An example of how to use the 'influ' package

Nokome Bentley

April 7, 2014

1 The basics

To begin, we use the same TRE 7 data, and the same generalized linear model, as described in Bentley et al (2011),

```
> model = glm(log(catch)~year+log(tows)+month+vessel+target+area,data=subset(tre7,catch>0
```

Load the influ package into the workspace the usual way, `library(influ)`, and create an influence object for the model,

```
>   infYear = Influence$new(model)
```

By default, `new` will use the first term in the model formula, in this case `year`, as the focus term (the term for which indices are calculated and for which influence is based). But you can also specifiy which term should be the focus term,

```
> infMonth = Influence$new(model,focus='month')
```

At present, the `influ` package does not work with `lm` models, but may do in the future.

Each `Influence` object needs to be initialised using the `calc` method before it can be used,

```
> infYear$calc()
```

Once that is done then it is possible to start looking at some plots and metrics. The `summary` attribute is a `data.frame` with a row for each term and columns describing the deviance explained and the influence metrics,

```
> infYear$summary
```

```
      term  k   logLike      aic r2      r2Dev    r2Negel    overall
1 intercept  1 -20311.43 40626.87 NA        NA         NA         NA
2      year 18 -20115.25 40270.50 NA 0.03845820 0.03913303         NA
3 log(tows)  1 -18755.66 37553.32 NA 0.22882729 0.23284257 0.10819353
4     month 11 -17962.59 35989.19 NA 0.10742045 0.10930538 0.07483488
5    vessel 26 -17221.65 34559.31 NA 0.08608236 0.08759286 0.19830290
6    target  1 -16862.87 33843.75 NA 0.03731810 0.03797293 0.10252278
7      area  5 -16673.81 33475.63 NA 0.01861424 0.01894087 0.07690408
        trend
1          NA
2          NA
```

```
3 -0.010744611
4  0.009599966
5  0.039357049
6  0.018372086
7  0.013885641
```
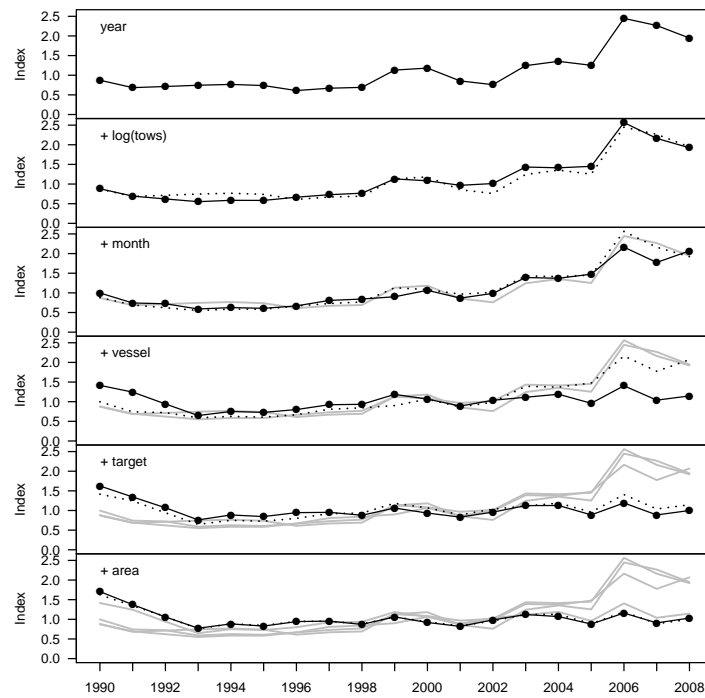
The `stanPlot` method provides a simple plot of both the standardised (with +/- 2 standard errors) and unstandardised indices for the focus term,

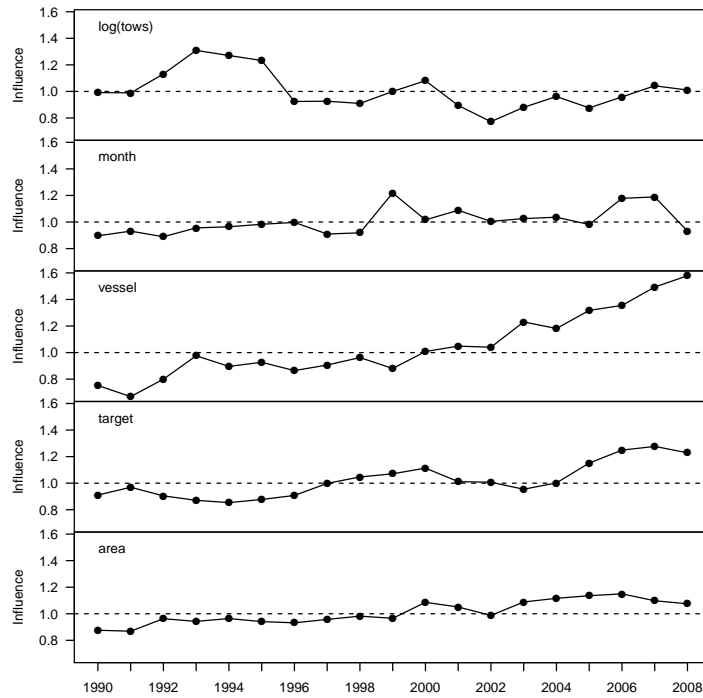```
> infYear$stanPlot()
```
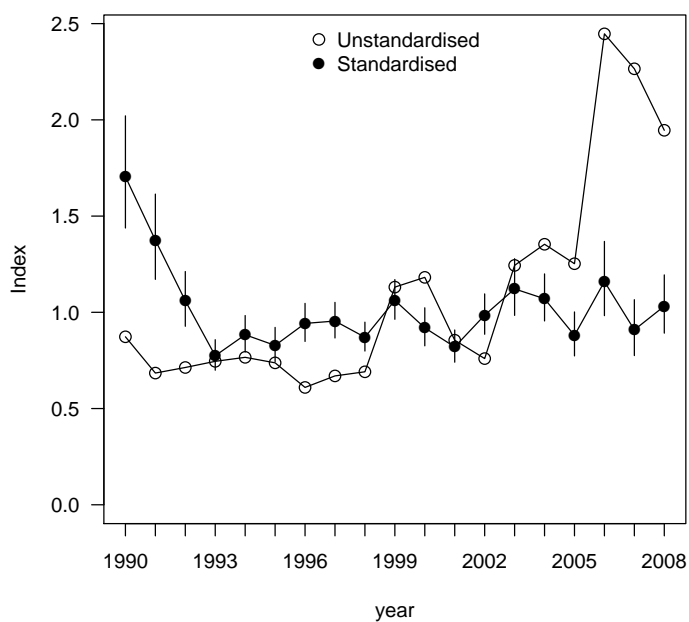


The `stepPlot` method shows standardised indices for the focus term when each of the terms is added to the model (in the order they are specified in the model formula),
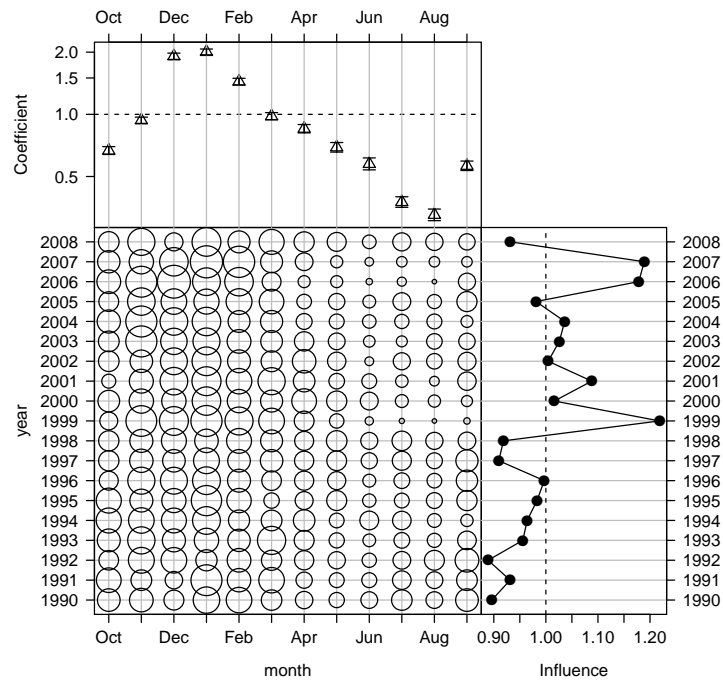
```
> infYear$stepPlot()
```

The `influePlot` method shows the influence of each term for each level of the focus term,

```
> infYear$influPlot()
```

Coefficient-distribution-influence (CDI) plots (see Bentley et al (2011) for a description) can be generated for a single term using the cdiPlot method,
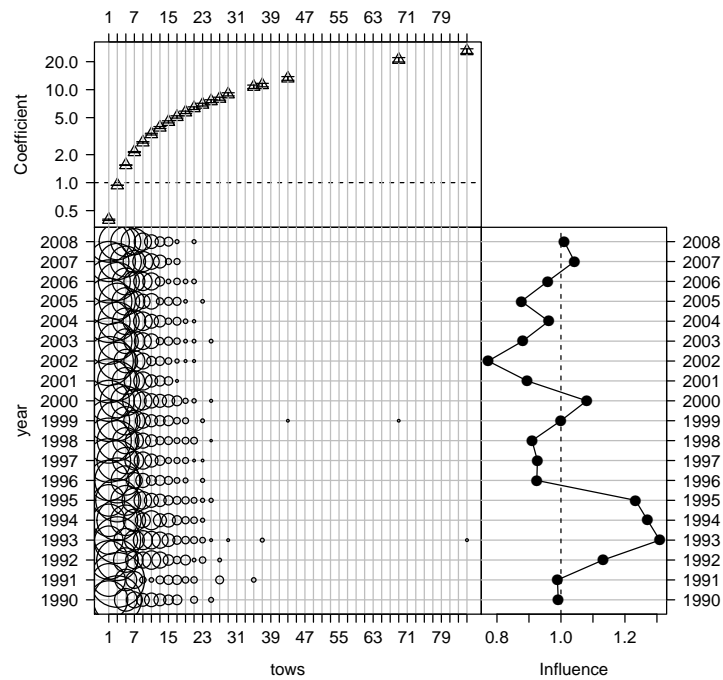
```
> infYear$cdiPlot('month')
```

For some terms in the model that are categorical factors (i.e. do not have a natural order to the levels of the factor) the CDI plot is more intuitive if you specify that the order is in to be based on the coefficients,

```
> infYear$orders['vessel'] = 'coef'
> infYear$cdiPlot('vessel')
```
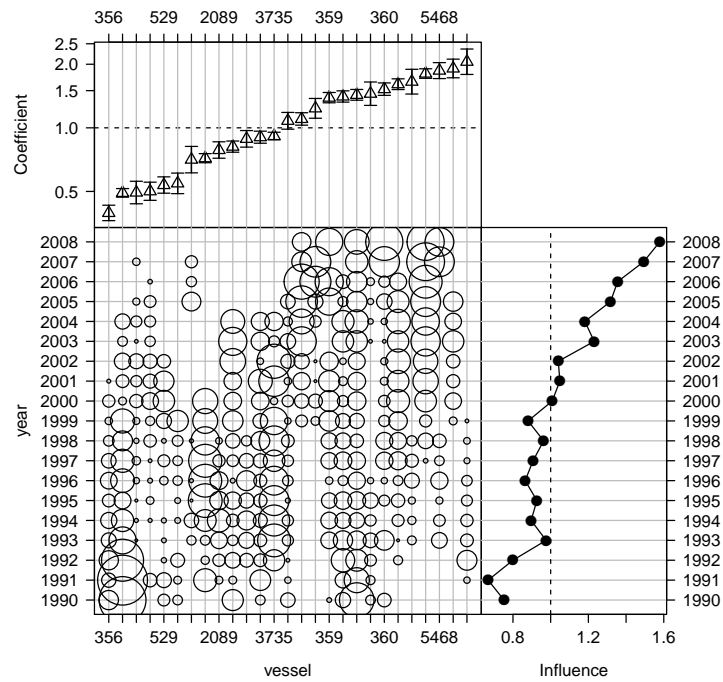
CDI plots can also be generated for continuous terms,

```
> infYear$cdiPlot('log(tows)')
```

Or you can call textttcdiPlotAll, optionally with a callback function, to produce CDI plots for allmodel terms,

```
> infYear$cdiPlotAll(function(term)dev.copy2pdf(file=paste("CDI ",term,".pdf",sep='')))
```

In the above example a PDF file is created for each CDI plot with the name "CDI month.pdf" etc. Note that the option `infYear$order['vessel'] = 'coef'` will still apply to vessel CDI plot created by this call.

2 Alternative types of models

The `influ` package can deal with a limited number of model types other than that produced by `glm`. In most cases you need to also specify the the `data` argument to `Influence$new` (see examples below)

2.1 Negative binomial model using package `MASS`

```
> library(MASS)
> influ = Influence$new(
+   glm.nb(catch~year+month+vessel+target+area,data=tre7),
+   data = tre7
+ )
> influ$calc()
> influ$cdiPlot('month')
```

2.2 Weibull model using package `survival`

When using `survreg` make sure you specify `model=T`:

```
> library(survival)
> influ = Influence$new(
+   survreg(Surv(catch)~year+month+vessel+target+area,data=subset(tre7,catch>0),model=T,di
+   data = subset(tre7,catch>0)
+ )
> influ$calc()
> influ$cdiPlot('month')
```

3 References

Bentley, N., Kendrick, T. H., Starr, P. J., Breen, P. A. (2011) Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardisations. ICES Journal of Marine Science, 69: 84-88.