Econometrics Lecture Notes III

Multiple Regression, Further Analysis (close to chapters 6-7, Wooldridge)

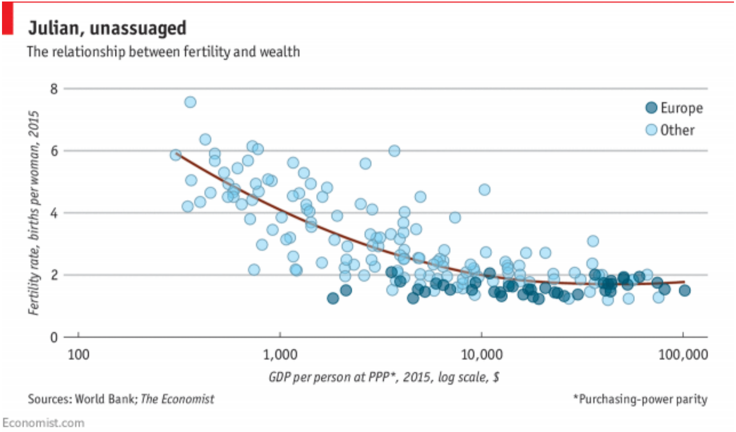
M Loecher

Matrix Notation

Overfitting

Nonlinearities

Nonlinearity I

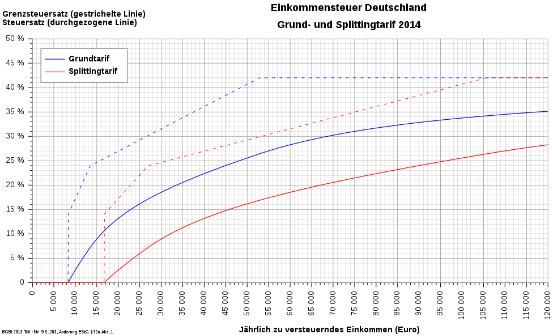


Nonlinearity II

Tax rate in Germany

Grundtabelle 2018 mit 9% KiSt

Zu verst. Einkommen	Einkommen-Steuer	Durchschn. Steuersatz	Grenz-Steuersatz
10.000,00	149,00	1,49 %	16,00 %
14.000,00	949,00	6,78 %	23,97 %
18.000,00	1.943,00	10,79 %	25,73 %
22.000,00	3.008,00	13,67 %	27,49 %
26.000,00	4.143,00	15,93 %	29,26 %
30.000,00	5.348,00	17,83 %	31,02 %
34.000,00	6.624,00	19,48 %	32,78 %
38.000,00	7.970,00	20,97 %	34,54 %
42.000,00	9.387,00	22,35 %	36,30 %
46.000,00	10.874,00	23,64 %	38,06 %
50.000,00	12.432,00	24,86 %	39,82 %
54.000,00	14.060,00	26,04 %	41,58 %
58.000,00	15.738,00	27,13 %	42,00 %
62.000,00	17.418,00	28,09 %	42,00 %
66.000,00	19.098,00	28,94 %	42,00 %



Nonlinearity III

Beer & Chips

Time with friends & kids

Energy Consumption & Temperature

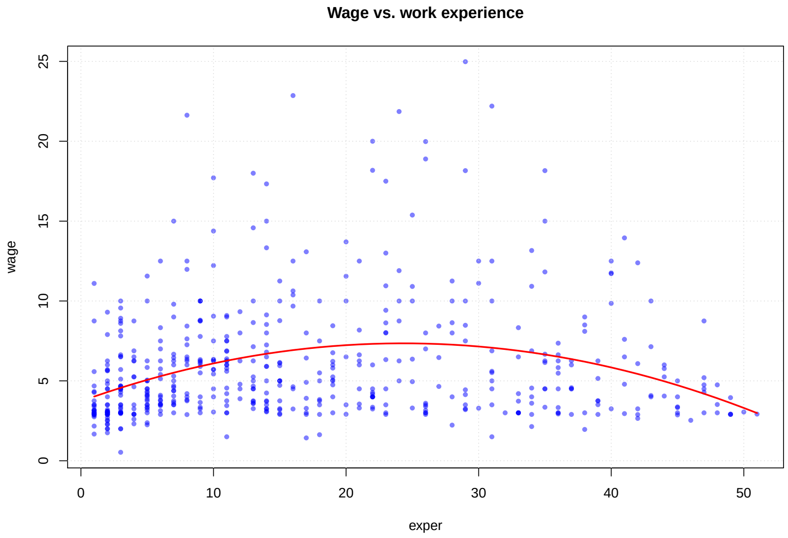
Using quadratic functional forms

Wage Equation

$$\widehat{\text{wage}} = \underset{(0.34594)}{3.72541} + \underset{(0.040966)}{0.298100}\,\text{exper} - \underset{(0.00090252)}{0.00612989}\,\text{exper}^2 \tag{1}$$

Marginal effect of experience

$$\frac{\partial wage}{\partial exper} = \hat{\beta}_1 + 2 \cdot \hat{\beta}_2 exper = 0.298 + 2 \cdot -0.006 \cdot exper$$



Wage vs. work experience
wage
exper
0
5
10
15
20
25
0
10
20
30
40
50

Interactions

Housing Prices

Model 1: OLS, using observations 1–506
Dependent variable: lprice

	Coefficient	Std. Error	t-ratio	p-value
const	9.555	0.257	37.118	0.000
crime	−0.013	0.002	−8.034	0.000
lproptax	−0.179	0.042	−4.272	0.000
nox	−0.048	0.013	−3.597	0.000
rooms	0.280	0.017	16.286	0.000

Mean dep var	9.94	S.D. dep var	0.41
Sum sq resid	32.83	S.E. of regr	0.26
R^2	0.61	Adjusted R^2	0.61
$F(4, 501)$	197.41	P-value(F)	0
Log-likelihood	−26.00186	Akaike criterion	62.00373
Schwarz criterion	83.14	Hannan–Quinn	70.29

Housing Prices w. interaction

Model 1: OLS, using observations 1–506
Dependent variable: lprice

	Coefficient	Std. Error	t-ratio	p-value
const	9.666	0.249	38.875	0.000
crime	0.131	0.023	5.727	0.000
lproptax	−0.236	0.041	−5.697	0.000
nox	−0.011	0.014	−0.808	0.419
rooms	0.284	0.017	17.109	0.000
crimeNOX	−0.021	0.003	−6.303	0.000

Mean dep var	9.94	S.D. dep var	0.41
Sum sq resid	30.42	S.E. of regr	0.25
R^2	0.64	Adjusted R^2	0.64
$F(5, 500)$	178.09	P-value(F)	0
Log-likelihood	−6.655958	Akaike criterion	25.31192
Schwarz criterion	50.67	Hannan–Quinn	35.26

Rescaling

Very difficult to interpret these coefficients: $\hat{\beta}_1 = 0.131$ effect of crime but at zero level of nox !

Reparametrization of interaction effects

$$y_i = \beta_0 + \beta_1 \cdot x_{i,1} + \beta_2 \cdot x_{i,2} + \beta_3 \cdot x_{i,1} \cdot x_{i,2} + u_i$$
$$y_i = \alpha_0 + \delta_1 \cdot x_{i,1} + \delta_2 \cdot x_{i,2} + \beta_3 \cdot (x_{i,1} - \mu_1) \cdot (x_{i,2} - \mu_2) + u_i$$

Now δ_2 is the effect of x_2 if all variables are at their mean values

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

Housing Prices w. interaction, II

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.09422    0.26235   38.48  < 2e-16 ***
## crime        0.01219    0.00426    2.86   0.0044 **
## lproptax    -0.23595    0.04142   -5.70  2.1e-08 ***
## nox         -0.08857    0.01433   -6.18  1.3e-09 ***
## rooms        0.28378    0.01659   17.11  < 2e-16 ***
## crime_nox   -0.02138    0.00339   -6.30  6.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.247 on 500 degrees of freedom
## Multiple R-squared:  0.64,   Adjusted R-squared:  0.637
## F-statistic:  178 on 5 and 500 DF,  p-value: <2e-16
```

Prediction Uncertainty

Variance of the Estimators

$$[2.57] \quad Var(\hat{\beta}_1) = \sigma^2 / SST_x \tag{2}$$

$$[2.58] \quad Var(\hat{\beta}_0) = \sigma^2 \cdot \overline{x^2} / SST_x \tag{3}$$

Prediction Error

Prediction at x_0:

$$
\begin{aligned}
[2.9] \qquad y_0 &= \beta_0 + \beta_1 x_0 + u_0 && (4)\\
[2.20] \qquad \hat{y}_0 &= \hat{\beta}_0 + \hat{\beta}_1 x_0 && (5)\\
[2.21] \Rightarrow \quad \hat{y}_0 - y_0 &= (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) x_0 - u_0 && (6)
\end{aligned}
$$

Qualitative Information

Titanic

Model 1: OLS, using observations 1–891
Dependent variable: Survived

	Coefficient	Std. Error	t-ratio	p-value
DPclass_1	0.630	0.031	20.198	0.000
DPclass_2	0.473	0.034	13.999	0.000
DPclass_3	0.242	0.021	11.722	0.000

Mean dep var	0.38	S.D. dep var	0.49
Sum sq resid	186.39	S.E. of regr	0.46
R^2	0.12	Adjusted R^2	0.11
$F(2, 888)$	57.96	P-value(F)	0
Log-likelihood	-567.2961	Akaike criterion	1140.592
Schwarz criterion	1154.97	Hannan–Quinn	1146.09

Titanic II

For categorical variables, the dummy coded columns are multi-collinear; hence one level needs to be removed.

Model 1: OLS, using observations 1–891
Dependent variable: Survived

	Coefficient	Std. Error	t-ratio	p-value
const	0.242	0.021	11.722	0.000
DPclass_1	0.387	0.037	10.353	0.000
DPclass_2	0.230	0.040	5.820	0.000

Mean dep var	0.38	S.D. dep var	0.49
Sum sq resid	186.39	S.E. of regr	0.46
R^2	0.12	Adjusted R^2	0.11
$F(2, 888)$	57.96	P-value(F)	0
Log-likelihood	−567.2961	Akaike criterion	1140.592
Schwarz criterion	1154.97	Hannan–Quinn	1146.09

Levels hiding

	Model 1	Model 2	Model 3	Model 4
(Intercept)	1.1250^{***}	1.1250^{***}		
	(0.0507)	(0.0507)		
Pclass2	-0.2077^{***}	-0.2077^{***}	0.9173^{***}	0.9173^{***}
	(0.0417)	(0.0417)	(0.0450)	(0.0450)
Pclass3	-0.4066^{***}	-0.4066^{***}	0.7184^{***}	0.7184^{***}
	(0.0383)	(0.0383)	(0.0380)	(0.0380)
Sexmale	-0.4795^{***}		-0.4795^{***}	
	(0.0307)		(0.0307)	
Age	-0.0055^{***}	-0.0055^{***}	-0.0055^{***}	-0.0055^{***}
	(0.0011)	(0.0011)	(0.0011)	(0.0011)
sex		-0.4795^{***}		-0.4795^{***}
		(0.0307)		(0.0307)
Pclass1			1.1250^{***}	1.1250^{***}
			(0.0507)	(0.0507)
R^2	0.3902	0.3902	0.6379	0.6379
Adj. R^2	0.3867	0.3867	0.6353	0.6353
Num. obs.	714	714	714	714
RMSE	0.3849	0.3849	0.3849	0.3849

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1:

Interactions with categorical variables

Different effects per level

Qual Interactions I

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	3.8860***	4.6116***		7.9830***	
	(0.2696)	(0.3060)		(0.2445)	
married1	1.3395***	−0.0457	4.5659***		
	(0.3097)	(0.4228)	(0.2917)		
male1	2.2944***	0.5564	0.5564		
	(0.3026)	(0.4736)	(0.4736)		
married1:male1		2.8607***	2.8607***		7.9830***
		(0.6076)	(0.6076)		(0.2445)
married0			4.6116***		
			(0.3060)		
married0:male0				−3.3714***	4.6116***
				(0.3916)	(0.3060)
married1:male0				−3.4171***	4.5659***
				(0.3806)	(0.2917)
married0:male1				−2.8150***	5.1680***
				(0.4363)	(0.3614)
R^2	0.1462	0.1810	0.7695	0.1810	0.7695
Adj. R^2	0.1429	0.1763	0.7678	0.1763	0.7678
Num. obs.	526	526	526	526	526
RMSE	3.4190	3.3518	3.3518	3.3518	3.3518

Qual Interactions II

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.9681^{***}		0.1354^{***}	
	(0.0392)		(0.0204)	
Pclass2	-0.0470	0.9211^{***}		
	(0.0586)	(0.0436)		
Pclass3	-0.4681^{***}	0.5000^{***}		
	(0.0504)	(0.0317)		
Sexmale	-0.5992^{***}	-0.5992^{***}		
	(0.0522)	(0.0522)		
Pclass2:Sexmale	-0.1644^{*}	-0.1644^{*}	0.0220	0.1574^{***}
	(0.0772)	(0.0772)	(0.0419)	(0.0366)
Pclass3:Sexmale	0.2347^{***}	0.2347^{***}		0.1354^{***}
	(0.0643)	(0.0643)		(0.0204)
Pclass1		0.9681^{***}		
		(0.0392)		
Pclass1:Sexfemale			0.8326^{***}	0.9681^{***}
			(0.0442)	(0.0392)
Pclass2:Sexfemale			0.7856^{***}	0.9211^{***}
			(0.0481)	(0.0436)
Pclass3:Sexfemale			0.3646^{***}	0.5000^{***}
			(0.0377)	(0.0317)
Pclass1:Sexmale			0.2334^{***}	0.3689^{***}

Binary Outcomes

Definitions

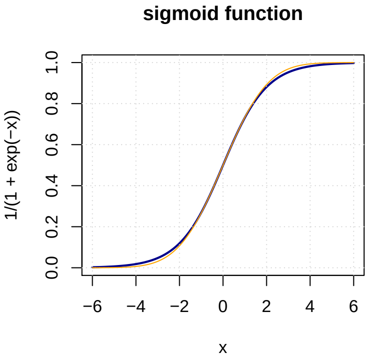
$$p = \text{Logistic(z)} = (\text{Sigmoid(z)}) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

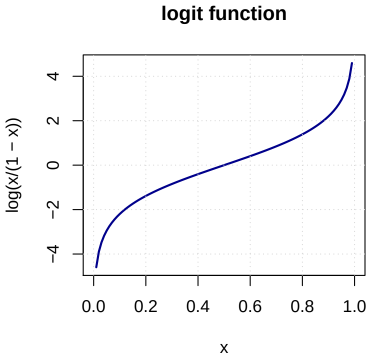
$$\text{Logit(p)} = \log\left(\frac{p}{1-p}\right) = \text{"log odds"}$$

The inverse of the sigmoid is the *logit* , which is defined as $log(p/(1-p))$. For the case where p is a probability we call the ratio $p/(1-p)$ the **probability odds**. Thus, the logit is the log of the odds and logistic regression models these *log-odds* as a linear combination of the values of x.

Logit

The sigmoid function, depicted below to the left, transforms the real axis to the interval $(0; 1)$ and can be interpreted as a probability.





Logistic Regression

Using linear regression as a starting point

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \epsilon_i$$

we modify the right hand side such that (i) the model is still basically a linear combination of the x_js but (ii) the output is -like a probability- bounded between 0 and 1. This is achieved by "wrapping" a sigmoid function $s(z) = 1/(1 + exp(-z))$ around the weighted sum of the x_js:

$$P(y_i = 1|x) = s(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \epsilon_i)$$

Titanic, Example

```
## [1] "Coefficients:"
## [2] "             Estimate Std. Error z value Pr(>|z|)
## [3] "(Intercept)  3.77701    0.40112    9.42  < 2e-16 **
## [4] "Age         -0.03699    0.00766   -4.83  1.4e-06 **
## [5] "Sexmale     -2.52278    0.20739  -12.16  < 2e-16 **
## [6] "Pclass2     -1.30980    0.27807   -4.71  2.5e-06 **
## [7] "Pclass3     -2.58063    0.28144   -9.17  < 2e-16 **
```

Interpretation

Finally, we can interpret the coefficients directly: the odds of a positive outcome are multiplied by a factor of $exp(\beta_j)$ for every unit change in x_j. (In that light, logistic regression is reminiscient of linear regression with logarithmically transformed dependent variable which also leads to multiplicative rather than additive effects.)

As an example, the coefficient for Pclass 3 is -2.581, which means that the odds of survival compared to the reference level Pclass 1 are reduced by a factor of $exp(-2.581) = 0.076$;with all other input variables unchanged. How does the relative change in odds translate into probabilities? It is relatively easy to memorize the to and forth relationship between odds and probability p:

$$odds = \frac{p}{1-p} \Leftrightarrow p = \frac{odds}{1+odds}$$

Why make life so complicated?

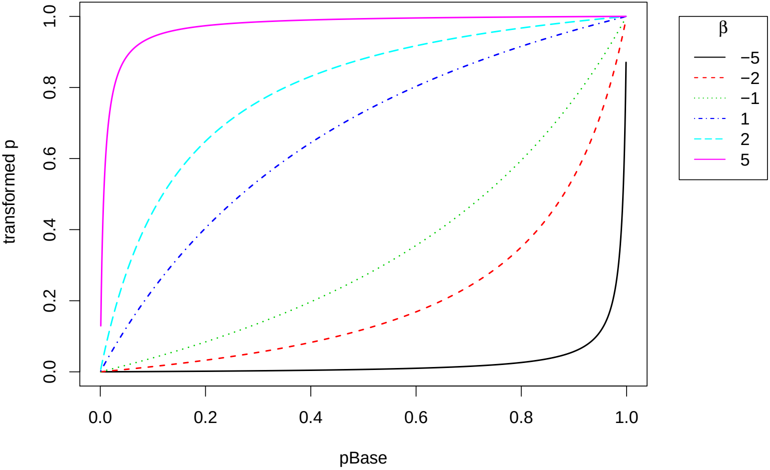
Do we have to go through the constant trouble of (i) exponentiating the coefficients, (ii) multiplying the odds and finally (iii) compute the resulting probabilities? Can we not simply transform the coefficients from the summary table such that we can read off their effects directly - just like in linear regression?

Partial effects are nonlinear and depend on the level of x!

$$\partial P(y_i = 1|x)/\partial x_j = g(\mathbf{x}\beta) \cdot \beta_j, \text{ with } g(z) = \partial s(z)/\partial z > 0$$

The following graph shows the effect of various coefficient values on a base/reference probability.

Non constant Effects



We immediately see that there is no straightforward additive or multiplicative effect that could be quantified.