NDK_PCR_FITTED

Last Modified on 03/14/2016 11:33 am CDT

- C/C++
- .Net

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

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,out]	**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 the return output 1. fitted values (default), 2. residuals, 3. standardized residuals, 4. leverage (H), 5. Cook's distance.

Remarks

1. li>The underlying model is described **here**.
2. The regression fitted (aka estimated) conditional mean is calculated as follows: \[\hat y_i = E \left[Y| x_i1\cdots x_ip \right] = \alpha + \hat \beta_1 \times x_i1 + \cdots + \beta_p \times x_ip\] Residuals are defined as follows: \[e_i = y_i - \hat y_i \] The standardized (aka studentized) residuals are calculated as follows: \[\bar e_i = \frac{e_i}{\hat \sigma_i} \] Where:
    - \(\hat y\)is the estimated regression value.
    - \(e\) is the error term in the regression.
    - \(\hat e\) is the standardized error term.
    - \(\hat \sigma_i \) is the standard error for the i-th observation.
3. For the influential data analysis, PCR_FITTED computes two values: leverage statistics and Cook's distance for observations in our sample data.
4. Leverage statistics describe the influence that each observed value has on the fitted value for that same observation. By definition, the diagonal elements of the hat matrix are the leverages. \[H = X \left(X^\top X \right)^{-1} X^\top\] \[L_i = h_{ii}\] Where:
    - \(H\) is the Hat matrix for uncorrelated error terms.
    - \(\mathbf{X}\) is a (N x p+1) matrix of explanatory variable where the first column is all ones.
    - \(L_i\) is the leverage statistics for the i-th observation.
    - \(h_{ii}\) is the i-th diagonal element in the hat matrix.
5. Cook's distance measures the effect of deleting a given observation. Data points with large residuals (outliers) and/or high leverage may distort the outcome and accuracy of a regression. Points with a large Cook's distance are considered to merit closer examination in the analysis. \[D_i = \frac{e_i^2}{p \ \mathrm{MSE}}\left[\frac{h_{ii}}{(1-h_{ii})^2}\right]\] Where:
    - \(D_i\) is the Cook's distance for the i-th observation.
    - \(h_{ii}\) is the leverage statistics (or the i-th diagonal element in the hat matrix).
    - \(\mathrm{MSE}\) is the mean square error of the regression model.
    - \(p\) is the number of explanatory variables.
    - \(e_i\) is the error term (residual) for the i-th observation.
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_FITTED 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")]