Reproducing Martínez-Miranda, Nielsen and Nielsen (2016) using the apc package

10 May 2016

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 1

2 Table: Deviance analysis 1

3 Table: Peak forecasts 1

4 Figure: forecasts 2

1 Introduction

The purpose of this vignette is to use the `apc` package version 1.2.1 to reproduce some the result in Martínez-Miranda, Nielsen and Nielsen (2016): *A simple benchmark for mesothelioma projection for Great Britain*, to appear in *Occupational and Environmental Medicine*. This is an update of Martínez Miranda, Nielsen and Nielsen (2015), for which there is also a vignette available. The `apc` package builds on the identification analysis and the forecast theory in Kuang, Nielsen and Nielsen (2008a,b), the development of deviance analysis for general data arrays in Nielsen (2014). The package is discussed in Nielsen (2015).

The data originates from the Health & Safety Executive, see http://www.hse.gov.uk/statistics/tables/index.htm#lung. The data consists of counts of mesothelioma deaths in the UK by age, $25-89$, and period $1968-2013$. This is modelling using a response-only Poisson regression using an age-period-cohort structure. The purpose of analysis is to forecast the future burden of mesothelioma deaths.

The data are available in the `apc` package. They can be called with the command

```
> library(apc)
> data <- data.asbestos.2013()
```

Here `data.asbestos.2013()` is a function that returns a `apc.data.list`. This includes a matrix with the cases (responses) as well as information about the period and age ranges. The original data include information about age groups $0-19$, $20-24$, $25, \ldots 94$, $95+$. The default is to drop the first two age groups and the last six age groups. To see the structure of the function use the code

```
> data.asbestos.2013
```

2 Table: Deviance analysis

The deviances can be reproduced by a single command

```
> apc.fit.table(data,"poisson.response")[1:4,1:6]
```

	deviance	df.residual	prob(>chi_sq)	LR vs.APC	df vs.APC	prob(>chi_sq)
APC	2763.570	2772	0.542	NaN	NaN	NaN
AP	8574.633	2880	0.000	5811.063	108	0.000
AC	2818.120	2816	0.485	54.550	44	0.132
PC	10544.749	2835	0.000	7781.180	63	0.000

3 Table: Peak forecasts

The peak forecasts are reproduced by first getting AC fit, then generating forecasts. When doing this, the most recent cohorts are removed from the data. We will truncate forecast by cohort 1966, corresponding to the last 22 cohorts. Thus, data is truncated by deleting the last 22 cohorts. There are 46 periods and 65 age groups, that is 110=46+65-1 cohorts. The first 46 cohorts are not forecast as they have been run-off. Thus we can potentially forecast 110-46=65-1=64 cohorts.

```
> data.trunc  <- apc.data.list.subset(data,0,0,0,0,0,22,suppress.warning=TRUE)
> fit.ac <- apc.fit.model(data.trunc,"poisson.response","AC")
> forecast <- apc.forecast.ac(fit.ac)
> cat("Peak forecast","\n")

Peak forecast

> print(forecast$response.forecast.per[1:6,])

         forecast       se  se.proc   se.est
per_2014 2056.316 47.70818 45.34663 14.82410
per_2015 2069.817 48.27965 45.49524 16.15883
per_2016 2076.532 48.80844 45.56898 17.48520
per_2017 2079.378 49.32851 45.60020 18.81288
per_2018 2073.990 49.78097 45.54108 20.10359
per_2019 2062.800 50.19340 45.41806 21.36768
```

4 Figure: forecasts

The forecast figure is a bit complex to generate as it compares forecasts from different methods by different authors.

First, we load the forecasts projections by the Health and Safety Executive based on data until 2006. These are from p24 in Tan and Warren (2009, p. 24). In the following matrix, the columns are period, point forecase, lower 5%, upper 95% forecast bands.

```
> v.WT2006    <- c(
+ 2007,    1791,    1715,    1864,
+ 2008,    1835,    1755,    1920,
+ 2009,    1869,    1788,    1953,
+ 2010,    1902,    1817,    1990,
+ 2011,    1926,    1842,    2015,
+ 2012,    1947,    1859,    2042,
+ 2013,    1964,    1874,    2062,
+ 2014,    1979,    1881,    2079,
+ 2015,    1988,    1886,    2099,
+ 2016,    1990,    1885,    2100,
+ 2017,    1988,    1875,    2100,
+ 2018,    1978,    1870,    2100,
+ 2019,    1966,    1851,    2083,
+ 2020,    1945,    1821,    2070,
+ 2021,    1916,    1790,    2045,
+ 2022,    1881,    1753,    2014,
```

```
+ 2023,    1841,    1709,    1984,
+ 2024,    1799,    1668,    1945,
+ 2025,    1745,    1612,    1893,
+ 2026,    1692,    1549,    1839,
+ 2027,    1625,    1485,    1780,
+ 2028,    1557,    1416,    1710,
+ 2029,    1486,    1338,    1639,
+ 2030,    1412,    1268,    1558)
> WT2006       <- matrix(data=v.WT2006, ncol=4,byrow=TRUE)
```

Second, we load the forecasts projections by the Health and Safety Executive based on data until 2010. These are from the file meso06.xls, downloaded Sep 2014 from `www.hse.gov.uk`

```
> v.WT2010     <- c(
+ 2011,    1942,    1866,    2022,
+ 2012,    1965,    1886,    2046,
+ 2013,    1983,    1901,    2069,
+ 2014,    1997,    1913,    2081,
+ 2015,    2003,    1918,    2099,
+ 2016,    2002,    1912,    2101,
+ 2017,    2000,    1904,    2093,
+ 2018,    1989,    1892,    2084,
+ 2019,    1974,    1874,    2076,
+ 2020,    1945,    1849,    2049,
+ 2021,    1916,    1817,    2017,
+ 2022,    1879,    1774,    1990,
+ 2023,    1842,    1740,    1948,
+ 2024,    1797,    1691,    1911,
+ 2025,    1738,    1631,    1849,
+ 2026,    1682,    1574,    1802,
+ 2027,    1614,    1510,    1730,
+ 2028,    1544,    1444,    1655,
+ 2029,    1471,    1364,    1591,
+ 2030,    1398,    1302,    1515)
> WT2010       <- matrix(data=v.WT2010, ncol=4,byrow=TRUE)
```

Third, we need forecasts from an AC model based on data until 2010.

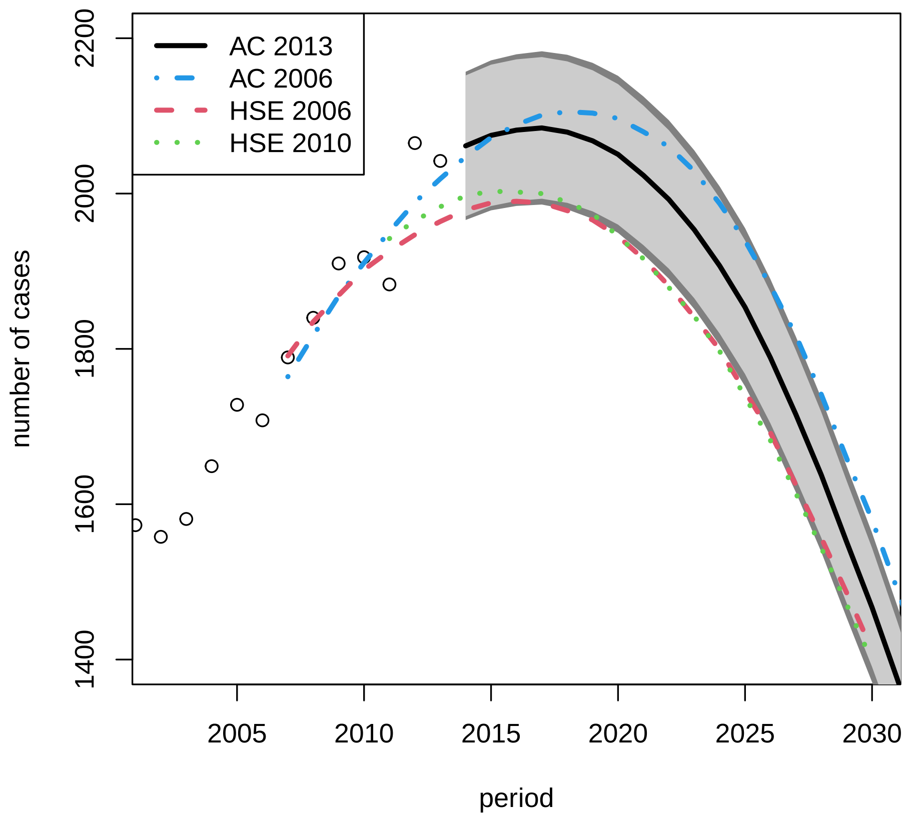
```
> data.trunc.2006 <- apc.data.list.subset(data,0,0,0,7,0,22,
+     suppress.warning=TRUE)
> fit.ac.2006      <- apc.fit.model(data.trunc.2006,
+     "poisson.response","AC")
> forecast.2006    <- apc.forecast.ac(fit.ac.2006)
```

Finally, we need data sums by period

```
> data.sum.per <- apc.data.sums(data.trunc)$sums.per
```

We can then produce the figure in colour

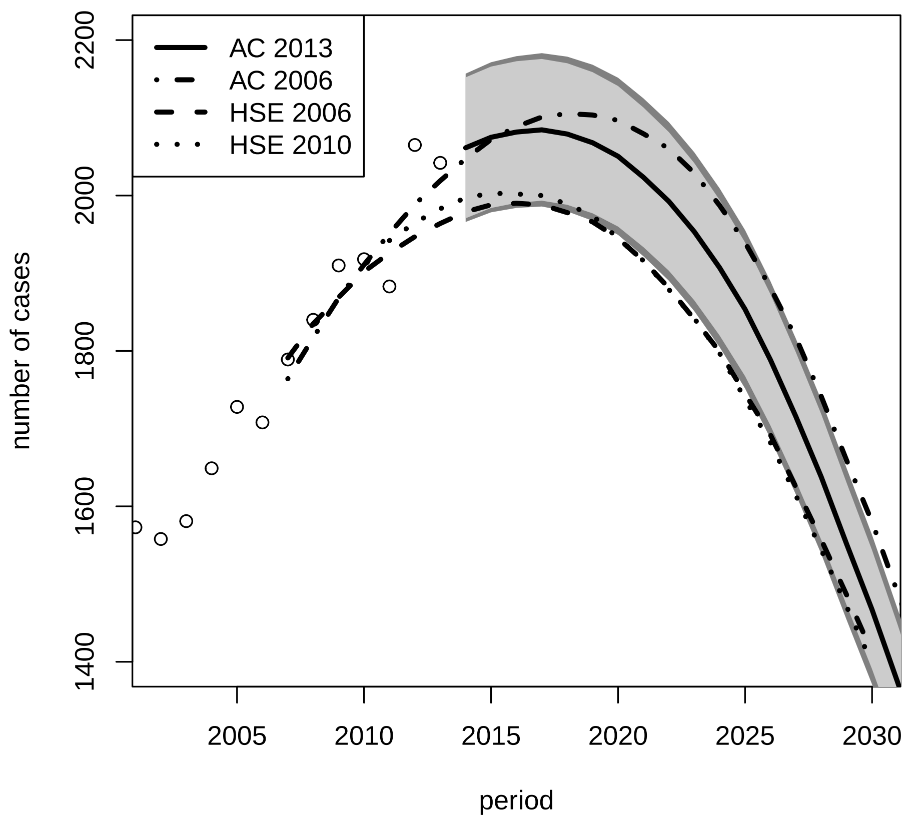
```
> plot(seq(1968,2013),data.sum.per,xlim=c(2002,2030),ylim=c(1400,2200),
+     xlab="period",ylab="number of cases")
> apc.polygon(forecast$response.forecast.per.ic,2013,TRUE,TRUE,
+     col.line=1,lwd.line=3)
> apc.polygon(forecast.2006$response.forecast.per.ic,2006,FALSE,
+     lty.line=4,col.line=4,lwd.line=3)
> apc.polygon(WT2006[,2:4],2006,FALSE,lty.line=2,col.line=2,lwd.line=3)
> apc.polygon(WT2010[,2:4],2010,FALSE,lty.line=3,col.line=3,lwd.line=3)
> legend("topleft",lty=c(1,4,2,3),col=c(1,4,2,3),lwd=3,
+     legend=c("AC 2013","AC 2006","HSE 2006","HSE 2010"))
```



and in black and white

```
> plot(seq(1968,2013),data.sum.per,xlim=c(2002,2030),ylim=c(1400,2200),
+     xlab="period",ylab="number of cases")
> apc.polygon(forecast$response.forecast.per.ic,2013,TRUE,TRUE,
+     col.line=1,lwd.line=3)
> apc.polygon(forecast.2006$response.forecast.per.ic,2006,FALSE,
+     lty.line=4,col.line=1,lwd.line=3)
```

```
> apc.polygon(WT2006[,2:4],2006,FALSE,lty.line=2,col.line=1,lwd.line=3)
> apc.polygon(WT2010[,2:4],2010,FALSE,lty.line=3,col.line=1,lwd.line=3)
> legend("topleft",lty=c(1,4,2,3),col=1,lwd=3,
+     legend=c("AC 2013","AC 2006","HSE 2006","HSE 2010"))
```



References

Kuang, D., Nielsen, B. and Nielsen, J.P. (2008a) Identification of the age-period-cohort model and the extended chain ladder model. *Biometrika* 95, 979-986. *Download*: Earlier version: http://www.nuffield.ox.ac.uk/economics/papers/2007/w5/KuangNielsenNielsen07.pdf.

Kuang, D., Nielsen, B. and Nielsen, J.P. (2008b) Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* 95, 987-991. *Download*: Earlier version: http://www.nuffield.ox.ac.uk/economics/papers/2008/w9/KuangNielsenNielsen_Forecast.pdf.

Martínez Miranda, M.D., Nielsen, B. and Nielsen, J.P. (2015) Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society* A 178, 29-55. *Download*: http://www.nuffield.ox.ac.uk/economics/papers/2013/Asbestos8mar13.pdf.

Martínez-Miranda, M.D., Nielsen, B. and Nielsen, J.P. (2016) A simple benchmark for mesothelioma projection for Great Britain. To appear in *Occupational and Environmental Medicine*. *Download*: https://www.nuffield.ox.ac.uk/economics/papers/2016/MartinezMirandaNielsenNielsen_AsbestosBenchmark.pdf.

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. (2015) apc: An R package for age-period-cohort analysis. *R Journal* 7, 52-64. *Download*: https://journal.r-project.org/archive/2015-2/nielsen.pdf.

Tan, E. and Warren, N. (2009) Projection of mesothelioma mortality in Great Britain. Health and Safety Executive, Research Report 728.