

Models In Epidemiology And Biostatistics

Gordon Hilton Fick

Session 7 : More On Nonlinearity

The Study to Understand Prognoses Preferences Outcomes and Risks of Treatment (SUPPORT) was a prospective observational study of hospitalized patients (Knaus et al., 1995). The file support.dta [from: biostat.mc.vanderbilt.edu/DataSets] contains many variables from the 9105 patients in this study.

Parabola, Piecewise Linear Curves and Restricted Cubic Splines will be illustrated with in-hospital death [death], an age grouping [a : age>62] and mean arterial pressure [b = meanbp in mmHg]

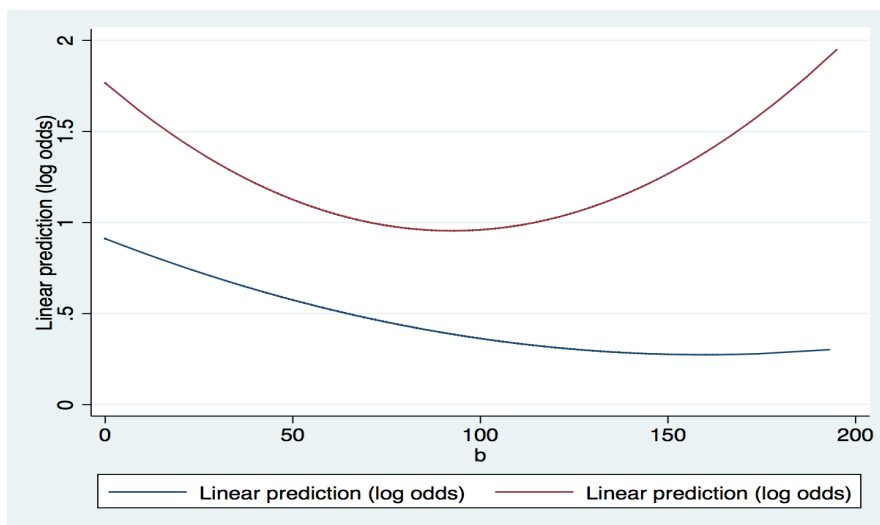
```
. gen b = meanbp
. gen b2 = b*b
. gen a = (age > 62)
. gen ab = a*b
. gen ab2 = a*b2

. logit death a b b2 ab ab2
```

Logistic regression	Number of obs	=	9,104
	LR chi2(5)	=	191.94
	Prob > chi2	=	0.0000
Log likelihood = -5604.0331	Pseudo R2	=	0.0168

death	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
a	.8544317	.3784705	2.26	0.024	.112643 1.59622
b	-.0079806	.0057968	-1.38	0.169	-.0193421 .0033809
b2	.000025	.000032	0.78	0.434	-.0000376 .0000876
ab	-.009552	.0088561	-1.08	0.281	-.0269096 .0078056
ab2	.0000698	.0000491	1.42	0.155	-.0000265 .000166
_cons	.9109971	.2499201	3.65	0.000	.4211627 1.400831

```
. predict l1,xb
. sort b
. twoway (line l1 b if a==0) (line l1 b if a==1)
```



```
. mkspline bls 7 = b,pctile
```

```
. gen abls1=a*bls1
. gen abls2=a*bls2
. gen abls3=a*bls3
. gen abls4=a*bls4
. gen abls5=a*bls5
. gen abls6=a*bls6
. gen abls7=a*bls7
```

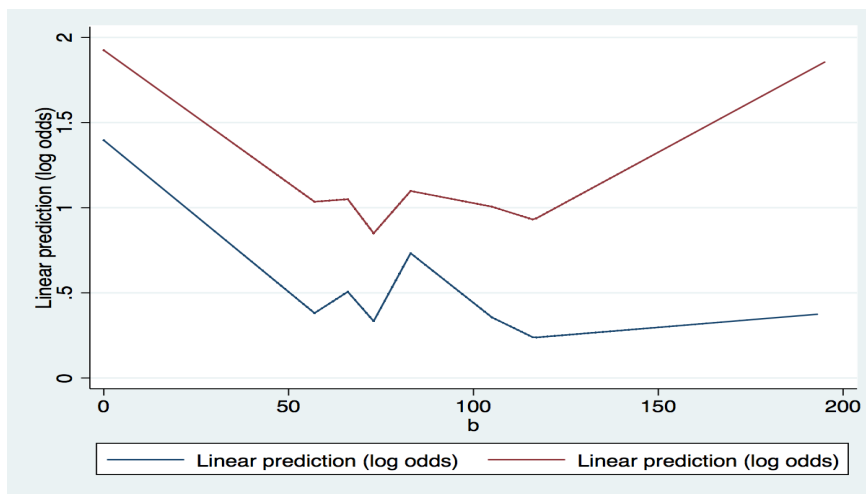
```
. logit death a bls* abls*
```

```
Iteration 0:   log likelihood = -5700.0028
Iteration 1:   log likelihood = -5596.2772
Iteration 2:   log likelihood = -5595.6746
Iteration 3:   log likelihood = -5595.6741
```

Logistic regression	Number of obs	=	9,104
	LR chi2(15)	=	208.66
	Prob > chi2	=	0.0000
Log likelihood = -5595.6741	Pseudo R2	=	0.0183

death	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
a	.5286011	.5520229	0.96	0.338	-.5533439	1.610546
bls1	-.0178062	.0071119	-2.50	0.012	-.0317453	-.0038672
bls2	.0139431	.0187267	0.74	0.457	-.0227606	.0506468
bls3	-.0247186	.0250618	-0.99	0.324	-.0738389	.0244016
bls4	.03993	.0185294	2.15	0.031	.003613	.0762471
bls5	-.0171298	.008083	-2.12	0.034	-.0329723	-.0012874
bls6	-.0105468	.0138597	-0.76	0.447	-.0377113	.0166176
bls7	.0017906	.0061972	0.29	0.773	-.0103557	.0139368
abls1	.0021998	.0109055	0.20	0.840	-.0191746	.0235742
abls2	-.0122932	.0255972	-0.48	0.631	-.0624628	.0378764
abls3	-.0038769	.0341337	-0.11	0.910	-.0707777	.0630239
abls4	-.0150693	.0258075	-0.58	0.559	-.065651	.0355124
abls5	.0129329	.011321	1.14	0.253	-.0092559	.0351217
abls6	.0037379	.0194402	0.19	0.848	-.0343642	.0418401
abls7	.0099731	.0095534	1.04	0.297	-.0087513	.0286974
_cons	1.395963	.3563556	3.92	0.000	.6975183	2.094407

```
. predict lo2,xb
. twoway (line lo2 b if a==0)(line lo2 b if a==1)
```



```
. mkspline brcs = b, cubic nknots(7) displayknots
```

```
-----+-----
          |      knot1      knot2      knot3      knot4      knot5      knot6      knot7
          +-----+-----+-----+-----+-----+-----+-----+
          b |      42      60      69      77      100      113      139
```

```
. gen abrcs1=a*brcs1
. gen abrcs2=a*brcs2
. gen abrcs3=a*brcs3
. gen abrcs4=a*brcs4
. gen abrcs5=a*brcs5
. gen abrcs6=a*brcs6
```

```
. logit death a brcs* abrcs*
```

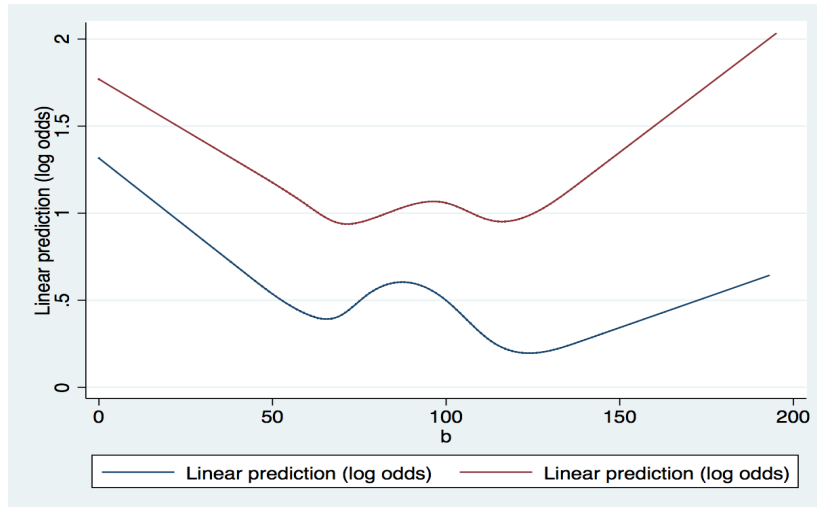
```
Iteration 0:   log likelihood = -5700.0028
Iteration 1:   log likelihood = -5597.6518
Iteration 2:   log likelihood = -5597.0801
Iteration 3:   log likelihood = -5597.0798
```

```
Logistic regression              Number of obs   =       9,104
                                LR chi2(13)      =       205.85
                                Prob > chi2       =       0.0000
Log likelihood = -5597.0798      Pseudo R2    =       0.0181
```

```
-----+-----
      death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
       a |   .4535419   .5666641     0.80   0.423    - .6570993    1.564183
    brcs1 |  -.0156754   .0076371    -2.05   0.040    - .0306439   -.0007069
    brcs2 |   .0702653   .1484817     0.47   0.636    - .2207535    .3612842
    brcs3 |   .322809    1.136871     0.28   0.776    -1.905416    2.551034
    brcs4 |  -1.293671   2.244449    -0.58   0.564    -5.692711    3.105369
    brcs5 |   .9232605   1.532782     0.60   0.547    -2.080938    3.927459
    brcs6 |   .3141737   .5888989     0.53   0.594    - .840047    1.468394
   abrcs1 |   .003818    .0117053     0.33   0.744    - .019124    .02676
   abrcs2 |  -.0906412   .2112104    -0.43   0.668    - .5046059    .3233235
   abrcs3 |   .1295184   1.581304     0.08   0.935    -2.96978    3.228817
   abrcs4 |   .5202307   3.102879     0.17   0.867    -5.5613    6.601761
   abrcs5 |  -.7145297   2.117496    -0.34   0.736    -4.864746    3.435687
   abrcs6 |   .1739235   .8234817     0.21   0.833    -1.440071    1.787918
    _cons |   1.315813   .3656957     3.60   0.000    .5990625    2.032563
-----+-----
```

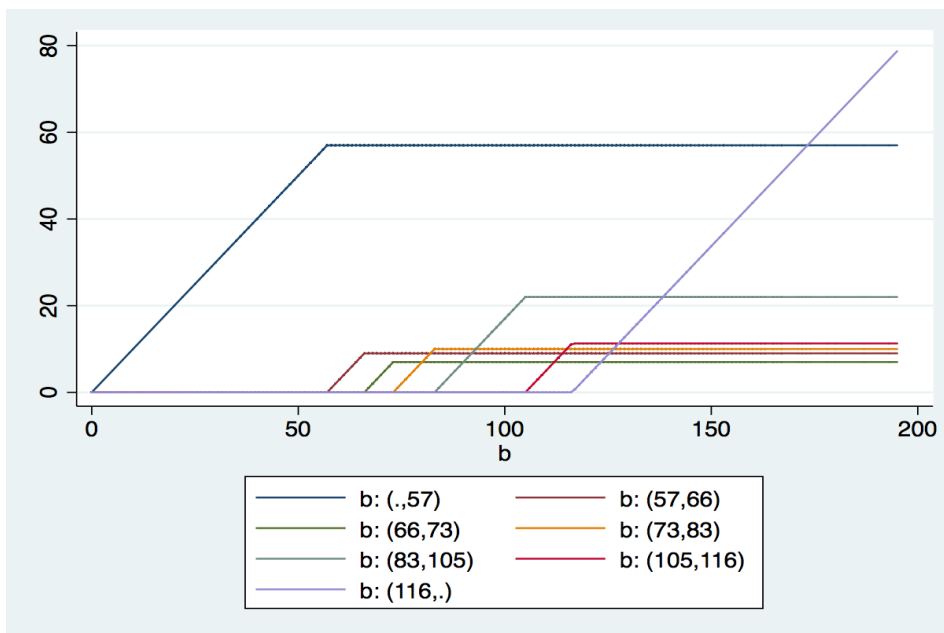
```
. predict lo3,xb
```

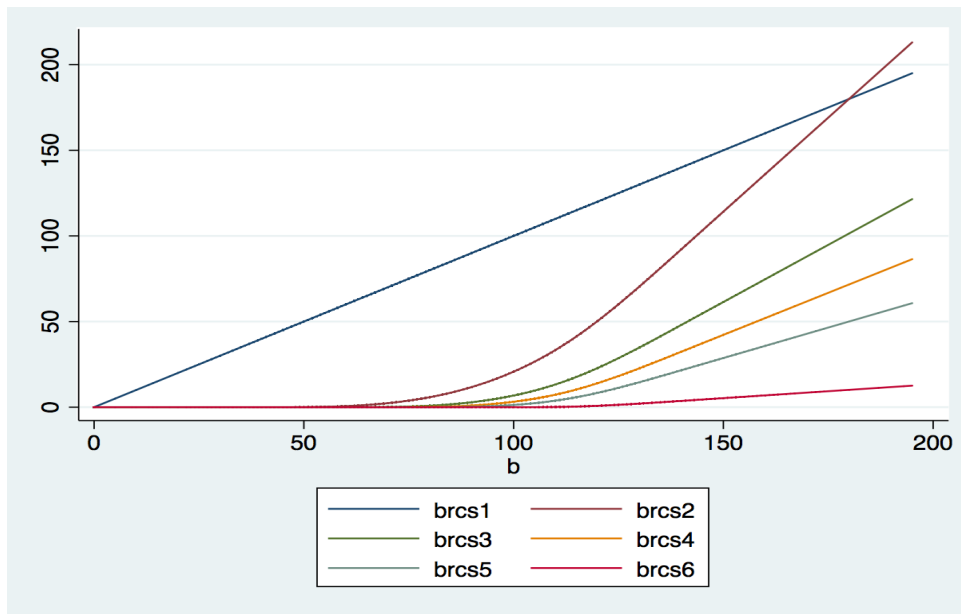
```
. twoway (line lo3 b if a==0) (line lo3 b if a==1)
```



So mkspline creates the variables that can then be used to build the additional variables to place in the logit models.

For the linear splines [aka piecewise linear curves] :





For the restricted cubic splines [aka natural splines] :