

# Quantitative Research Methods: Non-linearity

Brendan Halpin, Sociology, University of Limerick

February 2018

## 1 Birth rate and GNP example

```
do http://teaching.sociology.ul.ie/so5032/birth
sort gnp
label var bir "Birth Rate"
label var gnp "GNP Per Capita"
lowess bir gnp, title("Birth rate and GNP per capita for selected countries")
```

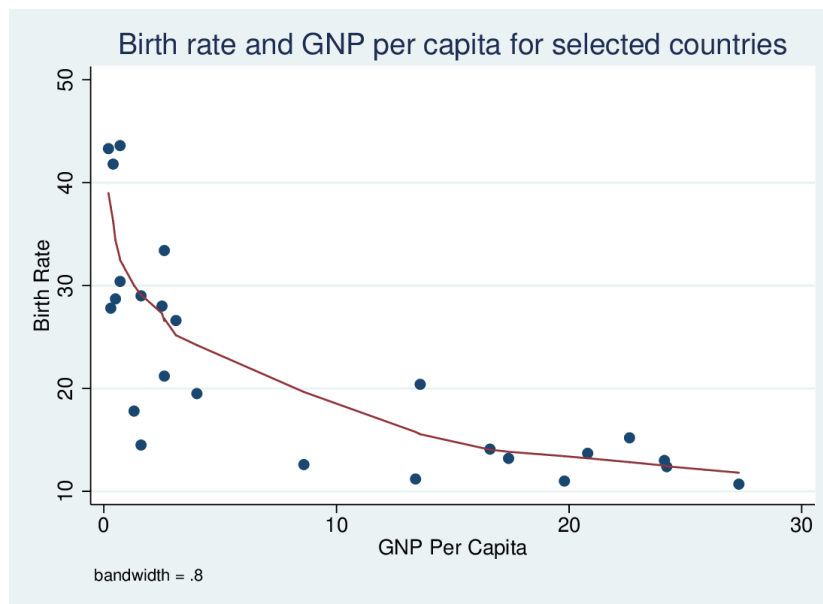


Figure shows a non-linear relationship.  
(Line is lowess)

```
reg bir gnp
```

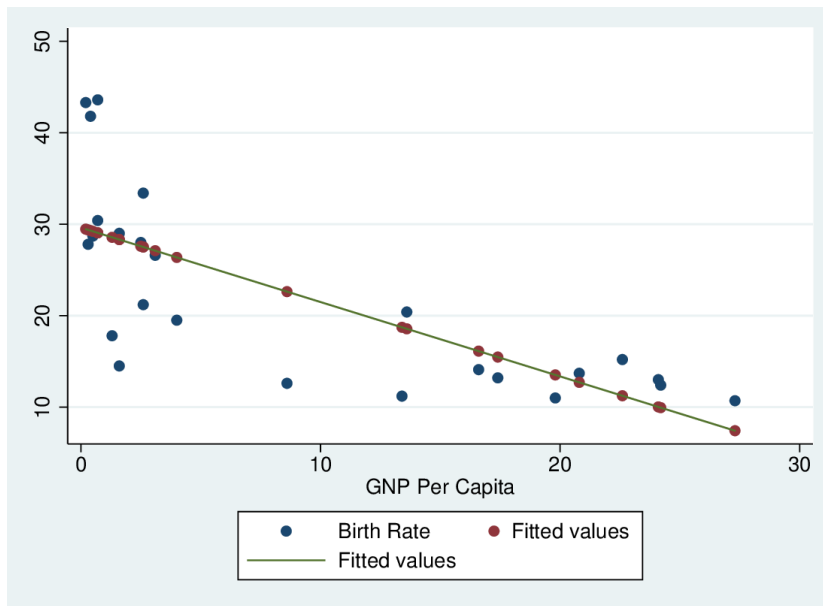
```
. reg bir gnp
```

Source	SS	df	MS	Number of obs	=	25
Model	1450.2603	1	1450.2603	F(1, 23)	=	27.52
Residual	1212.02523	23	52.696749	Prob > F	=	0.0000
				R-squared	=	0.5447
				Adj R-squared	=	0.5249
Total	2662.28552	24	110.928563	Root MSE	=	7.2593

bir	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
gnp	-.8133082	.155033	-5.25	0.000	-1.134018 - .4925981
_cons	29.6227	2.037416	14.54	0.000	25.40798 33.83742

```
predict plin
scatter bir plin gnp|| line plin gnp
```



Linear regression doesn't fit well  
Clearly, as GNP rises BIR falls, but the rate of fall declines  
Let's try quadratic:

```
reg bir c.gnp##c.gnp
sjlog close
```

```
. reg bir c.gnp##c.gnp
```

Source	SS	df	MS	Number of obs	=	25
Model	1665.82856	2	832.914278	F(2, 22)	=	18.39
Residual	996.456968	22	45.2934985	Prob > F	=	0.0000
				R-squared	=	0.6257
				Adj R-squared	=	0.5917
Total	2662.28552	24	110.928563	Root MSE	=	6.73

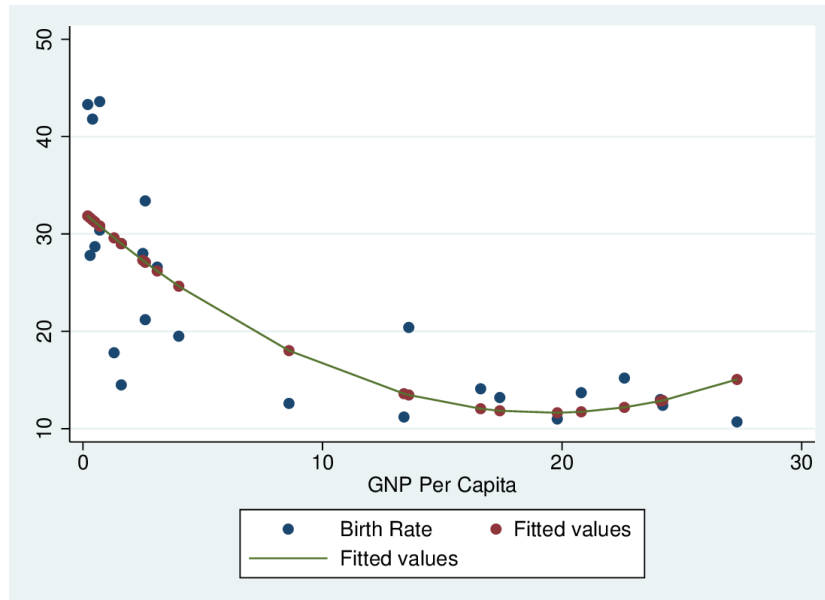
bir	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
gnp	-2.130192	.6205087	-3.43	0.002	-3.417048 - .8433351
c.gnp#c.gnp	.0549243	.0251762	2.18	0.040	.0027121 .1071366
_cons	32.27852	2.247195	14.36	0.000	27.61812 36.93892

```

predict pquad
scatter bir pquad gnp|| line pquad gnp

g2png, filename(birgnpquad) density(200) replace

```



Let's try square root of GNP:

```

gen sqg = sqrt(gnp)
reg bir sqg

```

```

. gen sqg = sqrt(gnp)
. reg bir sqg

```

Source	SS	df	MS	Number of obs	=	25
Model	1681.66084	1	1681.66084	F(1, 23)	=	39.44
Residual	980.624685	23	42.6358559	Prob > F	=	0.0000
Total	2662.28552	24	110.928563	R-squared	=	0.6317
				Adj R-squared	=	0.6156
				Root MSE	=	6.5296

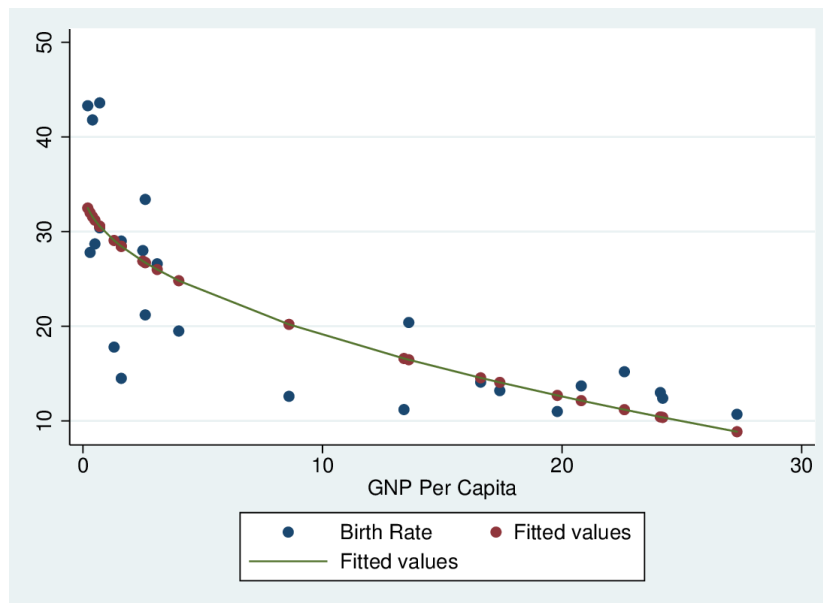
  

bir	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sqg	-4.945487	.7874579	-6.28	0.000	-6.574468 -3.316506
_cons	34.70314	2.391073	14.51	0.000	29.75683 39.64946

```

predict psqrt
scatter bir psqrt gnp|| line psqrt gnp

```



Let's try the log of GNP:

```
gen lgg = log(gnp)
reg bir lgg
```

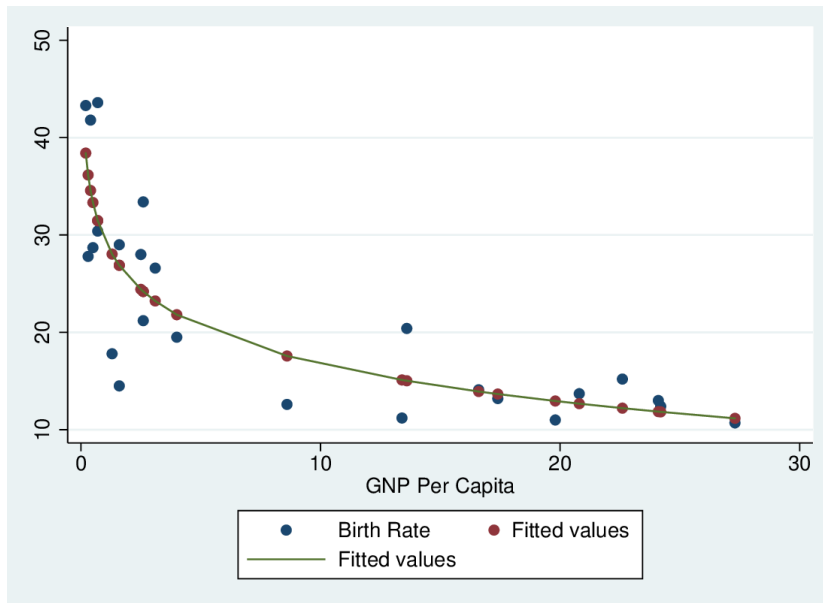
```
. gen lgg = log(gnp)
. reg bir lgg
```

Source	SS	df	MS	Number of obs	=	25
Model	1875.68482	1	1875.68482	F(1, 23)	=	54.84
Residual	786.600705	23	34.2000307	Prob > F	=	0.0000
Total	2662.28552	24	110.928563	R-squared	=	0.7045
				Adj R-squared	=	0.6917
				Root MSE	=	5.8481

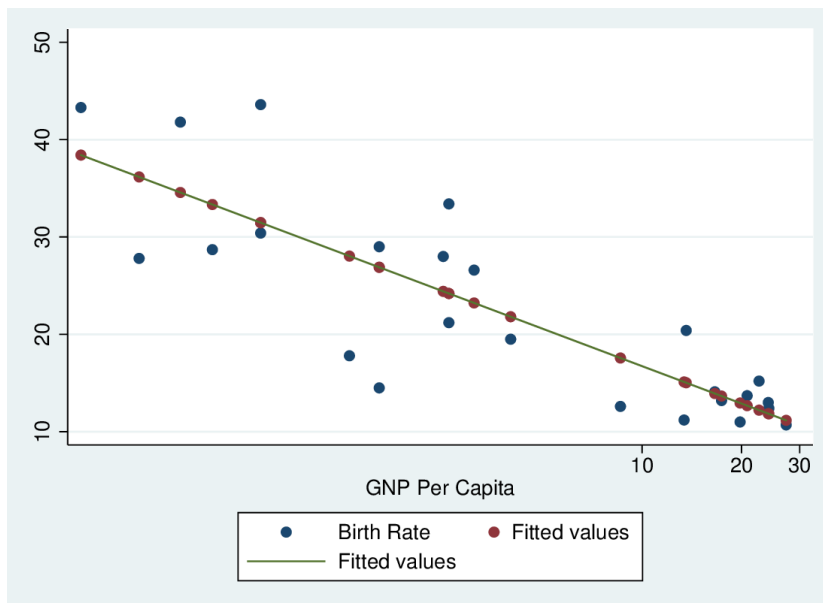
  

bir	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lgg	-5.542152	.748362	-7.41	0.000	-7.090257 -3.994047
_cons	29.49466	1.53576	19.21	0.000	26.3177 32.67162

```
predict plog
scatter bir plog gnp||line plog gnp
```



```
scatter bir plog gnp, xscale(log) || line plog gnp, xscale(log)
```



```
label var sqg "Sq Root GNP"
label var lg "Log of GNP"
scatter sqg lg gnp
```

