**Help information for “fg” functions for the Fewster GAM model.**

**Notes on programming**

I’ve separated the model fitting into two functions, fgdata that defines the model, and fgindex that fits it. This is not absolutely essential, but reduces the number of arguments for fgindex and also gives the potential for modifying the dataframe used for predictions, which may be useful for more complex models. We can convert the code to a more conventional R modelling function in the future, if preferred.

The functions are written around 3 dataframes:

* fgglob.data (global) contains the data used to fit the model
* fgglob.preddf (global) contains data for predictions
* newdf (local to fgindex) bootstrapped data

All three of these have the same column names so that they can be used with the same formula. The formula is initially supplied without the spline term (i.e. with year as a factor) and is manipulated within fgindex to give the appropriate spline term (or terms if several are compared).

Bootstrapping uses a for loop as this gives more flexibility than using apply() with no loss of speed.

I’ve used the equals sign for assignment in the programming, rather than the more traditional <- used by R programmers. This is purely for ease of typing and makes no difference to the results.

**Note on convergence (Jan 2020)**

Investigated using gam.control to speed iterations up. With waterways survey for England fitting spline model with usual covariates, standard fit took 48 seconds. Switching to epsilon=1e-05 reduced this to 30s with no change in covariate coefficients. 1e-04 reduced this to 24s, with coefficients only differing in 9th digit. Even 1e-03 (14s) is the same for first 5 or 6 digits. Also investigated the other 3 variables but negligible change to timings.

Therefore introduced an bepsilon argument that sets this for the bootstrap gam fits (not for the real model which keeps the standard settings). Also tried similar change for bootstrap glm, but gain much smaller so left as standard.

Items in red are essential items yet to be programmed. Items in orange are non-essential items which might be considered in the future.

|  |  |
| --- | --- |
| **fgdata** |  |

**Description**

Sets up data structures required to fit Fewster GAM models

**Usage**

fgdata(formula, distribution="poisson", offset)

**Arguments**

|  |  |
| --- | --- |
| formula | a formula describing the model to be fitted, with the first term on the right of the equation indicating the site factor (used to form bootstrap samples) and the second term the years (or other time periods). Note that the spline term should not be specified. A typical example would be count~sites+years+covariate. |
| distribution | The error distribution for the models. At present only “poisson” and “binomial” are supported, bur negative binomial will be added. Names can be abbreviated. |

**Details**

The function sets up a series of global variables which can be accessed by fgindex or fgregion:

* fgglob.data: dataframe containing all the variables required to fit the models, including the response variable.
* fgglob.preddf: similar dataframe with same columns as fgglob.data and one row for each site in each year. This is used to form predictions from the models.
* fgglob.distribution: distribution e.g. “poisson”
* fgglob.formula: formula as supplied to fgdata.
* fgglob.tvar: character strings containing details of the settings.

The function prints a brief summary of the information on the screen. This is primarily intended so that the user can check that the sites and year variables are appropriately set.

**Value**

The function returns the formula, but in practice there is no advantage in saving this. For the future there could be a case for returning a named list containing the items currently stored as global variables. This might give more flexibility, but would make it slightly more difficult to use.

**See Also**

fgindex, fgprint, fgplot.

**Examples**

fgdata(counts~site+year+cov1+cov2, distribution=”pois”)

|  |  |
| --- | --- |
| **fgindex** |  |

**Description**

Fits Fewster GAM models to the data specified using function fgdata.

**Usage**

fgindex(formula, distribution="poisson", offset)

**Arguments**

|  |  |
| --- | --- |
| print | What output to print; “summary” (default) for a summary of results; “model” for a summary of the model; “bootstrap” for details of bootstrap models (mainly intended for tracing bugs and requires nmonitor to be positive); “covariates” for covariate estimates; “none” for no printing. |
| weights | Weights for the analyses. Default is a vector of 1s (i.e. all observations equally weighted). |
| splinedf | Degrees of freedom for the spline term. Default is 0.3 times number of years. Leaving it unset or setting to a negative number gives the default. Note that a saturated model (i.e. a separate estimate for each year) is also fitted. The argument can be set to a list to examine a range of possible levels of smoothing. |
| confidence | Probability value for confidence limits. Default is 95% limits. |
| nboot | Number of bootstrap samples used to estimate variances and confidence limits. Default is zero (i.e. estimate model but with no variance estimates). |
| nmonitor | Which bootstrap samples to print monitoring information for, e.g. setting to 5 will print for samples 1, 6, 11 etc. Default is zero (no monitoring). |
| baseyear | Year to use as 100 in the index calculation. Default is the second year of the series. |
| plot | Whether to plot a graph. Default is FALSE, i.e. no plotting. |
| title | A character string giving a title for the plot and output |
| bsave | Whether to save the bootstrap estimates as fgglob.bsave. Default FALSE. This is used outside the function to create an Excel file to store indices from each bootstrap sample for use in calculating variances for year to year comparisons and the indicator. Up to version 1.2 the option was bfile and wrote the Excel file directly. |
| glmboot | Whether to bootstrap the GLM (saturated model) as well as the GAM model. Default FALSE, which reduces running time but gives no standard errors for the unsmoothed estimates. |
| offset | Offset for the model. The default is a vector of zeros, i.e. no offset. (Offsets are mainly useful where surveys may have variable length, e.g. car surveys). |

**Details**

The function fits the Fewster GAM model using the global variables created by function fgdata. Where the splinedf argument is set to a single value, two models are fitted, one a GAM with the specified degrees of freedom for the time variable (usually years), one a saturated model with years fitted as a factor. Where splinedf is set to several values these are all fitted to compare the fits, but no variance estimates are computed in this case. All spline models use the gam package, rather than package mgcv as originally proposed by Fewster (mgcv could be offered as an option, although it has performed less well in simulations).

Confidence limits are constructed by bootstrapping at the site level as described in the Fewster paper. For this to be appropriate the site variable must be correctly set by putting it first in the formula supplied to function fgdata; this should be carefully checked by looking at the output.

With large datasets the bootstrapping process can be very slow. It may therefore be helpful to set nmonitor to a positive value, indicating that a message should be printed for the first bootstrap sample and every nmonitor thereafter. By default this message just gives the sample number and the time, in order to see how quickly the process is proceeding. However, this can be combined with the setting print=”bootstrap” so that further details are printed, including the sites in the sample and a summary of the fitted model.

**Value**

The function returns a dataframe containing essential details of the index and its confidence limits and other ancillary information. This dataframe can be used as input for functions fgprint and fgplot for further printing and plotting of results. It can also be exported to Excel to produce graphs that can be easily modified by non-users of R. For future development the function could return an S3 object with printing/plotting via appropriate methods.

**See Also**

fgindex, fgprint, fgplot.

**Example**

#fgdata sets up data for fitting model

fgdata(counts~site+year+cov1+cov2, distribution=”pois”)

#fit model with 6 d.f., displaying progress every 10 bootstrap samples

gam6df=fgindex(splinedf=6, nboot=200,nmonitor=10,baseyear=1999,

title="6d.f. with covariates", print=c("model","summary"))

#plot graph (could also be done with plot=TRUE in above

fgplot(gam6df)

|  |  |
| --- | --- |
| **fgplot** |  |

**Description**

Plots results from fgindex.

**Usage**

fgplot(fgresults,xmin,xmax,ymin,yma)

**Arguments**

|  |  |
| --- | --- |
| fgresults | A dataframe of results produced by fgindex |
| xmin | Lowest value for x-axis (time/years) |
| xmax | Highest value for x-axis (time/years) |
| ymin | Lowest value for y-axis (index) |
| ymax | Highest value for y-axis (index) |

Other arguments could be added to control other aspects of the graph if needed – depends whether R or Excel will be used to produce the graphs for the report, and the R skills of those doing it (if highly skilled may prefer to hack the code themselves)

**Details**

The graph is plotted using the usual base graphics commands.

|  |  |
| --- | --- |
| **Fgprint** |  |

**Description**

Prints results from fgindex.

**Usage**

fgprint(fgresults,round)

**Arguments**

|  |  |
| --- | --- |
| fgresults | A dataframe of results produced by fgindex |
| round | Whether to round results before printing (default TRUE rounds the index values to 2 decimal places) |

|  |  |
| --- | --- |
| Fgxls |  |

This function was originally planned, but is no longer required as the output from fgindex can simply be written to Excel using Write.xlsx without the need for a special function.

|  |  |
| --- | --- |
| fgregion |  |

**Description**

Tests regional differences between trends fitted by Fewster GAM models using the data specified using function fgdata.

**Usage**

fgregion(formula, distribution="poisson", offset)

**Arguments**

|  |  |
| --- | --- |
| print | What output to print; “summary” (default) for a summary of results; “model” for a summary of the models. |
| weights | Weights for the analyses. Default is a vector of 1s (i.e. all observations equally weighted). |
| splinedf | Degrees of freedom for the spline term. Default is 0.3 times number of years. |
| pvalue | P-value for significance tests (used to calculate early exits) |
| nrand | Number of randomisations |
| ncheck | Number of randomisations to check for clearly non-significant results |
| nmonitor | Which randomisations to print monitoring information for, e.g. setting to 5 will print for 1, 6, 11 etc. |
| baseyear | Year to use as 100 in the index calculation. Default is the second year of the series. |
| region | Region to use for the tests. |
| plot | Whether to plot a graph. Default is FALSE, i.e. no plotting. |
| title | A character string giving a title for the plot and output |

**Details**

fgregion performs a randomisation test for differences in regions between GAM trends.

|  |  |
| --- | --- |
| fgwald |  |

**Description**

Performs a Wald test for the covariate estimates by accessing the structure fgglob.covest formed by the last call to fgindex.

**Usage**

fgwald(fgresults, covlabs,print=TRUE)

**Arguments**

|  |  |
| --- | --- |
| fgresults | Structure saving results of the model fitted by fgindex. |
| covlabs | Labels for the covariate terms of interest. |
| print | Whether to print output; “TRUE” (default) gives the relevant covariate estimates and Wald test statistics. |
|  |  |

**Details**

fgwald extracts the bootstrapped estimates of the relevant covariates from the structure fgglob.covest and uses them to create a variance-covariance matrix for the estimates, before calculating the Wald test.

|  |  |
| --- | --- |
| fgwald |  |

**Description**

Performs a Wald test for the covariate estimates by accessing the structure fgglob.covest formed by the last call to fgindex.

**Usage**

fgwald(fgresults, covlabs,print=TRUE)

**Arguments**

|  |  |
| --- | --- |
| fgresults | Structure saving results of the model fitted by fgindex. |
| covlabs | Labels for the covariate terms of interest. |
| print | Whether to print output; “TRUE” (default) gives the relevant covariate estimates and Wald test statistics. |
|  |  |

**Details**

fgwald extracts the bootstrapped estimates of the relevant covariates from the structure fgglob.covest and uses them to create a variance-covariance matrix for the estimates, before calculating the Wald test.