Generating new models from design matrix function
apc package

18 March 2015

Bent Nielsen   Department of Economics, University of Oxford
& Nuffield College
& Institute for Economic Modelling
bent.nielsen@nuffield.ox.ac.uk
http://users.ox.ac.uk/~nuff0078

Contents

1 Introduction

The `apc` package has a set number of models and hypotheses that can be explored. This document describes how alternative hypotheses can be analysed.

An important feature in the `apc` package is that it generates a design matrix. Normally this is kept in the background. The design matrix can, however, be called using the function `apc.get.design` and then modified.

2 Impose functional form on age effect for Belgian lung cancer data

Here we impose particular functional forms on the age effect for Belgian lung cancer data from Clayton and Schifflers (1987a). The analysis follows Nielsen (2014).

Initially we apply the standard analysis to the Belgian lung cancer data, focusing on the age drift model. Subsequently we construct a sub-models of the age drift model with the age-effect is restricted to be first cubic and then quadratic.

First we set up the data

```
> # attach apc library
> library(apc)
> # get data from precoded function
> data.list <- data.Belgian.lung.cancer()
```

The precoded deviance analysis can be run as follows.

```
> # Get a deviance table
> apc.fit.table(data.list,"poisson.dose.response")
```

We also get the fitted parameters of the age drift model.

```
> # Estimate selected model
> apc.fit.ad <- apc.fit.model(data.list,"poisson.dose.response","Ad")
> apc.fit.ad$coefficients.canonical
```

```
              Estimate Std. Error      z value       Pr(>|z|)
level       1.98218357 0.04901861 40.4373684 0.000000e+00
age slope   0.48079205 0.06025798  7.9788936 1.476508e-15
cohort slope 0.08871334 0.01158119  7.6601263 1.857504e-14
DD_age_35  -0.61853253 0.40273616 -1.5358257 1.245811e-01
DD_age_40   0.26923140 0.27445998  0.9809496 3.266176e-01
DD_age_45  -0.18763742 0.19544414 -0.9600565 3.370268e-01
DD_age_50  -0.16690515 0.14277769 -1.1689862 2.424092e-01
DD_age_55  -0.05010697 0.11368129 -0.4407671 6.593816e-01
DD_age_60  -0.08993029 0.09378346 -0.9589141 3.376020e-01
DD_age_65   0.02072825 0.08101869  0.2558452 7.980703e-01
DD_age_70  -0.04428379 0.07371656 -0.6007305 5.480195e-01
DD_age_75  -0.08566734 0.07233153 -1.1843706 2.362664e-01
```

We now want to construct a new design matrix. In order to do this we start by vectorising the data. We then get the corresponding design matrix for the age drift model.

```
> # Vectorise data
> index <- apc.get.index(data.list)
> v.response <- data.list$response[index$index.data]
> v.dose <- data.list$dose[index$index.data]
> # Get design matrix for "Ad" model
> get.design <- apc.get.design(index,"Ad")
> m.design.ad <- get.design$design
> p  <- ncol(m.design.ad)
```

As an aside we should think about the structure of the age drift design matrix. One approach is to inspect the canonical coefficient estimates above, which keep tract of parameter labels. The other approach is to think through the dimensions of the problem. The data has 11 age groups, hence we generate 9 age double differences. With the age drift model we do not generate period and cohort parameters. This is described by the object `get.design$difdif`. The linear plane of the model is unrestricted, hence it needs a level and 2 slopes. The two slopes are set in the age and cohort direction as indicates by `get.design$slopes`. In total we have $p = 12$ parameters.

```
> # Explore this design matrix
> index$age.max

[1] 11

> p

[1] 12

> get.design$difdif

[1] 1 0 0

> get.design$slopes

[1] 1 0 1
```

We achieve a *quadratic* age structure by imposing all age double difference parameters to be equal, see Nielsen and Nielsen (2014, equations 10, 80). We do this by post-multiplying the $(n \times p)$ age drift design matrix, X say with the following $(p \times 4)$-matrix, $H_\perp$ say. This gives the new design matrix $X_H = XH_\perp$. In the code we use the names `m.design.ad`, `m.rest.q`, and `m.design.adq` for X, $H_\perp$ for X_H

$$H_\perp = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}$$

```
> #  Quadractic age effect: restrict double differences to be equal
> m.rest.q  <- matrix(data=0,nrow=p,ncol=4)
> m.rest.q[1,1]          <- 1
> m.rest.q[2,2]          <- 1
> m.rest.q[3,3]          <- 1
> m.rest.q[4:p,4]         <- 1
> m.design.adq         <- m.design.ad %*% m.rest.q
```

Similarly, we achieve a *cubic* age structure by imposing that the age double difference parameters should grow linearly, see Nielsen and Nielsen (2014, equations 10, 81). We do by redefining the restriction matrix $H_\perp$ as a $(p \times 5)$-matrix.

$$
H_\perp = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 1 & 9
\end{pmatrix}
\tag{2}
$$

```
> #  Cubic age effect: restrict double differences to be linear
> m.rest.c          <- matrix(data=0,nrow=p,ncol=5)
> m.rest.c[1,1]          <- 1
> m.rest.c[2,2]          <- 1
> m.rest.c[3,3]          <- 1
> m.rest.c[4:p,4]         <- 1
> m.rest.c[4:p,5]         <- seq(1,p-3)
> m.design.adc         <- m.design.ad %*% m.rest.c
```

We can now refit the model with the new design matrices.

```
> # Poisson regression for dose-response and with log link
> fit.ad <- glm.fit(m.design.ad,v.response,
+                 family=poisson(link="log"),offset=log(v.dose))
> fit.adc <- glm.fit(m.design.adc,v.response,
+                 family=poisson(link="log"),offset=log(v.dose))
> fit.adq <- glm.fit(m.design.adq,v.response,
+                 family=poisson(link="log"),offset=log(v.dose))
```

From this we get deviance test statistics. These are asymptotically χ^2 under the Poisson assumption. So we need to find the corresponding degrees of freedom and compare with a χ^2 table, and fiddle a bit with the output.

```
> # Deviance test statistics
> dev.ad.c <- fit.adc$deviance - fit.ad$deviance
> dev.ad.q <- fit.adq$deviance - fit.ad$deviance
> # Degrees of freedom
> df.ad.c <- ncol(m.design.ad) - ncol(m.design.adc)
```

```
> df.ad.q <- ncol(m.design.ad) - ncol(m.design.adq)
> # p-values
> p.ad.c <- pchisq(dev.ad.c,df.ad.c,lower.tail=FALSE)
> p.ad.q <- pchisq(dev.ad.q,df.ad.q,lower.tail=FALSE)
> # Test for cubic restriction
> fit.tab<-matrix(nrow=2,ncol=3)
> colnames(fit.tab)<-c("LR.vs.Ad","df.vs.Ad","prob(>chi_sq)")
> rownames(fit.tab)<-c("cubic","quadratic")
> fit.tab[1,1:3] <- c(dev.ad.c,df.ad.c,p.ad.c)
> fit.tab[2,1:3] <- c(dev.ad.q,df.ad.q,p.ad.q)
> fit.tab

          LR.vs.Ad df.vs.Ad prob(>chi_sq)
cubic     4.984843        7     0.6618130
quadratic 12.863554       8     0.1166313
```

The estimated coefficients are reported below. Note that the three first coordinates determining the linear plane are only little changed.

```
> #  Coefficients
> fit.ad$coefficients

 [1]  1.98218357  0.48079205  0.08871334 -0.61853253  0.26923140 -0.18763742
 [7] -0.16690515 -0.05010697 -0.08993029  0.02072825 -0.04428379 -0.08566734

> fit.adc$coefficients

[1]  1.97341709  0.48757258  0.08877609 -0.14756909  0.01419983

> fit.adq$coefficients

[1]  1.93381602  0.51560821  0.08836867 -0.06213462
```

References

Clayton, D., Schifflers, E. (1987a) Models for temperoral variation in cancer rates. I: age-period and age-cohort models. *Statistics in Medicine* 6, 449-467.

Nielsen, B. (2014) Deviance analysis of age-period-cohort models. *Download*: http://www.nuffield.ox.ac.uk/economics/papers/2014/apc_deviance.pdf.

Nielsen, B., Nielsen, J.P. (2014) Identification and forecasting in mortality models. *Scientific World Journal.* vol. 2014, Article ID 347043, 24 pages. doi:10.1155/2014/347043. *Open access*: http://www.hindawi.com/journals/tswj/2014/347043/.