NDK_PCR_GOF

Last Modified on 03/14/2016 11:36 am CDT

- C/C++
- .Net

```
int __stdcall NDK_PCR_GOF ( double **  X,
                            size_t     nXSize,
                            size_t     nXVars,
                            LPBYTE     mask,
                            size_t     nMaskLen,
                            double *   Y,
                            size_t     nYSize,
                            double     intercept,
                            WORD       nRetType,
                            double *   retVal
                          )
```

Returns an array of cells for the i-th principal component (or residuals).

Returns

status code of the operation

Return values

NDK_SUCCESS	Operation successful
NDK_FAILED	Operation unsuccessful. See **Macros** for full list.

Parameters

[in]	**X**	is the independent variables data matrix, such that each column represents one variable
[in]	**nXSize**	is the number of observations (i.e. rows) in X
[in]	**nXVars**	is the number of variables (i.e. columns) in X
[in]	**mask**	is the boolean array to select a subset of the input variables in X. If missing (i.e. NULL), all variables in X are included.
[in]	**nMaskLen**	is the number of elements in mask
[in]	**Y**	is the response or the dependent variable data array (one dimensional array)
[in]	**nYSize**	is the number of elements in Y
[in]	**intercept**	is the constant or the intercept value to fix (e.g. zero). If missing (NaN), an intercept will not be fixed and is computed normally

[in]	**nRetType**	is a switch to select a fitness measure (1 = R-Square (default), 2 = Adjusted R Square, 3 = RMSE, 4 = LLF, 5 = AIC, 6 = BIC/SIC). 1. R-square (coefficient of determination) 2. Adjusted R-square 3. Regression Error (RMSE) 4. Log-likelihood (LLF) 5. Akaike information criterion (AIC) 6. Schwartz/Bayesian information criterion (SIC/BIC)
[out]	**retVal**	is the calculated goodness of fit measure

Remarks

1. The underlying model is described **here**.
2. The coefficient of determination, denoted R^2, provides a measure of how well observed outcomes are replicated by the model. $$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$
3. The adjusted R-square (denoted $\bar R^2$) is an attempt to take account of the phenomenon of the R^2 automatically and spuriously increasing when extra explanatory variables are added to the model. The $\bar R^2$ adjusts for the number of explanatory terms in a model relative to the number of data points. $$\bar R^2 = {1-(1-R^{2}){N-1 \over N-p-1}} = {R^{2}-(1-R^{2}){p \over N-p-1}} = 1 - \frac{\mathrm{SSE}/(N-p-1)}{\mathrm{SST}/(N-1)}$$ Where:
   - p is the number of explanatory variables in the model.
   - N is the number of observations in the sample.
4. The regression error is defined as the square root for the mean square error (RMSE): $$\mathrm{RMSE} = \sqrt{\frac{SSE}{N-p-1}}$$
5. The log likelihood of the regression is given as: $$\mathrm{LLF}=-\frac{N}{2}\left(1+\ln(2\pi)+\ln\left(\frac{\mathrm{SSR}}{N} \right) \right)$$ The Akaike and Schwarz/Bayesian information criterion are given as: $$\mathrm{AIC}=-\frac{2\mathrm{LLF}}{N}+\frac{2(p+1)}{N}$$ $$\mathrm{BIC} = \mathrm{SIC}=-\frac{2\mathrm{LLF}}{N}+\frac{(p+1)\times\ln(p+1)}{N}$$
6. The sample data may include missing values.
7. Each column in the input matrix corresponds to a separate variable.
8. Each row in the input matrix corresponds to an observation.
9. Observations (i.e. row) with missing values in X or Y are removed.
10. The number of rows of the response variable (Y) must be equal to the number of rows of the explanatory variables (X).
11. The MLR_GOF function is available starting with version 1.60 APACHE.

Requirements

```
Header
SFSDK.H

Library
SFSDK.LIB

DLL
SFSDK.DLL
```

References

Hamilton, J .D.; Time Series Analysis , Princeton University Press (1994), ISBN 0-691-04289-6

Tsay, Ruey S.; Analysis of Financial Time Series John Wiley & SONS. (2005), ISBN 0-471-690740

See Also

[template("related")]