Week 3 Questions

Dr M L Mackenzie

February 2018

Introduction

We are going to use the same data as for the last two weeks of the 'Weekly Questions' to primarily look at polynomials and B-splines.

```
       id              year          assets              capex         
 Min.   :  1050   Min.   :1990   Min.   :    0.72   Min.   :-401.609  
 1st Qu.: 10391   1st Qu.:1999   1st Qu.:   59.32   1st Qu.:   1.572  
 Median : 25390   Median :2004   Median :  216.73   Median :   8.203  
 Mean   : 54440   Mean   :2004   Mean   : 1434.82   Mean   :  74.055  
 3rd Qu.:104598   3rd Qu.:2009   3rd Qu.:  914.16   3rd Qu.:  40.154  
 Max.   :270705   Max.   :2014   Max.   :80033.55   Max.   :7150.000  
      ltd                ebitda                ppe          
 Min.   :    0.000   Min.   :-1134.000   Min.   :    0.00  
 1st Qu.:    0.000   1st Qu.:    2.293   1st Qu.:   15.13  
 Median :    3.129   Median :   21.395   Median :   67.45  
 Mean   :  293.976   Mean   :  199.175   Mean   :  768.24  
 3rd Qu.:  100.000   3rd Qu.:  123.826   3rd Qu.:  341.80  
 Max.   :24380.700   Max.   : 7723.890   Max.   :59762.83  
     sales                ads                 rd          
 Min.   :     0.05   Min.   :   0.000   Min.   :   0.000  
 1st Qu.:    58.35   1st Qu.:   0.537   1st Qu.:   0.000  
 Median :   229.86   Median :   3.055   Median :   3.996  
 Mean   :  1453.51   Mean   :  40.312   Mean   :  43.057  
 3rd Qu.:   992.47   3rd Qu.:  18.500   3rd Qu.:  23.247  
 Max.   :108465.00   Max.   :2840.000   Max.   :3146.829  
    bookval              mv              indclass        tobinsQ         
 Min.   :    0.04   Min.   :    0.10   Min.   : 1.00   Min.   :  0.00004  
 1st Qu.:   34.52   1st Qu.:   60.49   1st Qu.:15.00   1st Qu.:  1.23689  
 Median :  128.56   Median :  275.95   Median :36.00   Median :  2.10815  
 Mean   :  654.89   Mean   : 1759.15   Mean   :29.35   Mean   :  3.21056  
 3rd Qu.:  482.49   3rd Qu.: 1231.72   3rd Qu.:38.00   3rd Qu.:  3.58202  
 Max.   :41466.76   Max.   :50174.89   Max.   :49.00   Max.   :193.71598  
    ltdratio          capexratio           rdratio            adsratio       
 Min.   :0.00000   Min.   :-0.25973   Min.   :0.0000   Min.   :0.000000  
 1st Qu.:0.00000   1st Qu.: 0.01781   1st Qu.:0.0000   1st Qu.:0.005603  
 Median :0.03931   Median : 0.03282   Median :0.0290   Median :0.014364  
 Mean   :0.11473   Mean   : 0.05643   Mean   :0.0908   Mean   :0.032936  
 3rd Qu.:0.19905   3rd Qu.: 0.05886   3rd Qu.:0.1159   3rd Qu.:0.035927  
 Max.   :0.85654   Max.   : 5.04205   Max.   :7.0207   Max.   :1.757624  
    pperatio        ebitdaratio      
 Min.   :0.0000   Min.   :-4.06708  
 1st Qu.:0.1830   1st Qu.: 0.04678  
 Median :0.3484   Median : 0.11329  
 Mean   :0.4367   Mean   : 0.08921  
 3rd Qu.:0.6103   3rd Qu.: 0.17485  
 Max.   :3.7330   Max.   : 1.50841  
```

Polynomial-based models

```
deg <- 2

polyfit2 <- lm(tobinsQ ~ poly(ltdratio, degree = deg) + poly(capexratio,
    degree = deg) + poly(rdratio, degree = deg) + poly(adsratio,
    degree = deg) + poly(pperatio, degree = deg) + poly(ebitdaratio,
    degree = deg) + poly(year, degree = deg) + poly(assets, degree = deg) +
    poly(capex, degree = deg) + poly(ltd, degree = deg) + poly(ebitda,
    degree = deg) + poly(ppe, degree = deg) + poly(sales, degree = deg) +
    poly(ads, degree = deg) + poly(rd, degree = deg) + poly(bookval,
    degree = deg) + poly(mv, degree = deg) + as.factor(indclass),
    data = newdat)
stepfit2 <- step(polyfit2, trace = 0)
pander(Anova(polyfit2))
```

Table 1: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
poly(ltdratio, degree = deg)	4358	2	95.9	4.43e-42
poly(capexratio, degree = deg)	69.1	2	1.52	0.2187
poly(rdratio, degree = deg)	469.3	2	10.33	3.306e-05
poly(adsratio, degree = deg)	155.9	2	3.43	0.0324
poly(pperatio, degree = deg)	203.9	2	4.486	0.01128
poly(ebitdaratio, degree = deg)	2418	2	53.2	9.698e-24
poly(year, degree = deg)	918.4	2	20.21	1.724e-09
poly(assets, degree = deg)	700.5	2	15.41	2.06e-07
poly(capex, degree = deg)	25.59	2	0.5631	0.5695
poly(ltd, degree = deg)	87.51	2	1.926	0.1458
poly(ebitda, degree = deg)	1509	2	33.21	4.113e-15
poly(ppe, degree = deg)	564.2	2	12.41	4.107e-06
poly(sales, degree = deg)	410.2	2	9.027	0.0001209
poly(ads, degree = deg)	225.4	2	4.959	0.00703
poly(rd, degree = deg)	1583	2	34.82	8.223e-16
poly(bookval, degree = deg)	7988	2	175.8	4.411e-76
poly(mv, degree = deg)	33044	2	727.1	1.438e-300
as.factor(indclass)	5579	40	6.138	5.158e-31
Residuals	305631	13450	NA	NA

```
pander(Anova(stepfit2))
```

Table 2: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
poly(ltdratio, degree = deg)	4376	2	96.27	3.057e-42
poly(rdratio, degree = deg)	533.4	2	11.74	8.083e-06
poly(adsratio, degree = deg)	192.1	2	4.226	0.01463
poly(pperatio, degree = deg)	175.5	2	3.861	0.02107
poly(ebitdaratio, degree = deg)	2420	2	53.24	9.318e-24
poly(year, degree = deg)	1019	2	22.42	1.901e-10
poly(assets, degree = deg)	902	2	19.84	2.479e-09
poly(ebitda, degree = deg)	1650	2	36.31	1.877e-16

	Sum Sq	Df	F value	Pr(>F)
poly(ppe, degree = deg)	622.8	2	13.7	1.136e-06
poly(sales, degree = deg)	396.9	2	8.732	0.0001622
poly(ads, degree = deg)	255.4	2	5.619	0.003637
poly(rd, degree = deg)	1572	2	34.58	1.049e-15
poly(bookval, degree = deg)	10420	2	229.3	1.245e-98
poly(mv, degree = deg)	33920	2	746.3	4.36e-308
as.factor(indclass)	5615	40	6.177	2.673e-31
Residuals	305797	13456	NA	NA

```
pander(vif(polyfit2))
```

	GVIF	Df	GVIF^(1/(2*Df))
poly(ltdratio, degree = deg)	1.941	2	1.18
poly(capexratio, degree = deg)	1.587	2	1.122
poly(rdratio, degre = deg)	2.119	2	1.207
poly(adsratio, degree = deg)	1.513	2	1.109
poly(pperatio, degree = deg)	2.605	2	1.27
poly(ebitdaratio, degree = deg)	1.717	2	1.145
poly(year, degree = deg)	1.211	2	1.049
poly(assets, degree = deg)	918	2	5.504
poly(capex, degree = deg)	22.95	2	2.189
poly(ltd, degree = deg)	39.68	2	2.51
poly(ebitda, degree = deg)	36.05	2	2.45
poly(ppe, degree = deg)	39.79	2	2.512
poly(sales, degree = deg)	14.59	2	1.954
poly(ads, degree = deg)	4.043	2	1.418
poly(rd, degree = deg)	3.696	2	1.387
poly(bookval, degree = deg)	144	2	3.464
poly(mv, degree = deg)	8.955	2	1.73
as.factor(indclass)	12.79	40	1.032

```
pander(vif(stepfit2))
```

	GVIF	Df	GVIF^(1/(2*Df))
poly(ltdratio, degree = deg)	1.68	2	1.139
poly(rdratio, degre = deg)	1.855	2	1.167
poly(adsratio, degree = deg)	1.466	2	1.1
poly(pperatio, degree = deg)	2.518	2	1.26
poly(ebitdaratio, degree = deg)	1.714	2	1.144
poly(year, degree = deg)	1.178	2	1.042
poly(assets, degree = deg)	130.3	2	3.378
poly(ebitda, degree = deg)	34.1	2	2.416
poly(ppe, degree = deg)	21.54	2	2.154
poly(sales, degree = deg)	12.3	2	1.873
poly(ads, degree = deg)	3.781	2	1.394
poly(rd, degree = deg)	3.14	2	1.331
poly(bookval, degree = deg)	63.64	2	2.824
poly(mv, degree = deg)	8.568	2	1.711
as.factor(indclass)	10.37	40	1.03

```r
require(pander)
pander(AIC(polyfit2, stepfit2))
```

	df	AIC
polyfit2	76	80703
stepfit2	70	80698

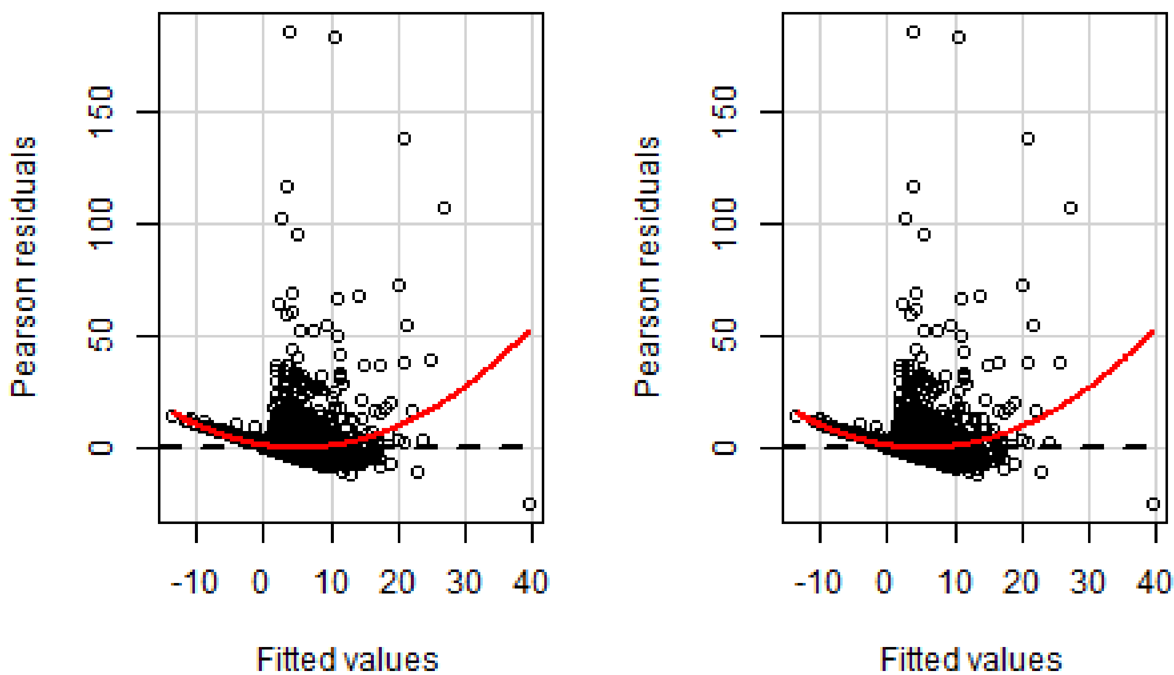
```r
pander(BIC(polyfit2, stepfit2))
```

	df	BIC
polyfit2	76	81274
stepfit2	70	81224

```r
par(mfrow = c(1, 2))

residualPlots(polyfit2, terms = ~1)
```

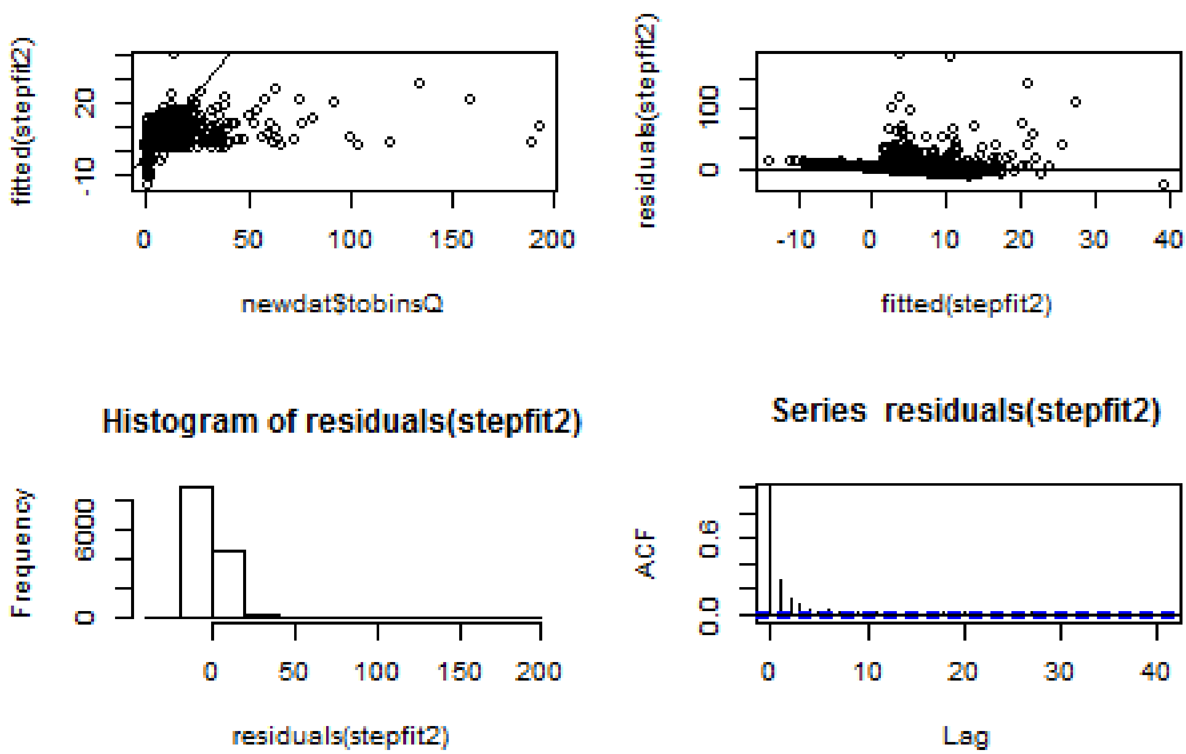
```
           Test stat Pr(>|t|)
Tukey test    27.454        0
```

```r
residualPlots(stepfit2, terms = ~1)
```




```
           Test stat Pr(>|t|)
Tukey test    27.261        0
```

```r
par(mfrow = c(2, 2))
plot(newdat$tobinsQ, fitted(stepfit2))
abline(0, 1)
plot(fitted(stepfit2), residuals(stepfit2))
abline(h = 0)
hist(residuals(stepfit2))
acf(residuals(stepfit2))
```



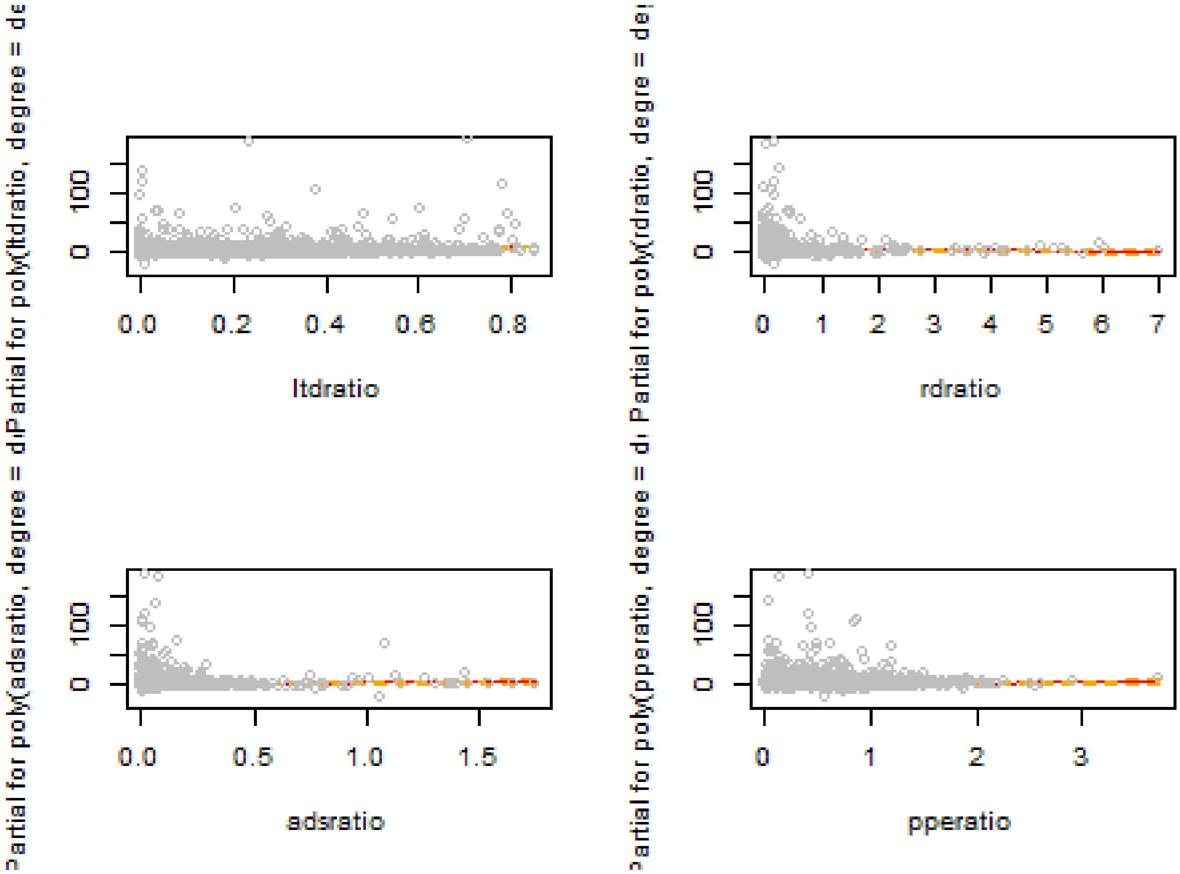

```r
ncvTest(stepfit2)
```

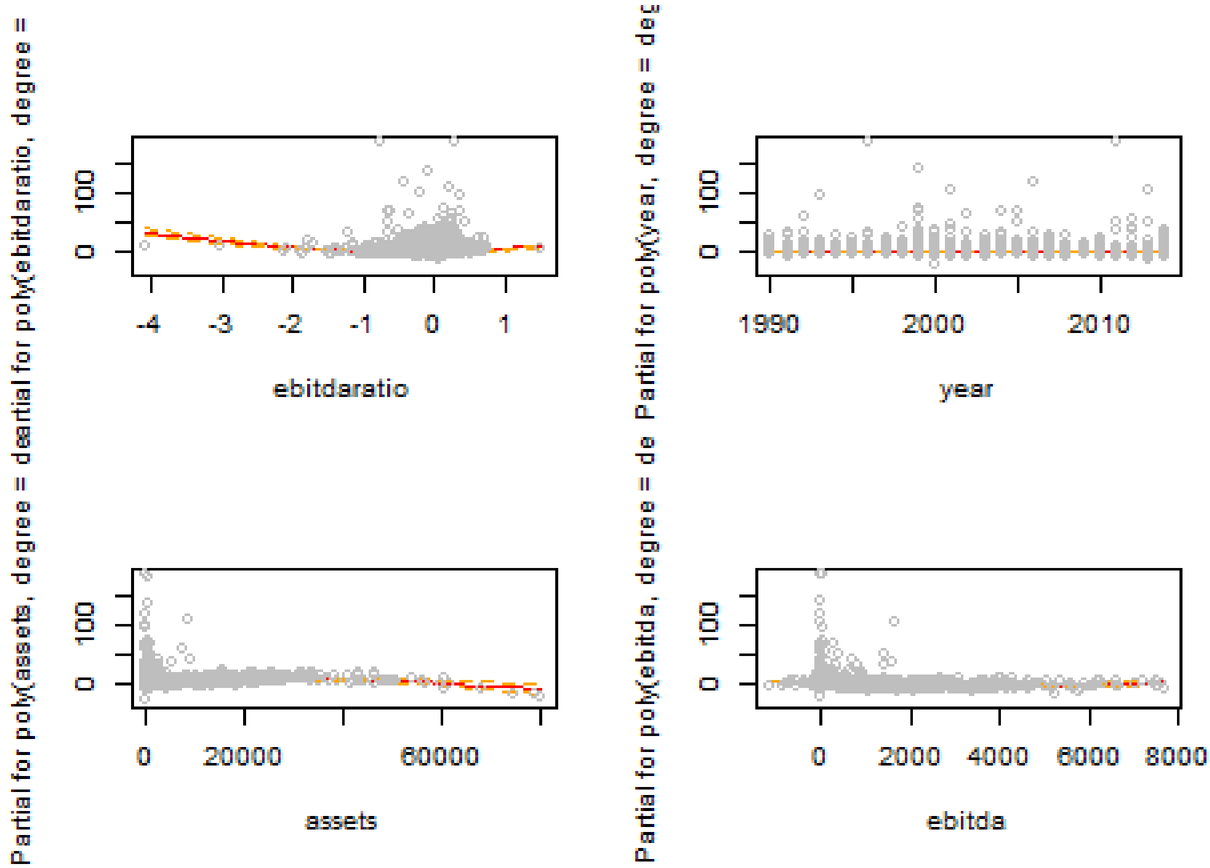
```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 31700.44    Df = 1     p = 0
```

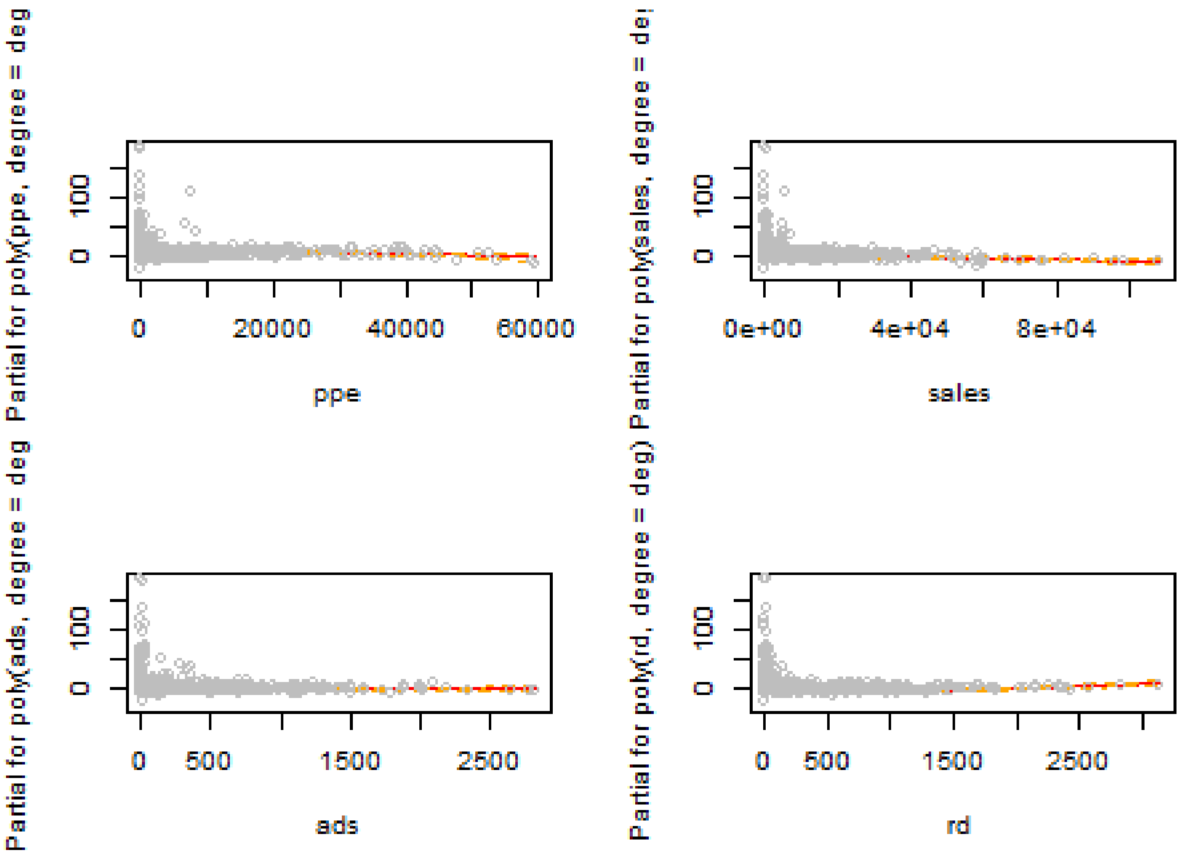
```r
runs.test(residuals(stepfit2))
```

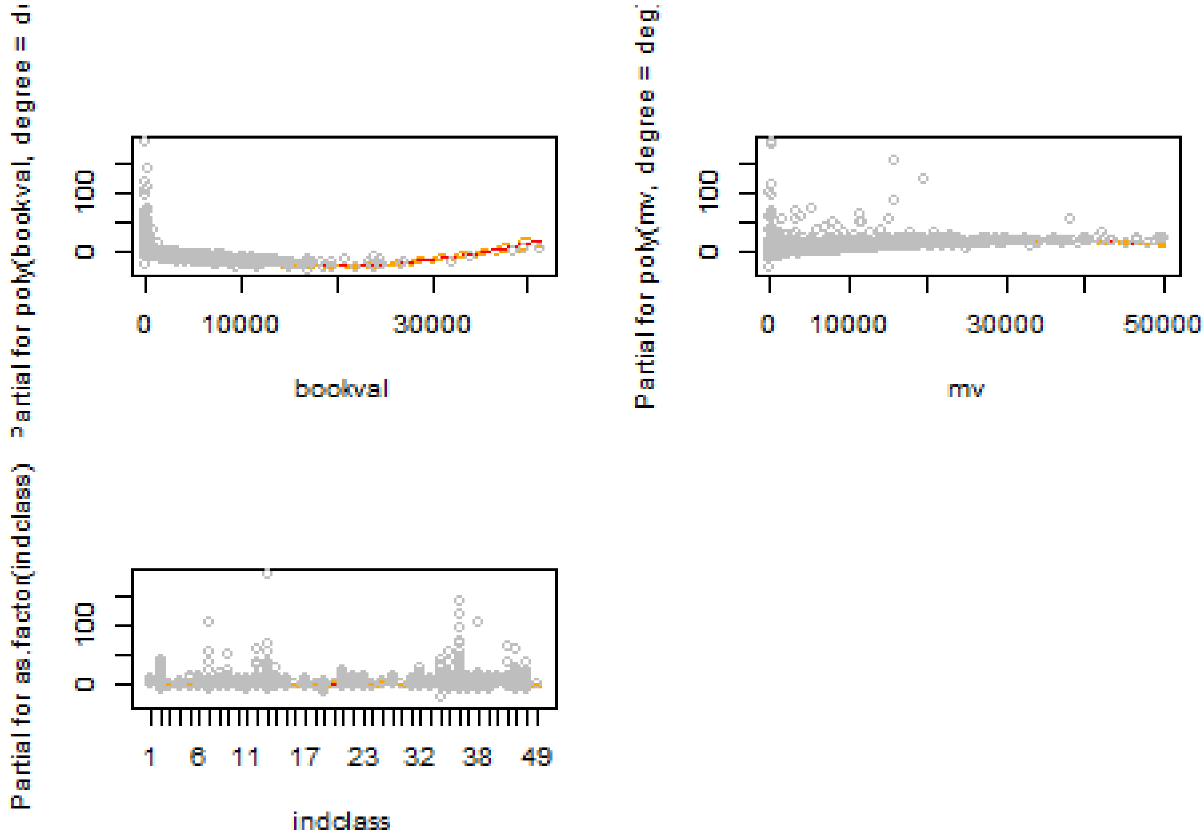
```

    Runs Test - Two sided

data:  residuals(stepfit2)
Standardized Runs Statistic = -62.334, p-value < 2.2e-16
```

```r
par(mfrow = c(2, 2))
termplot(stepfit2, se = T, partial.resid = TRUE)
```


ltdratio
rdratio
adsratio
pperatio


Partial for poly(ebitdaratio, degree = de
Partial for poly(year, degree = deg
Partial for poly(assets, degree = de
Partial for poly(ebitda, degree = de
ebitdaratio
year
assets
ebitda


Partial for poly(ppe, degree = deg
ppe
0
20000
40000
60000
Partial for poly(sales, degree = de
sales
0e+00
4e+04
8e+04
Partial for poly(ads, degree = deg
ads
0
500
1500
2500
Partial for poly(rd, degree = deg)
rd
0
100




```r
deg <- 3
polyfit3 <- lm(tobinsQ ~ poly(ltdratio, degree = deg) + poly(capexratio,
    degree = deg) + poly(rdratio, degree = deg) + poly(adsratio,
    degree = deg) + poly(pperatio, degree = deg) + poly(ebitdaratio,
    degree = deg) + poly(year, degree = deg) + poly(assets, degree = deg) +
    poly(capex, degree = deg) + poly(ltd, degree = deg) + poly(ebitda,
    degree = deg) + poly(ppe, degree = deg) + poly(sales, degree = deg) +
    poly(ads, degree = deg) + poly(rd, degree = deg) + poly(bookval,
    degree = deg) + poly(mv, degree = deg) + as.factor(indclass),
    data = newdat)

stepfit3 <- step(polyfit3, trace = 0)
pander(Anova(polyfit3))
```

Table 7: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
poly(ltdratio, degree = deg)	4985	3	77.9	5.973e-50
poly(capexratio, degree = deg)	133.9	3	2.093	0.09886
poly(rdratio, degree = deg)	481.7	3	7.528	4.976e-05
poly(adsratio, degree = deg)	168.1	3	2.627	0.04859
poly(pperatio, degree = deg)	291.9	3	4.562	0.003377
poly(ebitdaratio, degree = deg)	5299	3	82.8	4.497e-53
poly(year, degree = deg)	520	3	8.126	2.105e-05
poly(assets, degree = deg)	982.5	3	15.35	5.721e-10
poly(capex, degree = deg)	223.1	3	3.486	0.01508

	Sum Sq	Df	F value	Pr(>F)
poly(ltd, degree = deg)	304.8	3	4.762	0.002548
poly(ebitda, degree = deg)	2608	3	40.76	3.275e-26
poly(ppe, degree = deg)	1210	3	18.92	3.084e-12
poly(sales, degree = deg)	130.6	3	2.041	0.1058
poly(ads, degree = deg)	184.3	3	2.88	0.03451
poly(rd, degree = deg)	1029	3	16.08	1.975e-10
poly(bookval, degree = deg)	12470	3	194.9	1.053e-123
poly(mv, degree = deg)	41184	3	643.6	0
as.factor(indclass)	4140	40	4.852	6.527e-22
Residuals	286549	13433	NA	NA

```r
pander(Anova(stepfit3))
```

Table 8: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
poly(ltdratio, degree = deg)	4985	3	77.9	5.973e-50
poly(capexratio, degree = deg)	133.9	3	2.093	0.09886
poly(rdratio, degre = deg)	481.7	3	7.528	4.976e-05
poly(adsratio, degree = deg)	168.1	3	2.627	0.04859
poly(pperatio, degree = deg)	291.9	3	4.562	0.003377
poly(ebitdaratio, degree = deg)	5299	3	82.8	4.497e-53
poly(year, degree = deg)	520	3	8.126	2.105e-05
poly(assets, degree = deg)	982.5	3	15.35	5.721e-10
poly(capex, degree = deg)	223.1	3	3.486	0.01508
poly(ltd, degree = deg)	304.8	3	4.762	0.002548
poly(ebitda, degree = deg)	2608	3	40.76	3.275e-26
poly(ppe, degree = deg)	1210	3	18.92	3.084e-12
poly(sales, degree = deg)	130.6	3	2.041	0.1058
poly(ads, degree = deg)	184.3	3	2.88	0.03451
poly(rd, degree = deg)	1029	3	16.08	1.975e-10
poly(bookval, degree = deg)	12470	3	194.9	1.053e-123
poly(mv, degree = deg)	41184	3	643.6	0
as.factor(indclass)	4140	40	4.852	6.527e-22
Residuals	286549	13433	NA	NA

```r
pander(AIC(polyfit3, stepfit3))
```

	df	AIC
polyfit3	93	79865
stepfit3	93	79865

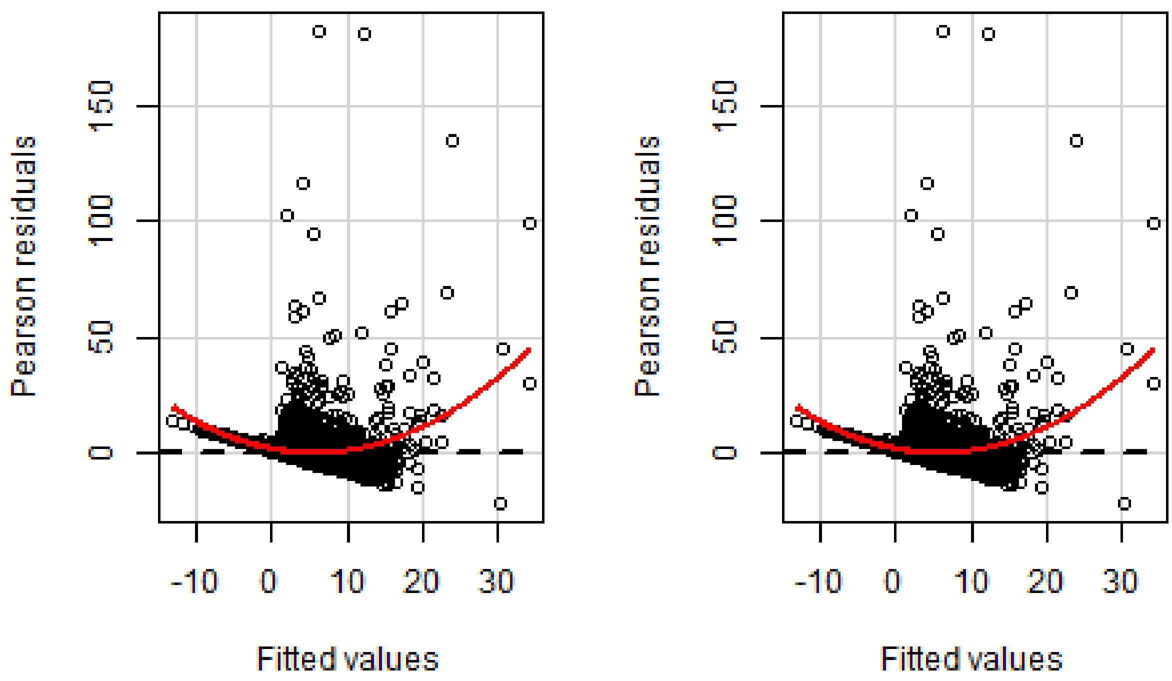
```r
pander(BIC(polyfit3, stepfit3))
```

	df	BIC
polyfit3	93	80564
stepfit3	93	80564

```r
par(mfrow = c(1, 2))
residualPlots(polyfit3, terms = ~1)
```

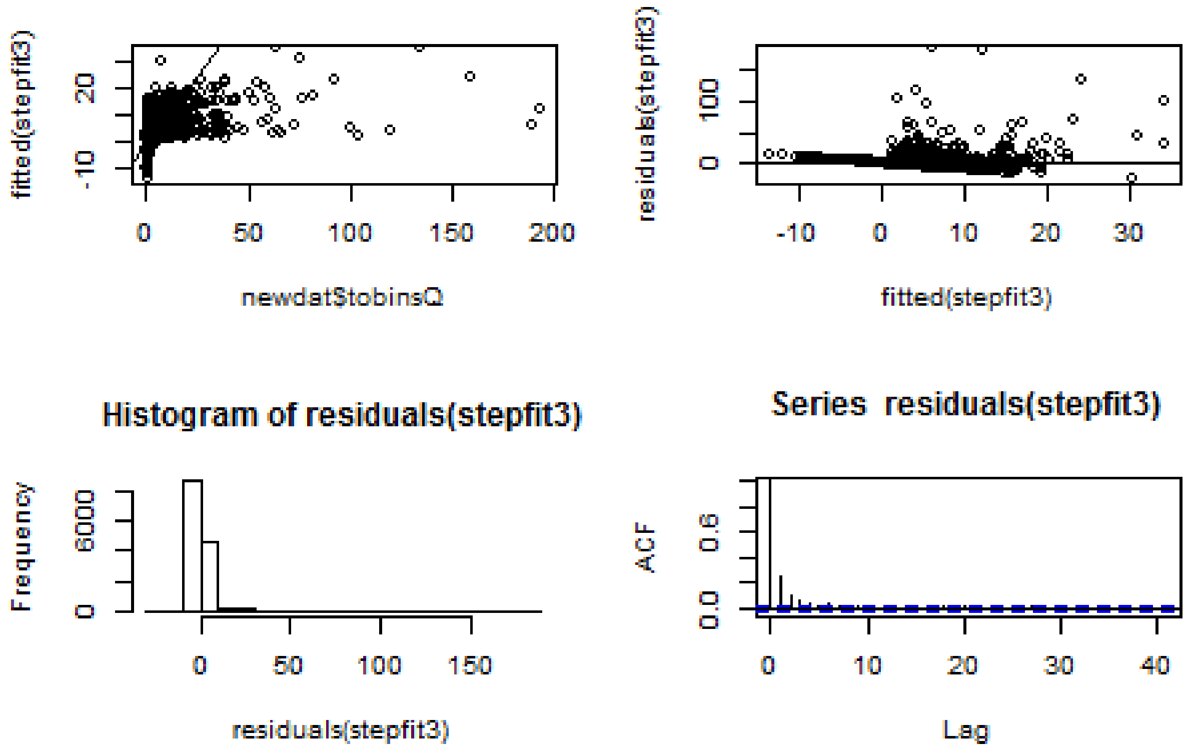
```
           Test stat Pr(>|t|)
Tukey test    35.332        0
```

```r
residualPlots(stepfit3, terms = ~1)
```




```
           Test stat Pr(>|t|)
Tukey test    35.332        0
```

```r
par(mfrow = c(2, 2))
plot(newdat$tobinsQ, fitted(stepfit3))
abline(0, 1)
plot(fitted(stepfit3), residuals(stepfit3))
abline(h = 0)
hist(residuals(stepfit3))
acf(residuals(stepfit3))
```




```
ncvTest(stepfit3)
```

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 41794.4    Df = 1     p = 0
```

```
runs.test(residuals(stepfit3))
```

```

    Runs Test - Two sided

data:  residuals(stepfit3)
Standardized Runs Statistic = -61.853, p-value < 2.2e-16
```

B-spline based models

```
deg <- 2
bsfit3 <- lm(tobinsQ ~ bs(ltdratio, degree = deg, knots = mean(ltdratio)) +
    bs(capexratio, degree = deg, knots = mean(capexratio)) +
    bs(rdratio, degre = deg, knots = mean(rdratio)) + bs(adsratio,
    degree = deg, knots = mean(adsratio)) + bs(pperatio, degree = deg,
    knots = mean(pperatio)) + bs(ebitdaratio, degree = deg, knots = mean(ebitdaratio)) +
    bs(year, degree = deg, knots = mean(year)) + bs(assets, degree = deg,
    knots = mean(assets)) + bs(capex, degree = deg, knots = mean(capex)) +
    bs(ltd, degree = deg, knots = mean(ltd)) + bs(ebitda, degree = deg,
    knots = mean(ebitda)) + bs(ppe, degree = deg, knots = mean(ppe)) +
```

```r
    bs(sales, degree = deg, knots = mean(sales)) + bs(ads, degree = deg,
    knots = mean(ads)) + bs(rd, degree = deg, knots = mean(rd)) +
    bs(bookval, degree = deg, knots = mean(bookval)) + bs(mv,
    degree = deg, knots = mean(mv)) + as.factor(indclass), data = newdat)

stepfitbs3 <- step(bsfit3, trace = 0)
pander(Anova(bsfit3))
```

Table 11: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
bs(ltdratio, degree = deg, knots = mean(ltdratio))	1811	3	31.4	3.202e-20
bs(capexratio, degree = deg, knots = mean(capexratio))	185.5	3	3.217	0.02182
bs(rdratio, degree = deg, knots = mean(rdratio))	101.7	3	1.765	0.1516
bs(adsratio, degree = deg, knots = mean(adsratio))	89.28	3	1.548	0.1997
bs(pperatio, degree = deg, knots = mean(pperatio))	41.93	3	0.7272	0.5356
bs(ebitdaratio, degree = deg, knots = mean(ebitdaratio))	2880	3	49.94	4.317e-32
bs(year, degree = deg, knots = mean(year))	72.86	3	1.264	0.285
bs(assets, degree = deg, knots = mean(assets))	838.2	3	14.54	1.884e-09
bs(capex, degree = deg, knots = mean(capex))	80.34	3	1.393	0.2427
bs(ltd, degree = deg, knots = mean(ltd))	366.5	3	6.356	0.0002662
bs(ebitda, degree = deg, knots = mean(ebitda))	1012	3	17.55	2.294e-11
bs(ppe, degree = deg, knots = mean(ppe))	503.3	3	8.73	8.806e-06
bs(sales, degree = deg, knots = mean(sales))	417.1	3	7.235	7.576e-05
bs(ads, degree = deg, knots = mean(ads))	119.9	3	2.079	0.1006
bs(rd, degree = deg, knots = mean(rd))	606.5	3	10.52	6.601e-07
bs(bookval, degree = deg, knots = mean(bookval))	23693	3	410.9	2.056e-255
bs(mv, degree = deg, knots = mean(mv))	66171	3	1148	0
as.factor(indclass)	1641	40	2.135	4.093e-05
Residuals	258170	13433	NA	NA

```r
pander(Anova(stepfitbs3))
```

Table 12: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
bs(ltdratio, degree = deg, knots = mean(ltdratio))	1837	3	31.84	1.686e-20
bs(capexratio, degree = deg, knots = mean(capexratio))	149	3	2.584	0.0515
bs(ebitdaratio, degree = deg, knots = mean(ebitdaratio))	3514	3	60.92	4.138e-39
bs(assets, degree = deg, knots = mean(assets))	988.1	3	17.13	4.235e-11
bs(ltd, degree = deg, knots = mean(ltd))	330.9	3	5.737	0.000643
bs(ebitda, degree = deg, knots = mean(ebitda))	1040	3	18.03	1.125e-11
bs(ppe, degree = deg, knots = mean(ppe))	716.6	3	12.42	4.137e-08
bs(sales, degree = deg, knots = mean(sales))	445.5	3	7.724	3.752e-05
bs(rd, degree = deg, knots = mean(rd))	554.2	3	9.608	2.474e-06
bs(bookval, degree = deg, knots = mean(bookval))	24054	3	417	4.768e-259
bs(mv, degree = deg, knots = mean(mv))	68757	3	1192	0
as.factor(indclass)	1940	40	2.523	3.867e-07
Residuals	258628	13451	NA	NA

```
pander(AIC(bsfit3, stepfitbs3))
```

	df	AIC
bsfit3	93	78455
stepfitbs3	75	78443

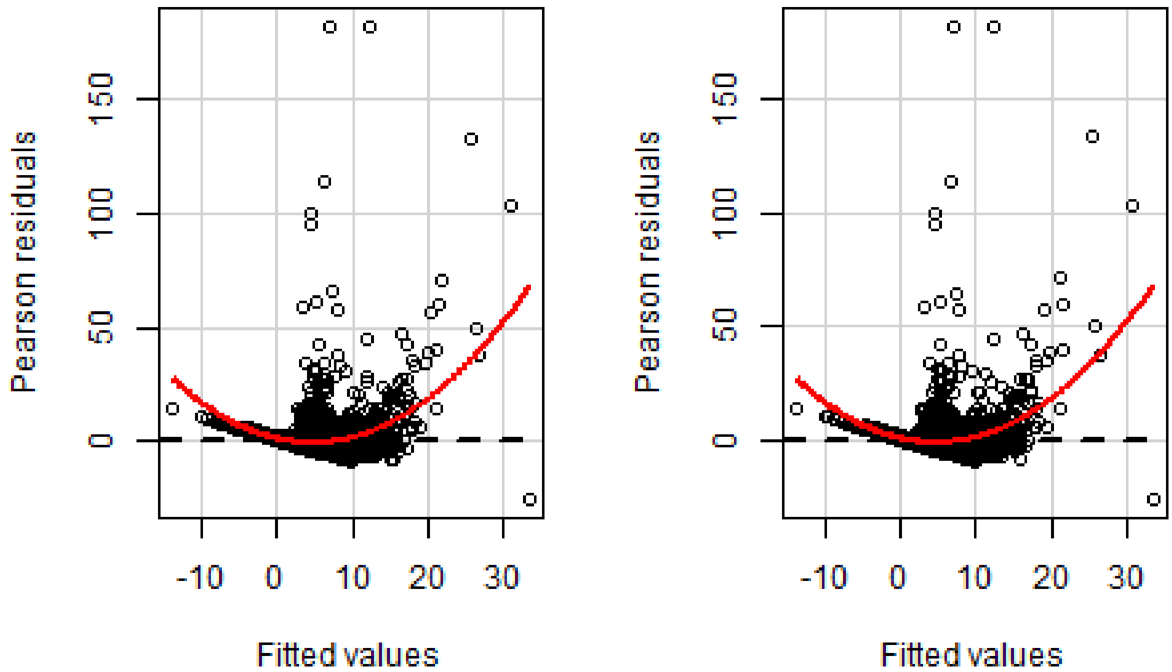
```
pander(BIC(bsfit3, stepfitbs3))
```

	df	BIC
bsfit3	93	79153
stepfitbs3	75	79006

```
par(mfrow = c(1, 2))
residualPlots(bsfit3, terms = ~1)
```

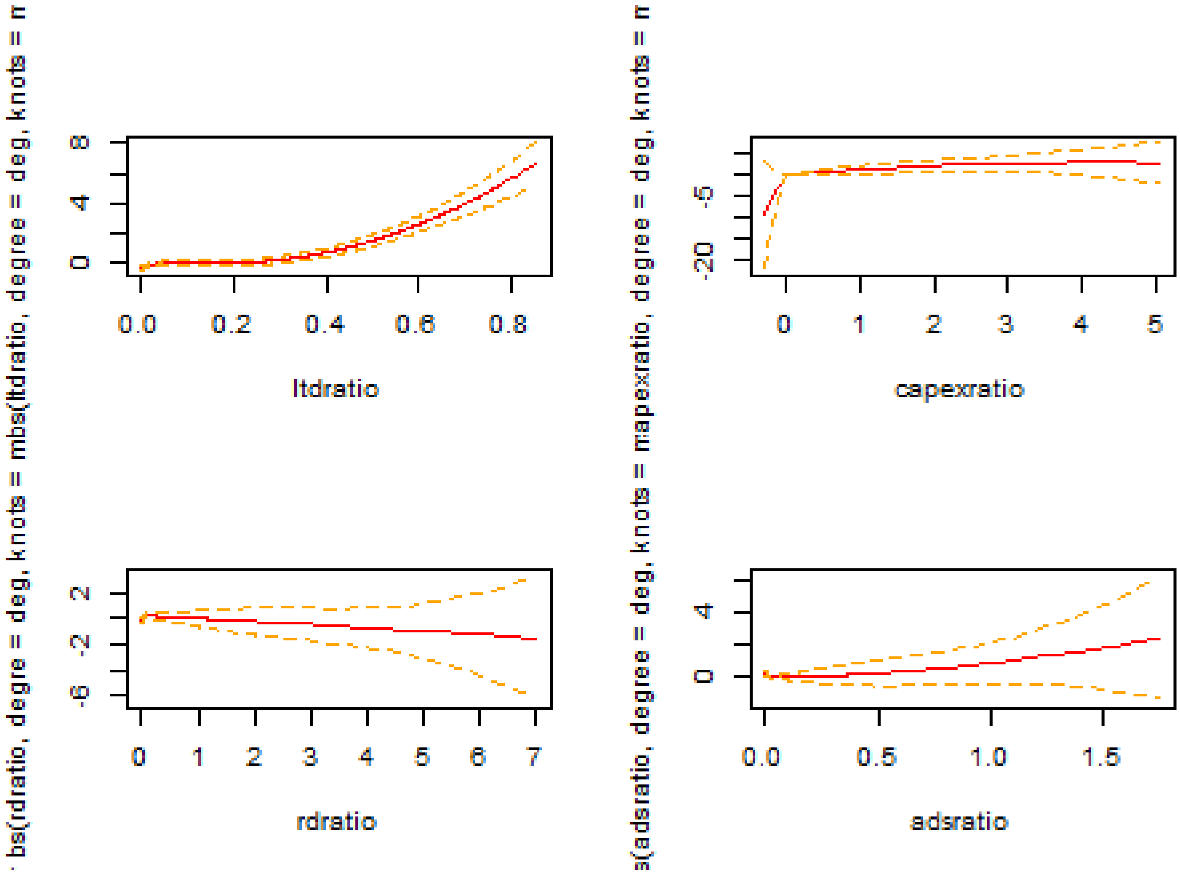
```
           Test stat Pr(>|t|)
Tukey test     59.46        0
```

```r
residualPlots(stepfitbs3, terms = ~1)
```

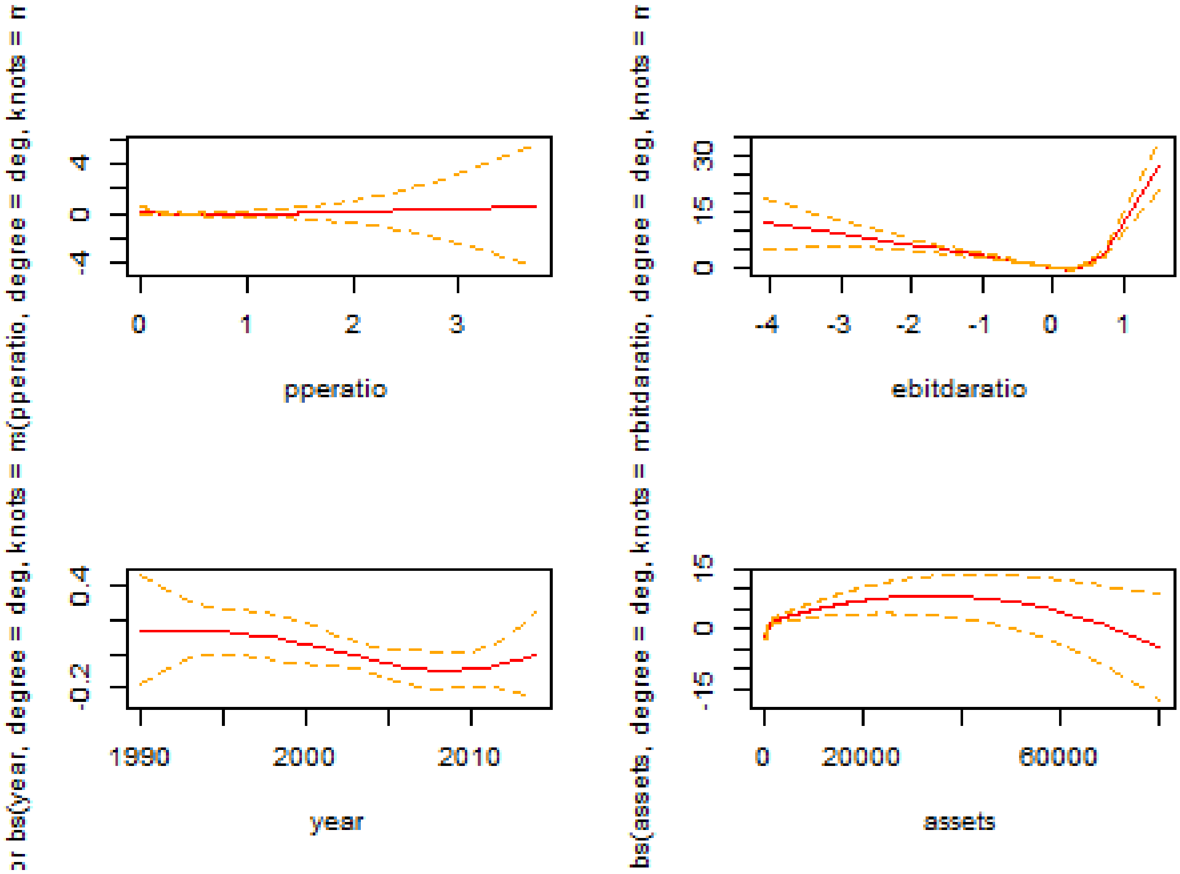



```
           Test stat Pr(>|t|)
Tukey test    58.571        0
```

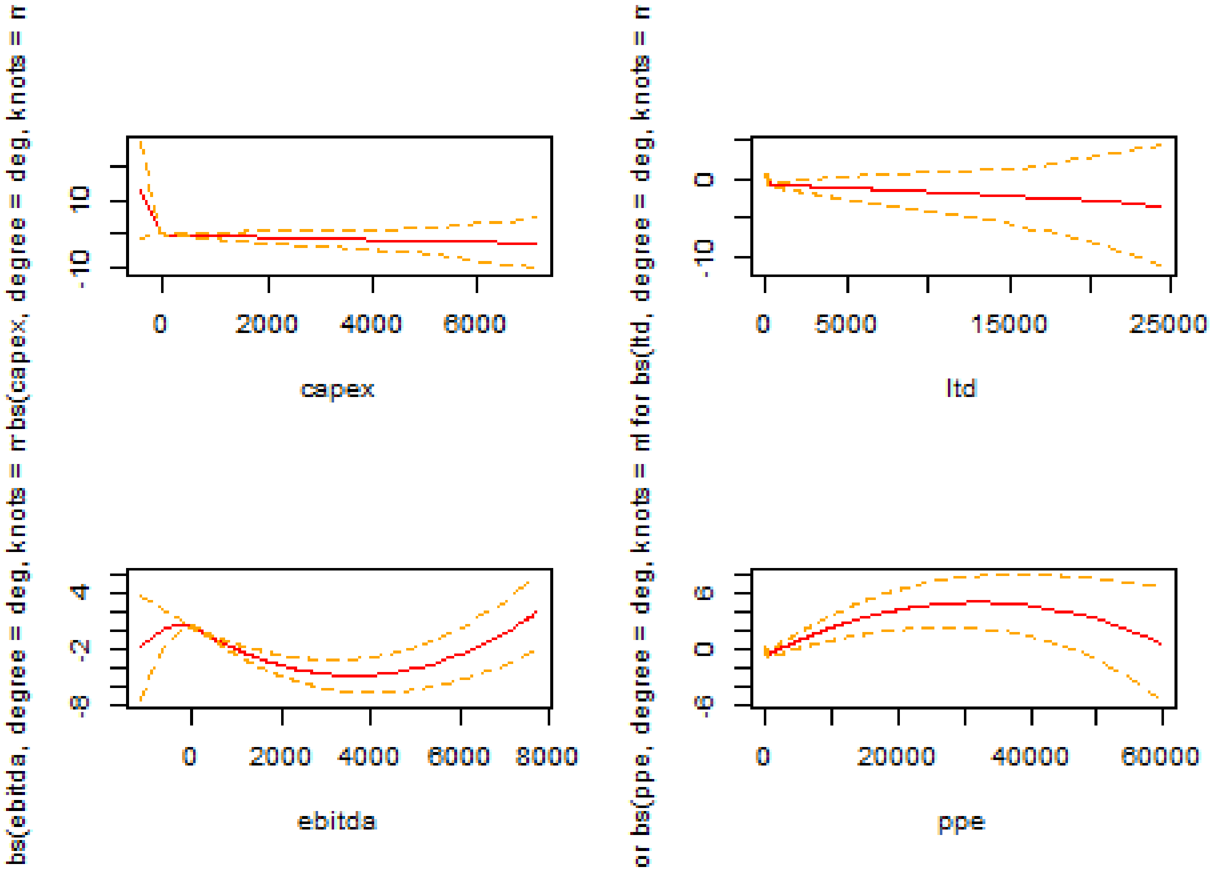
```r
par(mfrow = c(2, 2))

termplot(bsfit3, se = T, ylim = "free")
```

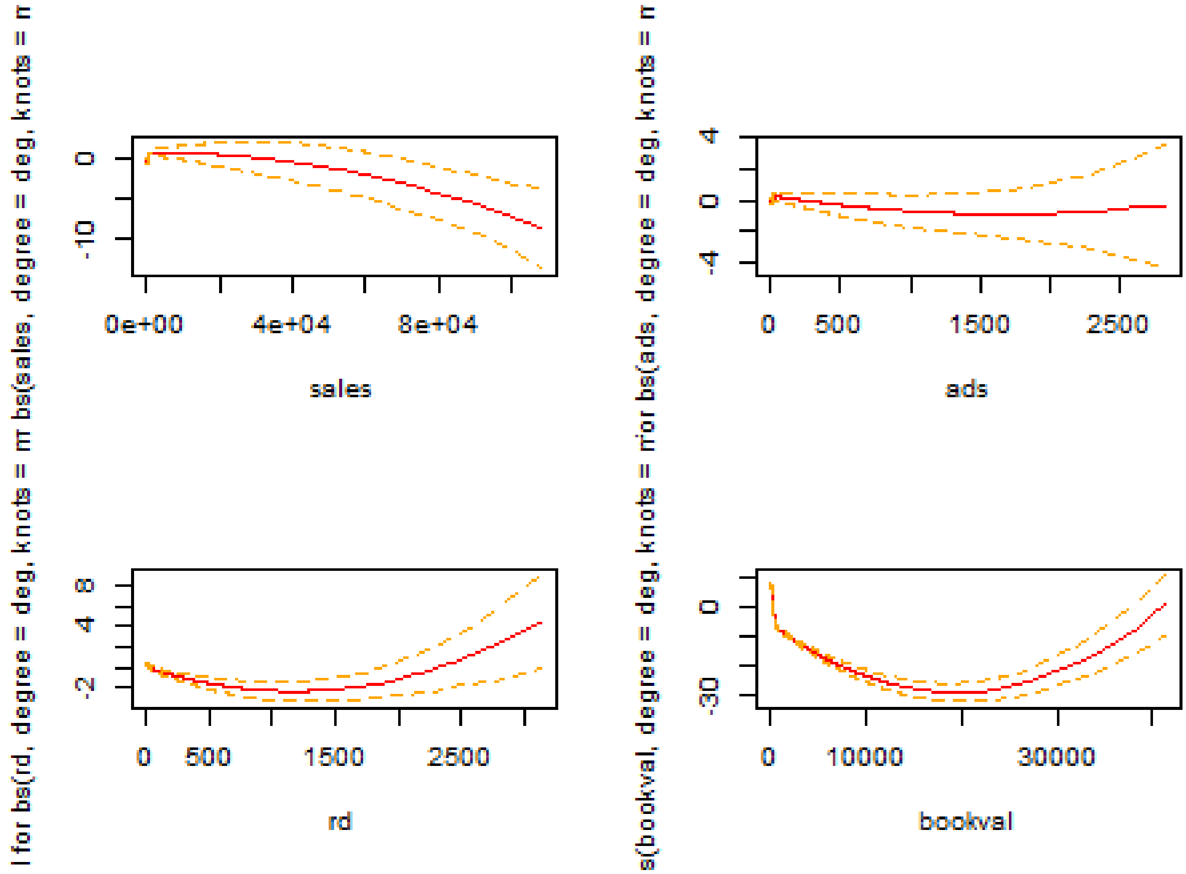


bs(ltdratio, degree = deg, knots = mbs(ltdratio, degree = deg, knots = n
0
4
8
0.0
0.2
0.4
0.6
0.8
ltdratio
bs(rdratio, degree = deg, knots =
-6
-2
2
0
1
2
3
4
5
6
7
rdratio
bs(adsratio, degree = deg, knots = mapexratio, degree = deg, knots = n
-20
-5
0
1
2
3
4
5
capexratio
0
4
0.0
0.5
1.0
1.5
adsratio

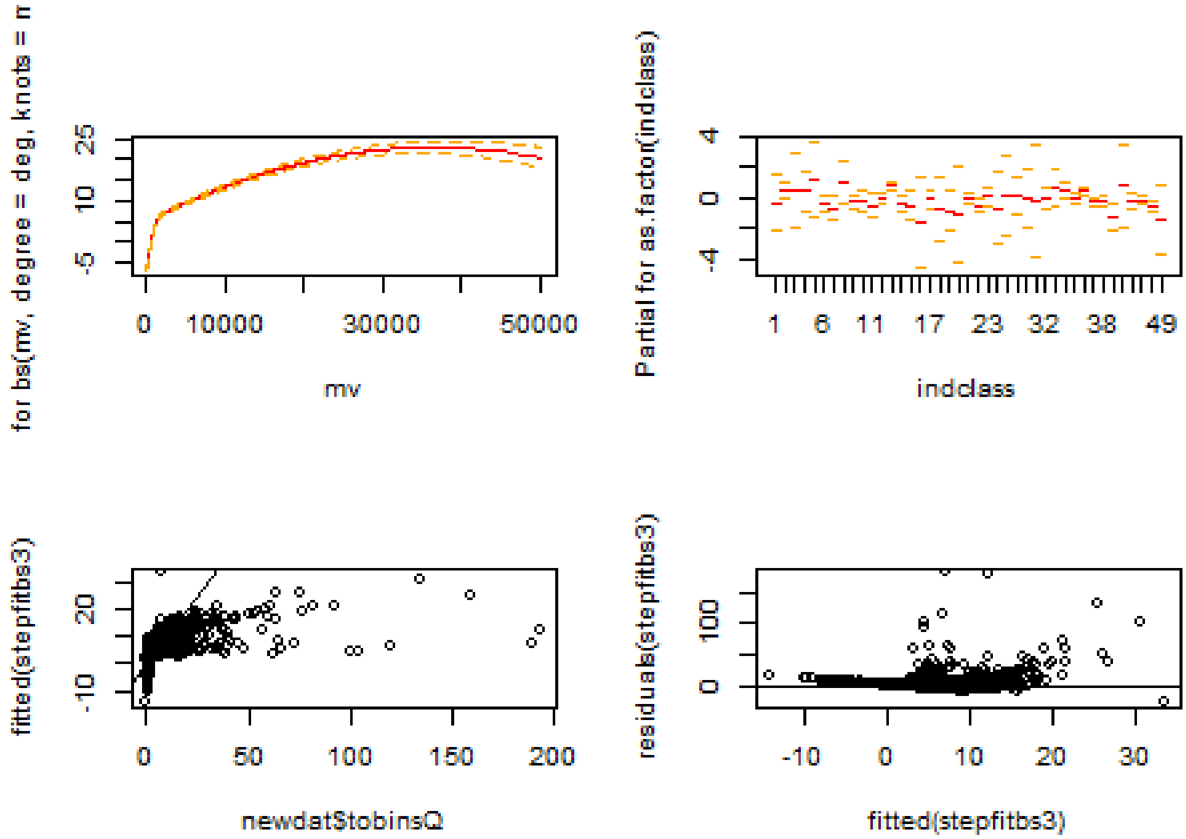


pperatio
ebitdaratio
year
assets


bs(capex, degree = deg, knots = rr
capex
0
2000
4000
6000
-10
10
bs(ltd, degree = deg, knots = rr
ltd
0
5000
15000
25000
-10
0
bs(ebitda, degree = deg, knots = rr
ebitda
0
2000
4000
6000
8000
-8
-2
4
or bs(ppe, degree = deg, knots = rn
ppe
0
20000
40000
60000
-6
0
6




```r
plot(newdat$tobinsQ, fitted(stepfitbs3))
abline(0, 1)
plot(fitted(stepfitbs3), residuals(stepfitbs3))
abline(h = 0)
```

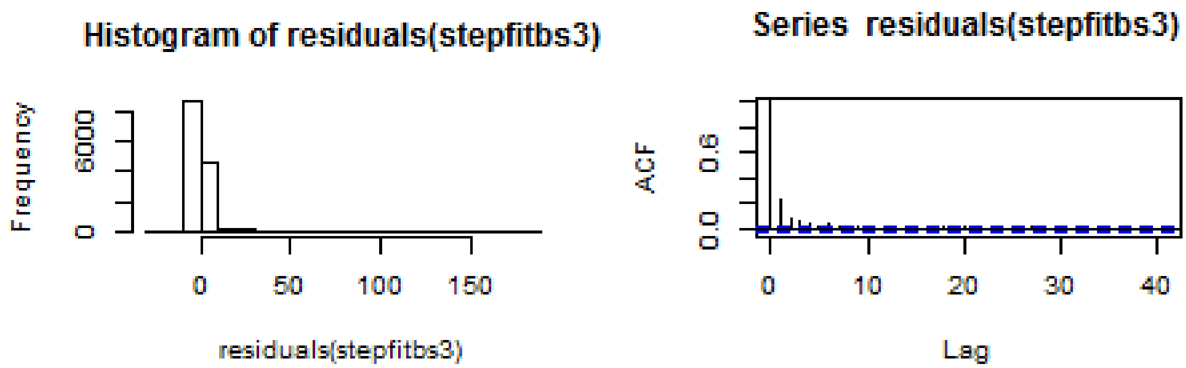



```r
hist(residuals(stepfitbs3))
acf(residuals(stepfitbs3))
ncvTest(stepfitbs3)
```

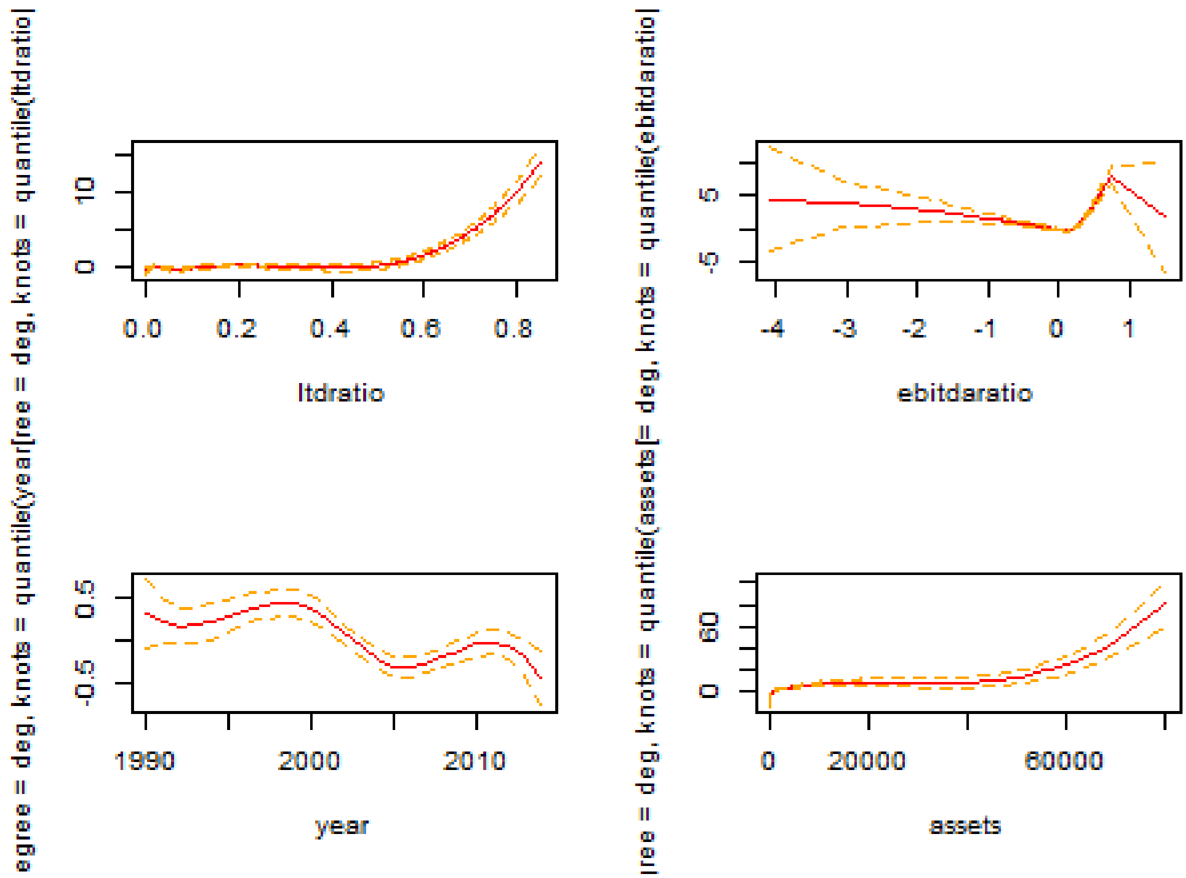
```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 44797.37    Df = 1     p = 0
```

```r
runs.test(residuals(stepfitbs3))
```

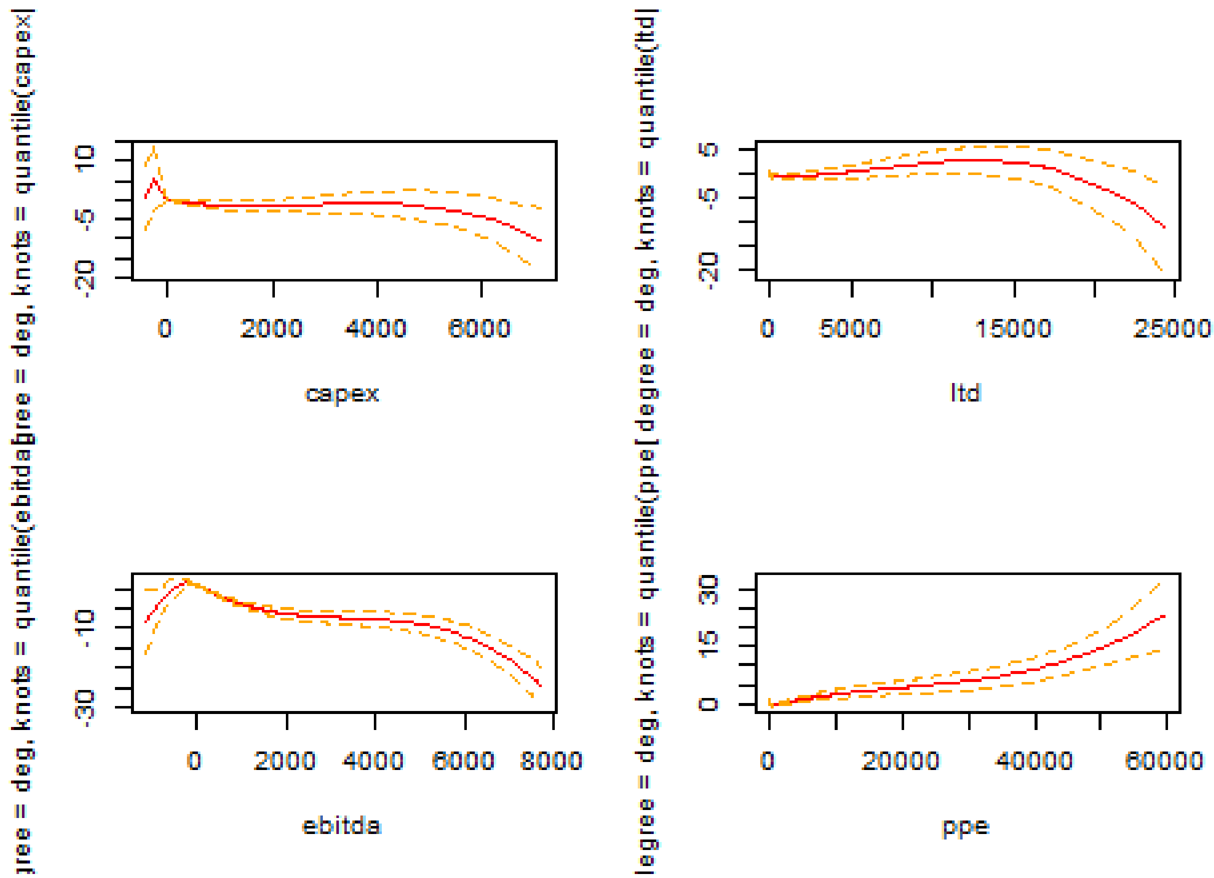
```

    Runs Test - Two sided

data:  residuals(stepfitbs3)
Standardized Runs Statistic = -55.885, p-value < 2.2e-16
```

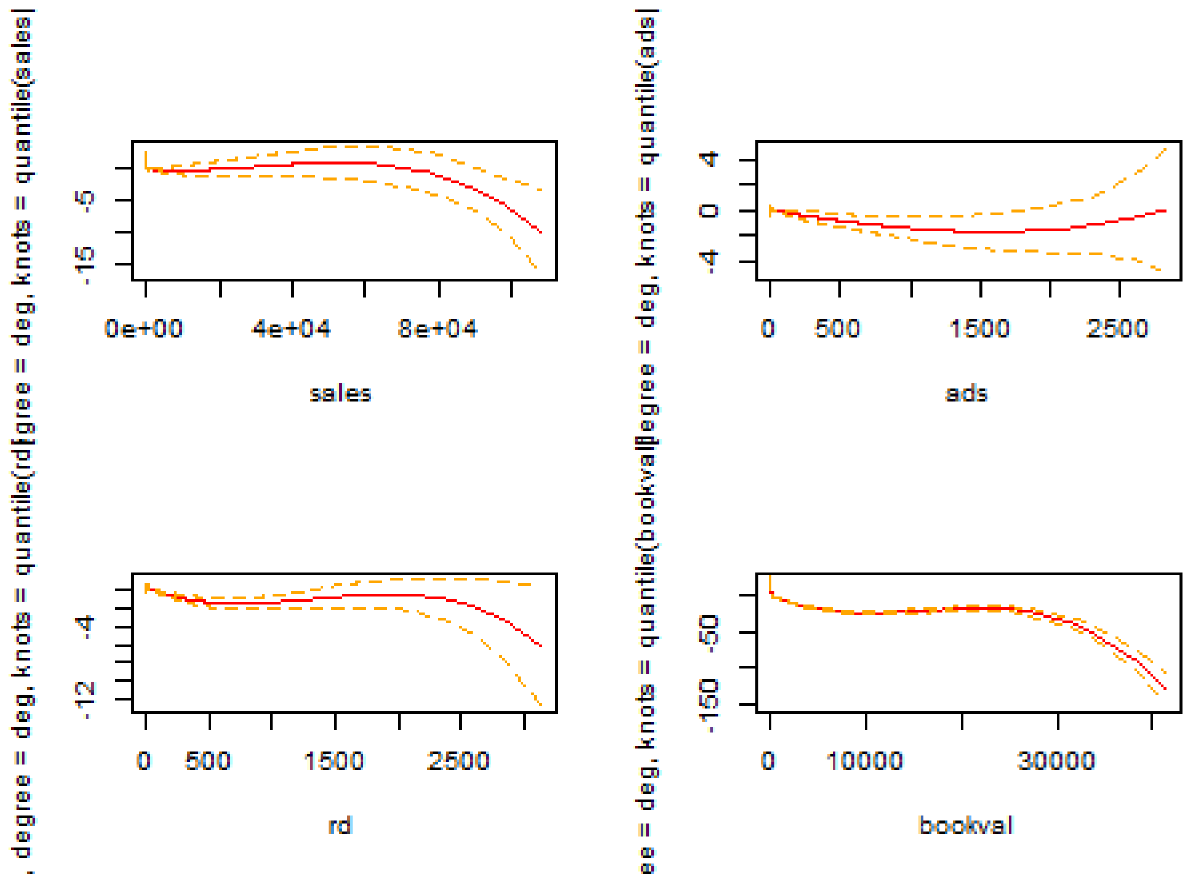



```
deg <- 3
bsfit4 <- lm(tobinsQ ~ bs(ltdratio, degree = deg, knots = quantile(ltdratio[ltdratio >
    min(ltdratio)])[2:3]) + bs(capexratio, degree = deg, knots = quantile(capexratio[capexratio >
    min(capexratio)])[2:3]) + bs(rdratio, degree = deg, knots = quantile(rdratio[rdratio >
    min(rdratio)])[2:3]) + bs(adsratio, degree = deg, knots = quantile(adsratio[adsratio >
    min(adsratio)])[2:3]) + bs(pperatio, degree = deg, knots = quantile(pperatio[pperatio >
    min(pperatio)])[2:3]) + bs(ebitdaratio, degree = deg, knots = quantile(ebitdaratio[ebitdaratio >
    min(ebitdaratio)])[2:3]) + bs(year, degree = deg, knots = quantile(year[year >
    min(year)])[2:3]) + bs(assets, degree = deg, knots = quantile(assets[assets >
    min(assets)])[2:3]) + bs(capex, degree = deg, knots = quantile(capex[capex >
    min(capex)])[2:3]) + bs(ltd, degree = deg, knots = quantile(ltd[ltd >
    min(ltd)])[2:3]) + bs(ebitda, degree = deg, knots = quantile(ebitda[ebitda >
    min(ebitda)])[2:3]) + bs(ppe, degree = deg, knots = quantile(ppe[ppe >
    min(ppe)])[2:3]) + bs(sales, degree = deg, knots = quantile(sales[sales >
    min(sales)])[2:3]) + bs(ads, degree = deg, knots = quantile(ads[ads >
    min(ads)])[2:3]) + bs(rd, degree = deg, knots = quantile(rd[rd >
    min(rd)])[2:3]) + bs(bookval, degree = deg, knots = quantile(bookval[bookval >
    min(bookval)])[2:3]) + bs(mv, degree = deg, knots = quantile(mv[mv >
    min(mv)])[2:3]) + as.factor(indclass), data = newdat)

stepfitbs4 <- step(bsfit4, trace = 0)

par(mfrow = c(2, 2))
termplot(stepfitbs4, se = T, ylim = "free")
```



0
10
0.0
0.2
0.4
0.6
0.8
ltdratio
-5
5
-4
-3
-2
-1
0
1
ebitdaratio
-0.5
0.5
1990
2000
2010
year
0
60
0
20000
60000
assets


gree = deg, knots = quantile(capex
10
-5
-20
0
2000
4000
6000
capex
degree = deg, knots = quantile(ltd
5
-5
-20
0
5000
15000
25000
ltd
gree = deg, knots = quantile(ebitda
-10
-30
0
2000
4000
6000
8000
ebitda
degree = deg, knots = quantile(ppe
30
15
0
0
20000
40000
60000
ppe




```r
pander(Anova(bsfit4))
```

Table 15: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
bs(ltdratio, degree = deg, knots = quantile(ltdratio[ltdratio > min(ltdratio)])[2:3])	3429	5	40.77	8.888e-42
bs(capexratio, degree = deg, knots = quantile(capexratio[capexratio > min(capexratio)])[2:3])	54.21	5	0.6447	0.6656
bs(rdratio, degree = deg, knots = quantile(rdratio[rdratio > min(rdratio)])[2:3])	132.1	5	1.571	0.1644
bs(adsratio, degree = deg, knots = quantile(adsratio[adsratio > min(adsratio)])[2:3])	91.92	5	1.093	0.3618
bs(pperatio, degree = deg, knots = quantile(pperatio[pperatio > min(pperatio)])[2:3])	57.45	5	0.6832	0.6362
bs(ebitdaratio, degree = deg, knots = quantile(ebitdaratio[ebitdaratio > min(ebitdaratio)])[2:3])	3039	5	36.14	6.783e-37
bs(year, degree = deg, knots = quantile(year[year > min(year)])[2:3])	611.6	5	7.273	8.193e-07

	Sum Sq	Df	F value	Pr(>F)
bs(assets, degree = deg, knots = quantile(assets[assets > min(assets)])[2:3])	5245	5	62.37	1.625e-64
bs(capex, degree = deg, knots = quantile(capex[capex > min(capex)])[2:3])	259.8	5	3.089	0.008645
bs(ltd, degree = deg, knots = quantile(ltd[ltd > min(ltd)])[2:3])	401.5	5	4.775	0.0002314
bs(ebitda, degree = deg, knots = quantile(ebitda[ebitda > min(ebitda)])[2:3])	2650	5	31.51	5.083e-32
bs(ppe, degree = deg, knots = quantile(ppe[ppe > min(ppe)])[2:3])	1723	5	20.5	1.909e-20
bs(sales, degree = deg, knots = quantile(sales[sales > min(sales)])[2:3])	262.2	5	3.118	0.008137
bs(ads, degree = deg, knots = quantile(ads[ads > min(ads)])[2:3])	232.7	5	2.768	0.01671
bs(rd, degree = deg, knots = quantile(rd[rd > min(rd)])[2:3])	486.8	5	5.789	2.405e-05
bs(bookval, degree = deg, knots = quantile(bookval[bookval > min(bookval)])[2:3])	44053	5	523.9	0
bs(mv, degree = deg, knots = quantile(mv[mv > min(mv)])[2:3])	68811	5	818.3	0
as.factor(indclass)	1298	40	1.929	0.0003886
Residuals	225348	13399	NA	NA

```r
pander(Anova(stepfitbs4))
```

Table 16: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
bs(ltdratio, degree = deg, knots = quantile(ltdratio[ltdratio > min(ltdratio)])[2:3])	3374	5	40.02	5.567e-41
bs(ebitdaratio, degree = deg, knots = quantile(ebitdaratio[ebitdaratio > min(ebitdaratio)])[2:3])	3158	5	37.45	2.822e-38
bs(year, degree = deg, knots = quantile(year[year > min(year)])[2:3])	763.6	5	9.056	1.315e-08
bs(assets, degree = deg, knots = quantile(assets[assets > min(assets)])[2:3])	6174	5	73.22	6.583e-76
bs(capex, degree = deg, knots = quantile(capex[capex > min(capex)])[2:3])	225.7	5	2.677	0.02007
bs(ltd, degree = deg, knots = quantile(ltd[ltd > min(ltd)])[2:3])	382.9	5	4.54	0.0003876
bs(ebitda, degree = deg, knots = quantile(ebitda[ebitda > min(ebitda)])[2:3])	2895	5	34.34	5.375e-35

	Sum Sq	Df	F value	Pr(>F)
bs(ppe, degree = deg, knots = quantile(ppe[ppe > min(ppe)])[2:3])	1633	5	19.37	2.889e-19
bs(sales, degree = deg, knots = quantile(sales[sales > min(sales)])[2:3])	448.2	5	5.316	6.963e-05
bs(ads, degree = deg, knots = quantile(ads[ads > min(ads)])[2:3])	209.8	5	2.488	0.02928
bs(rd, degree = deg, knots = quantile(rd[rd > min(rd)])[2:3])	786.8	5	9.331	6.914e-09
bs(bookval, degree = deg, knots = quantile(bookval[bookval > min(bookval)])[2:3])	45926	5	544.7	0
bs(mv, degree = deg, knots = quantile(mv[mv > min(mv)])[2:3])	76835	5	911.2	0
Residuals	226974	13459	NA	NA

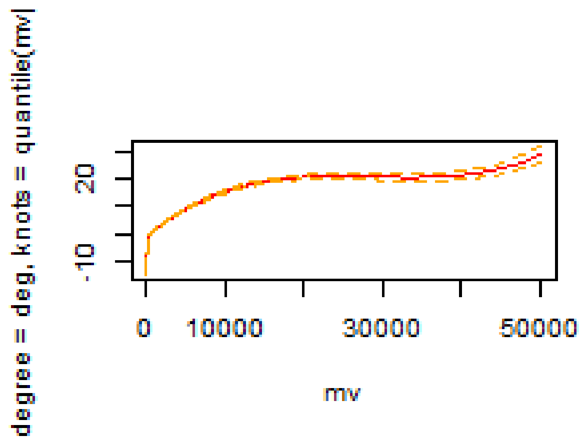
```
pander(AIC(bsfit4, stepfitbs4))
```

	df	AIC
bsfit4	127	76684
stepfitbs4	67	76661

```
pander(BIC(bsfit4, stepfitbs4))
```

	df	BIC
bsfit4	127	77638
stepfitbs4	67	77164

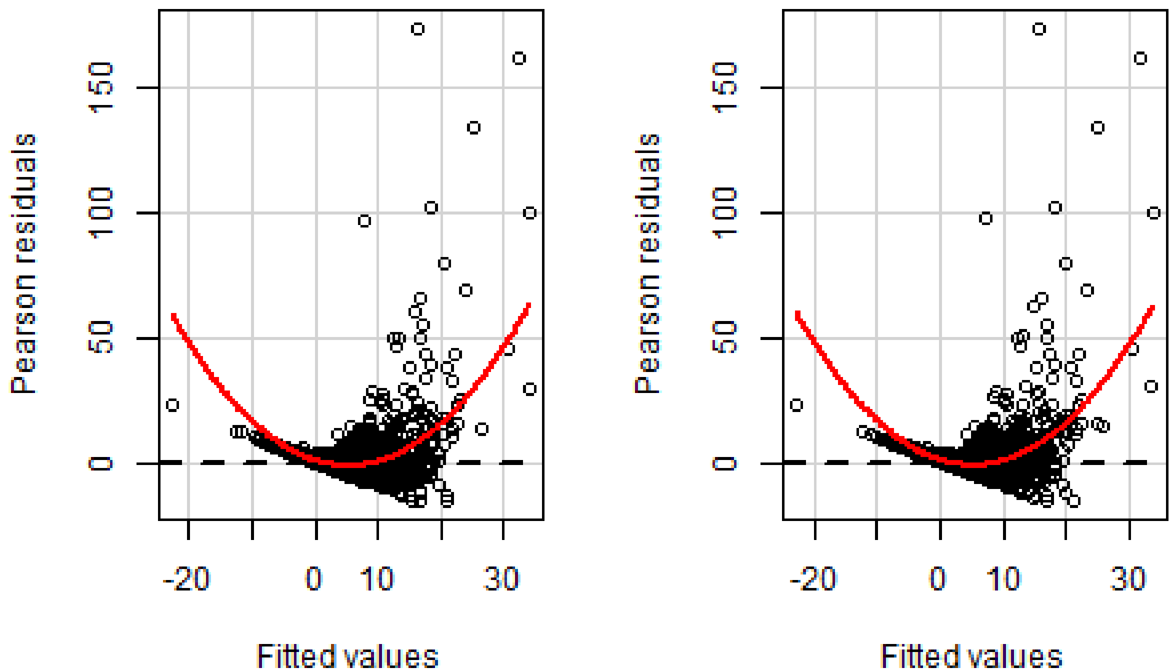
```
par(mfrow = c(1, 2))
```




```r
residualPlots(bsfit4, terms = ~1)
```

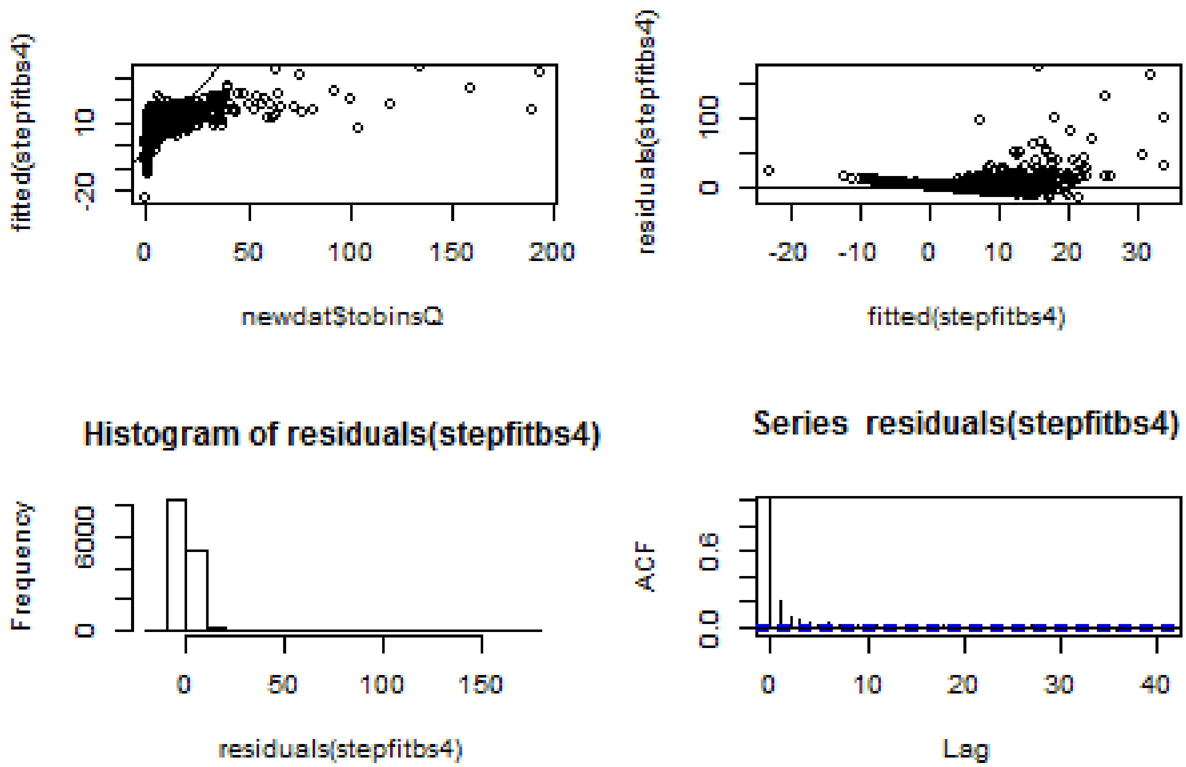
```
           Test stat Pr(>|t|)
Tukey test     85.44        0
```

```r
residualPlots(stepfitbs4, terms = ~1)
```




```
           Test stat Pr(>|t|)
Tukey test    83.616        0
```

```
par(mfrow = c(2, 2))
plot(newdat$tobinsQ, fitted(stepfitbs4))
abline(0, 1)
plot(fitted(stepfitbs4), residuals(stepfitbs4))
abline(h = 0)
hist(residuals(stepfitbs4))
acf(residuals(stepfitbs4))
```




```
ncvTest(stepfitbs4)
```

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 110418.6    Df = 1      p = 0
```

P 小，non-constant

```
runs.test(residuals(stepfitbs4))
```

```
    Runs Test - Two sided

data:  residuals(stepfitbs4)
Standardized Runs Statistic = -50.278, p-value < 2.2e-16
```

Questions

1. Which of the following is FALSE?

符合

   - The models fitted here are inappropriate in that they do not return fitted values which conform to the natural (hard) boundary of the response range. One possible remedy to this is to use a link function which prevents negative predictions being returned.
   - Interestingly, while the covariates that were dropped as part of the stepwise selection procedure which resulted in `polyfit2` did not have concerning `GVIF^(1/(2*Df))` values (e.g. greater than 5), the omission of those variables resulted in a reduction of these `GVIF^(1/(2*Df))` values for some of the other covariates.
   - Despite the use of quadratic polynomials, the `stepfit2` model still shows evidence of ill-fitting behaviour; the Tukey's test for non-additivity for the `stepfit2` model has a small test statistic and a small p-value.
   - The differences in the `stepfit2` model compared with the `polyfit2` model are not suprising when we consider the p-values exhibited by the terms in the `polyfit2` model.

2. TRUE or FALSE? Both the AIC and BIC indicate that the `stepfit2` model is a better model than the `polyfit2` model. While the difference in the number of parameters between these two models is 6, the cost of including these parameters in the model (as quantified using the penalty for the AIC/BIC) outweigh the fit benefits of including these parameters.

3. Which of the following is FALSE?

   - In keeping with all models fitted to date for these data, the largest response values are underpredicted.
   - For all models fitted with graphical diagnostics shown in this document, there is evidence of non-constant error variance both visually and quantified numerically using the Breusch-Pagan test.
   - For all models fitted with graphical diagnostics shown in this document, there is evidence of non-normality in the errors.
   - For all models fitted with graphical diagnostics shown in this document, there is no evidence of correlation in the errors both visually and quantified numerically using the runs test.

4. Which of the following is FALSE?

   - Due to overplotting, the partial relationships for the `stepfit2` model are largely obscured.
   - Despite the `stepfit3` model having 23 additional parameters compared with the `stepfit2` model, both the AIC and BIC scores favour the `stepfit3` model (even though the BIC has a larger penalty than the AIC in this case).
   - While the stepwise model selection results based on the AIC for the quadratic polynomial-based model resulted in the loss of three covariates (compared to the 'full' model), these covariates were retained in the model when cubic polynomials were used instead. This reminds us that model selection does not depend on how each covariate (i.e. in what form) enters the model.
   - The model selection results for the `stepfit3` model and the associated p-values for each term are not surprising since the stepwise results are based on the AIC and it is entirely possible to have terms retained under the AIC (or BIC) when the p-values are in excess of 5%.

5. Which of the following is FALSE?

   - The model fitted using quadratic B-splines fitted with one internal knot at the mean for each covariate returned a better model fit than the corresponding model (with the same covariates) fitted using cubic polynomicals instead. This is true even though both models estimated the same number of parameters which indicates the gains that can be made when parameters are used more wisely.
   - While some covariates in the `bsfit3` model were omitted when subjected to stepwise selection based on the AIC, the `stepfitbs3` model retained at least one covariate with a p-value $> 5\%$.
   - The stepwise results for the model fitted using quadratic B-splines fitted with one internal knot at the mean resulted in the loss of 5 covariates from the model and still the AIC and BIC scores

suggest the model is favoured when these are omitted.
- There are some differences apparent in the fitted functions using B-splines with knots placed at the mean compared with those with knots placed at quantiles.

6. TRUE or FALSE? While knots placed at quantiles and using degree 3 B-splines 'cost' an additional 2 parameters per covariate in this case (compared with a model with degree 2 B-splines and one internal knot placed at the mean of each covariate), the stepwise selected model for the model fitted with splines with knots of this sort, still indicated this model was preferred over the model selected when a knot was placed at the mean for each covariate.

7. TRUE or FALSE? There was a substantial loss of model covariates when B-splines were fitted with knots placed at quantiles compared to the loss in the number of covariates seen with other less complicated models when subjected to stepwise selection. Despite this loss, all terms which remained were statistically significant at the 5% level.