Package 'murphydiagram'

October 13, 2022

Type Package

Title Murphy Diagrams for Forecast Comparisons

Version 0.12.2

Date 2019-12-06

Author Alexander Jordan, Fabian Krueger

Maintainer Fabian Krueger <Fabian.Krueger83@gmail.com>

Description Data and code for the paper by Ehm, Gneiting, Jordan and Krueger ('Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings', JRSS-B, 2016 <DOI:10.1111/rssb.12154>).

License GPL-3

URL https://sites.google.com/site/fk83research/code

RoxygenNote 7.0.0

NeedsCompilation no

Repository CRAN

Date/Publication 2019-12-06 15:30:08 UTC

R topics documented:

`Data sets` *Data sets with forecasts and realizations*

Description

Data sets with forecasts and corresponding realizations, as used in the paper by Ehm et al (2016). In the inflation_mean data, the outcome variable is continuous; in the recession_probability data, the outcome is binary.

Usage

```
data(inflation_mean)
data(recession_probability)
```

Format

Both data sets are data frames, with the following layout: First column contains the quarterly date, in string format (e.g. "1998Q4" for the fourth quarter of 1998). The second and third columns contain forecasts by two alternative methods. The fourth column contains realizations.

Source

Forecasts are generated as described in Section 4 of Ehm et al (2016).

Data sources: Inflation - "spf" forecasts and realizations based on data from the Federal Reserve Bank of Philadelphia, http://www.phil.frb.org/research-and-data/real-time-center/ (individual-level CPI forecasts, and real-time data for CPI realizations). "michigan" forecasts based on data from the Michigan Survey of Consumers, https://data.sca.isr.umich.edu/tables.php, Table 32. Recessions - "spf" forecasts and realizations based on data from the Federal Reserve Bank of Philadelphia, http://www.phil.frb.org/research-and-data/real-time-center/ ("anxious index" and real-time data for real GDP growth). The Probit forecasts uses the same real-time data on GDP growth, as well as interest rate data from the Federal Reserve Bank of St. Louis, http://research.stlouisfed.org/fred2/ (series TB3MS and GS10).

Disclaimer: The providers of the raw data take no responsibility for the accuracy of the forecast and realization data sets posted here. Furthermore, the raw data may be revised over time, and the websites linked above should be consulted for the official, most recent versions.

Code and raw data to construct the two data sets can be found at https://sites.google.com/site/fk83research/code.

References

Ehm, W., Gneiting, T., Jordan, A. and Krueger, F. (2016): Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings. *Journal of the Royal Statistical Society (Series B)* 78, 1-29. doi: 10.1111/rssb.12154 (open access).

Examples

```
## Not run:

# Load inflation forecasts
data(inflation_mean)

# Make numeric time axis
tm <- as.numeric(substr(inflation_mean$dt, 1, 4)) +
      0.25*(as.numeric(substr(inflation_mean$dt, 6, 6))-1)

# Plot
matplot(x = tm, y = inflation_mean[,2:4], type = "l", bty = "n",
        xlab = "Time", ylab= "Inflation (percent)", col = 3:1)
legend("topright", legend = c("SPF", "Michigan", "Actual"), fill = 3:1, bty = "n")


## End(Not run)
```

`Fluctuation Test` *Fluctuation test*

Description

Test to analyze whether the ranking of two forecasts is stable over time. The variant implemented here has been proposed in Proposition 1 of Giacomini and Rossi (2010); the critical values are tabulated in their Table 1. The null hypothesis of the test is that both forecasting methods perform equally well (same expected score) at all time points. The alternative is that their performance differs in at least one time point.

Usage

```
fluctuation_test(loss1, loss2, mu = 0.5, dmv_fullsample = TRUE,
                 lag_truncate = 0, time_labels = NULL,
                 conf_level = 0.05)
```

Arguments

`loss1, loss2`	Vectors of losses corresponding to two forecast methods (smaller losses correspond to better forecasts).
`mu`	Size of the rolling window (relative to evaluation sample). Must be in 0.1, 0.2, ..., 0.9.
`dmv_fullsample`	Logical; if `TRUE` (the default), the full sample is used to estimate the variance of the Diebold-Mariano type statistic employed in the test. See page 14/footnote 16 in the working paper version of Rossi (2013).
`lag_truncate`	Truncation lag used when estimating the variance of the Diebold-Mariano type test statistic.

`time_labels` Vector of labels to be used for the time axis. If `NULL` (the default), integer labels are used.

`conf_level` Confidence level, either `0.05` or `0.1` (two-sided test).

Value

List with two elements: 1) Data frame containing the time path of the test statistic, and 2) the relevant critical values. In addition, the function draws a plot which illustrates the test.

Author(s)

Fabian Krueger

References

Giacomini, R. and Rossi, B. (2010): Forecast Comparisons in Unstable Environments. Journal of Applied Econometrics 25, 595-620. doi: 10.1002/jae.1177

Rossi, B. (2013): Advances in Forecasting under Model Instability. In: Handbook of Economic Forecasting, vol. 2, Graham Elliott and Alan Timmermann (eds), pp. 1203-1324. doi: 10.1016/b9780444627315.00021x

Examples

```
# Comparison of Inflation Forecasts:
# Survey of  Professional Forecasters (SPF)
# versus Michigan Survey of Consumers

data(inflation_mean)

# Compute extremal scores of SPF/Michigan (theta = 3)
score_spf <- extremal_score(x = inflation_mean$spf,
                            y = inflation_mean$rlz, theta = 3)
score_michigan <- extremal_score(x = inflation_mean$michigan,
                                 y = inflation_mean$rlz, theta = 3)

# Make simplified label for time axis
tml <- as.numeric(substr(inflation_mean$dt, 1, 4))

# Fluctuation test
fluct_test <- fluctuation_test(score_spf, score_michigan,
                               time_labels = tml, lag_truncate = 4)
```

`Murphy Diagrams` *Murphy diagrams to visualize forecast comparisons*

Description

Visual comparisons of two forecasting methods, allowing to study whether the ranking is robust across the class of elementary or extremal scoring functions. See Ehm et al (2016, esp. Sections 3 and 4) for details.

Usage

```
murphydiagram(f1, f2, y, functional = "expectile", alpha = 0.5,
labels = c("Method 1", "Method 2"), colors =  NULL,
equally_spaced = FALSE)

murphydiagram_diff(f1, f2, y, functional = "expectile",
alpha = 0.5, equally_spaced = FALSE, lag_truncate = 0,
conf_level = 0.95)
```

Arguments

`f1, f2`	Vectors of point forecasts
`y`	Vector of realizing observations.
`functional`	Either "expectile" (the default) or "quantile". Note that the probability of a binary event is an expectile at level `alpha` = 0.5 (see below).
`alpha`	Level of the expectile or quantile, must be between 0 and 1. Defaults to 0.5, which is the mean (if functional is set to "expecile") or median (if functional is set to "quantile").
`labels`	Method labels for murphydiagram to be used in plot legend. Character vector of length two, or `NULL` (in order to omit labels).
`colors`	Colors used. Defaults to NULL, such that the colors are as in Ehm et al (2016). Alternative colors can be specified as a character vector of length two.
`equally_spaced`	Method for choosing the grid of values on the horizontal axis. If set to FALSE (the default), the set of points that is relevant for dominance (c.f. Section 3.4 of the paper) is chosen. This can be somewhat time consuming for large data sets. If set to TRUE, an auxiliary grid of equally spaced points is used.
`lag_truncate`	Largest order of autocorrelation that is accounted for in the variance estimator for murphydiagram_diff (defaults to zero).
`conf_level`	Level of the confidence bands plotted in murphydiagram_diff, defaults to 0.95.

Value

None, used for the effect of creating a plot. `murphydiagram` plots the extremal scores of two forecasting methods. `murphydiagram_diff` plots the difference in the extremal scores of two forecasting methods, together with a confidence interval.

Author(s)

Fabian Krueger

References

Ehm, W., Gneiting, T., Jordan, A. and Krueger, F. (2016): Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings. *Journal of the Royal Statistical Society (Series B)* 78, 1-29. doi: 10.1111/rssb.12154 (open access).

Examples

```
# Comparison of Inflation Forecasts: Survey of Professional Forecasters (SPF)
# versus Michigan Survey of Consumers

data(inflation_mean)
murphydiagram(inflation_mean$spf, inflation_mean$michigan,
inflation_mean$rlz, labels = c("SPF", "Michigan"))
murphydiagram_diff(inflation_mean$spf, inflation_mean$michigan,
inflation_mean$rlz, lag_truncate = 4)
```

`Scoring functions` *Scoring functions*

Description

Implementations of some scoring functions discussed in the paper.

Usage

```
extremal_score(x, y, theta, functional = "expectile", alpha = 0.5)

apl_score(x, y, alpha = 0.5)
ase_score(x, y, alpha = 0.5)
```

Arguments

`x`	Numeric vector of forecasts
`y`	Numeric vector of realizations (same length as `x`)
`theta`	Threshold parameter for extremal score (must be a numeric scalar)
`functional`	String, either "expectile" or "quantile"
`alpha`	Level of the quantile or expectile, must be a numeric scalar in the (0,1) interval

Value

All functions return a vector of scores (same length as `x` and `y`). Smaller scores correspond to better forecasts.

`extremal_score` is the scoring function defined in Equations (10) and (12) of Ehm et al (2016). `apl_score` is the asymmetric piecewise scoring function for quantiles, see Equation (6) in Ehm et al (2016). `ase_score` is the asymmetric squared error for expectiles, see Equation (8) in Ehm et al (2016).

Author(s)

Fabian Krueger

References

Ehm, W., Gneiting, T., Jordan, A. and Krueger, F. (2016): Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings. *Journal of the Royal Statistical Society (Series B)* 78, 1-29. doi: 10.1111/rssb.12154 (open access).

`Synthetic Example` *Analytical Expressions from the Synthetic Example in Section 3.3 and Appendix B*

Description

Functions to compute the analytical expressions in Table 3 of the paper by Ehm et al (2016). These expressions yield the expected score of various forecasters, given the synthetic setup studied in Section 3.3 and Appendix B of the paper. The expressions can be used to replicate Figure 2 in the paper.

Usage

```
expected_score_mean(theta, forecaster = "P")
expected_score_quantile(theta, alpha, forecaster = "P")
```

Arguments

`theta`	Value of the parameter $theta$, indexing the extremal score
`alpha`	Quantile level, between zero and one
`forecaster`	ID of the forecaster, string of length one. Either "P" (perfect forecaster), "C" (climatological forecaster), "U" (unfocused forecaster), or "SR" (sign-reversed forecaster).

Value

Expected value of the extremal score, given the synthetic setup described in Section 3.3 of Ehm et al (2016).

Author(s)

Alexander Jordan, Fabian Krueger

References

Ehm, W., Gneiting, T., Jordan, A. and Krueger, F. (2016): Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings. *Journal of the Royal Statistical Society (Series B)* 78, 1-29. doi: 10.1111/rssb.12154 (open access).

Examples

```
## Not run:
# Color palette, obtained from http://www.cookbook-r.com/Graphs/Colors_
cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73")
cbbPalette <- cbbPalette[c(1, 4, 2, 3)]

# Labeling stuff
forecasters <- c("P", "C", "U", "SR")
names <- c("Perfect", "Climatological", "Unfocused", "Sign-Reversed")
x_label <- expression(paste("Parameter ", theta))

# Figure 2, top left

# Grid for theta
theta_grid1 <- seq(-3, 3, 0.01)
# Expected scores for all forecasters
scores1 <- sapply(forecasters, expected_score_mean, theta = theta_grid1)
# Plot
matplot(x = theta_grid1, y = scores1[, 4:1], type = "l", lty = 1, col = cbbPalette[4:1],
        lwd = 2, bty = "n", xlab = x_label, ylab = expression("Expected Score"))
legend("topright", names, col = cbbPalette, lwd = 2, bty = "n")

## End(Not run)
```

Index