. The L functions $f_1(x_{i1}...\ x_{iL}), f_2(x_{i1}...\ x_{iL}), ..., f_L(x_{i1}...\ x_{iL})$ map the K variables into L new variables $z_{1i},\ z_{2i},\ ...,\ z_{Li}$. The function $g(y_i)$ is a function of the dependent variable. The non-linear econometric model

$$g(y_i) = \beta_0 + \beta_1 f_1(x_{i1}...\ x_{iL}) + \cdots + \beta_L f_L(x_{i1}...\ x_{iL}) + u_i$$

can therefore be written as

$$\begin{aligned} g(y_i) &= \beta_0 + \beta_1 z_{1t} + \beta_2 z_{2t} + \cdots + \beta_L z_{iL} + u_i \\ &= z_i'\beta + u_i\ . \end{aligned}$$

The latter is the usual multiple linear regression model with $L+1$ regressors as long as all necessary assumptions about the error term and the *transformed* independent variables $z_i = (1 z_{i1}\ z_{i2}\ ...\ z_{iL})$ are satisfied. All properties of OLS are therefore preserved.

Note: While the original model is potentially *non-linear in the variables* x, it is *linear in the parameters* β. Also note that the error term u_i is already additive in the original model.

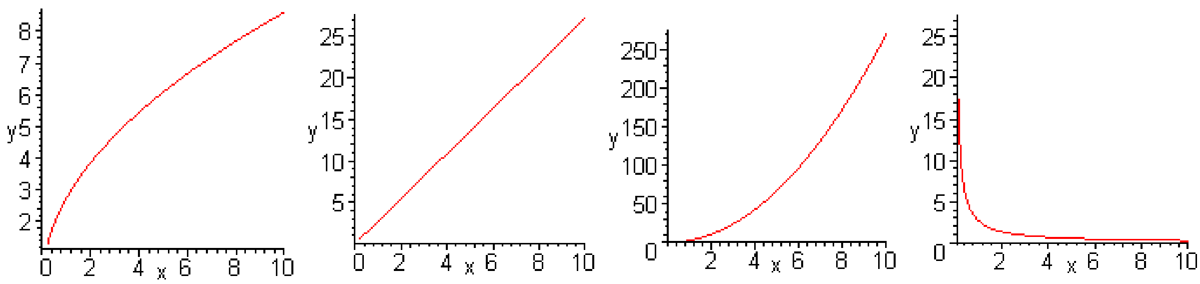
Version: 11-4-2012, 21:37

2 Some Examples

2.1 Log-Linear

Functional form:

$$\ln y_i = \beta_1 + \beta_2 \ln x_i + u_i$$



$\beta_1 = 1, \beta_2 = 0.5$ $\beta_1 = 1, \beta_2 = 1$ $\beta_1 = 1, \beta_2 = 2$ $\beta_1 = 1, \beta_2 = -1$

Marginal effect of x on y, deterministic (omitting index i):

$$\frac{\partial y}{\partial x} = \mathrm{e}^{\beta_1 + \beta_2 \ln x} \beta_2 \frac{1}{x} = \beta_2 x^{\beta_2 - 1} \mathrm{e}^{\beta_1}$$

The marginal effect depends on the value of the independent variable.

Note: this is the effect in the deterministic part of the model and not generally equal to $\partial Ey/\partial x$ as $E(log(y)) \neq log(E(y))$ (see Wooldridge p. 211).

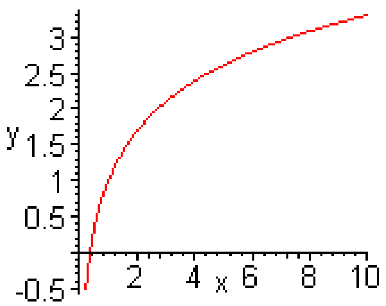
Elasticity of y w.r.t. x, deterministic:

$$\frac{d \ln y}{d \ln x} = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{dy/y}{dx/x} = \beta_2$$

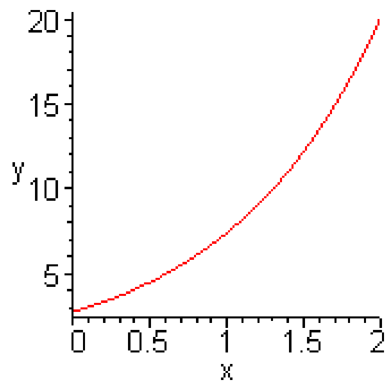
2.2 Semi-log

Functional form (two alternatives):

$$y_i = \beta_1 + \beta_2 \ln x_i + u_i \qquad \ln y_i = \beta_1 + \beta_2 x_i + u_i$$



$\beta_1 = 1, \beta_2 = 1$



$\beta_1 = 1, \beta_2 = 1$

Marginal effects of x on Ey and y, respectively:

$$\frac{\partial Ey}{\partial x} = \frac{\beta_2}{x} \qquad \frac{\partial y}{\partial x} = \beta_2 \mathrm{e}^{(\beta_1 + \beta_2 x)}$$

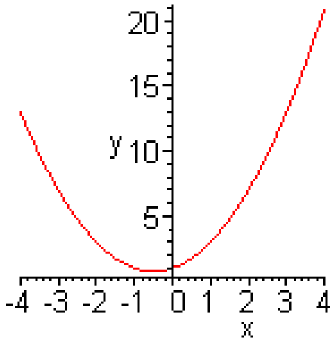
"Percentage effects" of x on Eyx/y, respectively:

$$\frac{\partial Ey}{\partial lnx} = \frac{\partial Ey}{\partial x / x} = \beta_2 \qquad \frac{\partial lny}{\partial x} = \frac{\partial y / y}{\partial x} = \beta_2$$

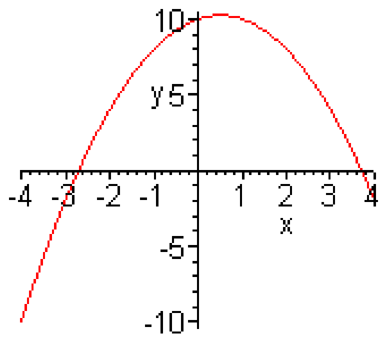
2.3 Polynomial

Functional form (e.g. order 3):

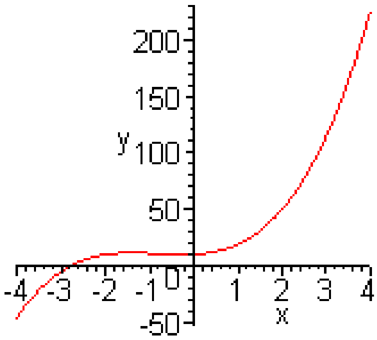
$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \beta_4 x_i^3 + u_i$$



$\beta_1 = 1, \beta_2 = 1$
$\beta_3 = 1, \beta_4 = 0$



$\beta_1 = 10, \beta_2 = 1$
$\beta_3 = -1, \beta_4 = 0$



$\beta_1 = 10, \beta_2 = 2$
$\beta_3 = 5, \beta_4 = 2$

Marginal effect of x on Ey:

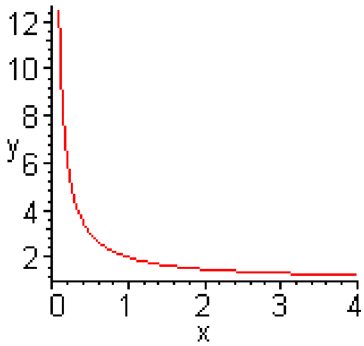
$$\frac{\partial Ey}{\partial x} = \beta_2 + 2\beta_3 x + 3\beta_4 x^2$$

Note that the marginal effect depends on the value of the independent variable. The individual parameters β_2, β_3 and β_4 have often now meaning of their own.

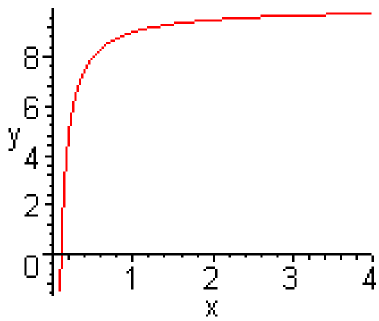
2.4 Inverse

Functional form:

$$y_i = \beta_1 + \beta_2 \frac{1}{x_i} + u_i$$



$\beta_1 = 1, \beta_2 = 1$



$\beta_1 = 10, \beta_2 = -1$

Marginal effect of x on Ey:

$$\frac{\partial Ey}{\partial x} = -\frac{\beta_2}{x^2}$$

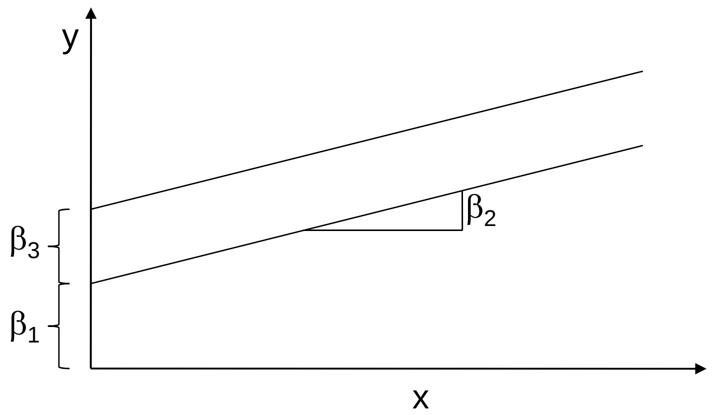
Note that a positive sign of β_2 means a negative relationship and vice-versa.

2.5 Dummy Variables

Functional form:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 d_i + u_i$$

where the dummy variable d_i is either 1 or 0.



Note that the marginal effects for the two groups (implicitly defined by the dummy variable) are equal but the constant terms differ.

2.6 Interaction Terms

Functional form:

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \beta_4 (x_{1i} \cdot x_{2i}) + u_i$$

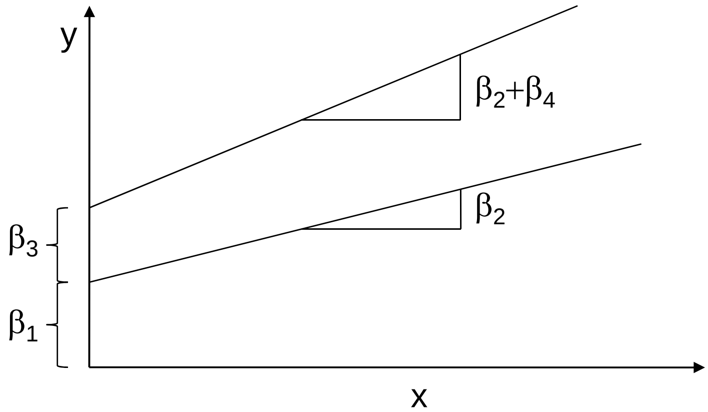
Marginal effects of x_1 and x_2 on Ey:

$$\frac{\partial Ey}{\partial x_1} = \beta_2 + \beta_4 x_2 \quad \text{and} \quad \frac{\partial Ey}{\partial x_2} = \beta_3 + \beta_4 x_1$$

The interpretation of these effects and of the individual parameters is very specific to theoretical model behind the relationship.

Special case, interaction with a dummy variable d_i:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 d_i + \beta_4 (x_i \cdot d_i) + u_i$$



Marginal effect of x on Ey:

$$\frac{\partial Ey}{\partial x_1} = \begin{cases} \beta_2 & \text{if } d_i = 0 \\ \beta_2 + \beta_4 & \text{if } d_i = 1 \end{cases}$$

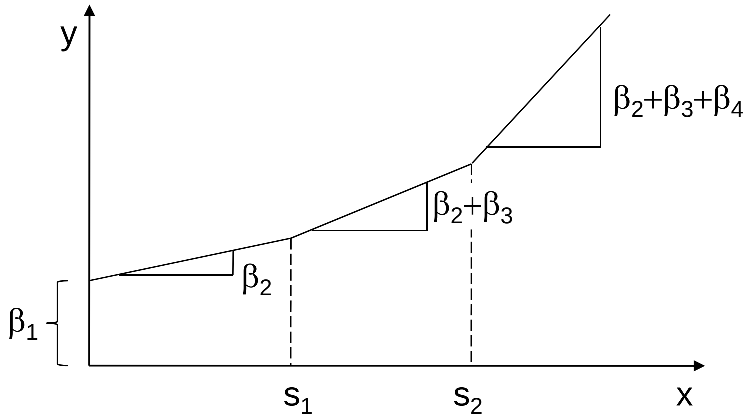
The interaction of all variables (here constant and one independent variable x) with the dummy variable allows to estimate separate linear relationships for the two groups defined by d_i. However this is different from two separate regression models as the error term is assumed to have identical variance across groups.

2.7 Spline Functions

Functional form:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 d_{1i}(x_i - s_1) + \beta_4 d_{2i}(x_i - s_2) + u_i$$

where $d_{1i} = 1$ if $x_i \geq s_1$ and $d_{2i} = 1$ if $x_i \geq s_2$. s_1 and s_2 are known thresholds.



Marginal effect of x on Ey:

$$\frac{\partial Ey}{\partial x_1} = \begin{cases} \beta_2 & \text{if } x_i < s_1 \\ \beta_2 + \beta_3 & \text{if } s1 < x_i < s_2 \\ \beta_2 + \beta_3 + \beta_4 & \text{if } x_i > s_1 \end{cases}$$

3 Implementation in Stata 12.0

Non-linear functional forms can be estimated with OLS by generating the transformed variables. For example,

```
webuse auto7.dta
generate mpg2 = mpg^2
reg price mpg mpg2
```

estimates a second order polynomial.

Dummy variables are easily created from categorical variables with the `xi` command. For example,

```
xi i.manufacturer
reg price _Imanufactu_*
```

creates 23 dummy variables for the 24 categories in the variable `manufacturer` (excluding the first one for use as reference category) and regresses `price` on all 23 dummy variables plus a constant. This can also be done in one step,

```
xi: reg price i.manufacturer
```

Interactions with dummy variables are also directly created with the `xi` command. For example,

```
xi: reg price i.foreign*mpg
```

estimates separate slopes and intercepts for foreign and domestic cars. As of version 11, dummy variables and interactions can also be formed as "factor variables". The above example is then

```
reg price i.foreign##c.mpg
```

The variables used for spline functions are conveniently created with the `mkspline` command. For example,

```
mkspline mpg_1 20 mpg_2 25 mpg_3 = mpg
reg price mpg_*
```

regresses `price` on `mpg` using a piecewise linear function. Also consider the option `marginal`.

4 Implementation in R

Non-linear functional forms can be estimated with OLS by specifying the functional form in the estimated model. For example,

```
> data(mtcars)
> lm(mpg~wt+I(wt^2), data=mtcars)
```

estimates a second order polynomial for the variable `wt`. Note that most mathematical functions need to be wrapped within the `I()` function.

Categorical variables are automatically included as a set of dummy variables if they are defined as factor variables. For example,

```
> mtcars$carbf <- factor(carb)
> lm(mpg~wt+carbf, data=mtcars)
```

regresses `mpg` on `wt` and on 5 dummy variables for 5 categories in `carbf` (excluding the first category for use as reference group) plus a constant.

Interactions with dummy variables from categorical variables can be directly estimated when the categorical variable is defined as factor variable. For example,

```
> mtcars$amf <- factor(mtcars$am, labels=c("automatic", "manual"))
> lm(mpg~amf+wt:amf, data=mtcars)
```

estimates separate slopes of `wt` and intercepts for cars with automatic and manual transmission. Altenatively,

```
> summary(lm(mpg~amf+wt/am, data=mtcars))
```

reports the difference between the two slopes. This is equivalent to

```
> lm(mpg~am+wt+wt:am, data=mtcars)
```

which does not use factor variables.

Linear (and polynomial) spline functions are implemented in the `splines` package. See the help for details,

```
> library("splines")
> ?splines
```

References

Stock, James H. and Mark W. Watson (2007), Introduction to Econometrics, 2nd ed., Pearson Addison-Wesley. Chapter 8.

Wooldridge, Jeffrey M. (2009), Introductory Econometrics: A Modern Approach, 4th ed. South-Western. Section 6.2 and 6.4.

Jaccard James and Robert Turrisi (2003), Interaction Effects in Multiple Regression, 2nd ed., Quantitative Applications in the Social Sciences 07-72, Sage.

Kennedy, Peter (2003), A Guide to Econometrics, 5th ed., Blackwell Publishing, chapter 7.