**Package ‘PDMIF’**

March 30, 2021

**Title** Heterogeneous panel data models with interactive effects

**Version** 0.1.0

**Date** December 5th 2021

**Description** Fits heterogeneous panel data models with interactive effects for linear regression, logit, poisson, probit, quantile, and clustering.

**URL** <https://github.com/Tomohiro-Ando/PDMIF>

**Author** Tomohiro Ando [aut, cre]

Hani Fayad [aut]

**Maintainer** Tomohiro Ando [<T.Ando@mbs.edu>](mailto:%3cT.Ando@mbs.edu%3e)

**Repository** RCAN

Functions in the PDMIF Package**:**

HOMTEST [3](#_bookmark14)

HOMTESTGLM [4](#_bookmark14)

HYPTEST [5](#_bookmark14)

PDMIFCLUST [6](#_bookmark14)

PDMIFCLUST [7](#_bookmark14)

PDMIFCOUNT 8

PDMIFGLM 9

PDMIFLIN 10

PDMIFLING 11

PDMIFLOGIT 1[2](#_bookmark14)

PDMIFPROBIT 13

PDMIFQUANTILE [14](#_bookmark15)

PDMIFQUANTILE [15](#_bookmark15)

HOMTEST *Testing homogeneity of panel data models with interactive effects*

# Description

This function tests homogeneity of the regression coefficients in heterogeneous panel data models with interactive effects.

# Usage

HOMTEST (X, Y, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

Nfactors A pre-specified number of common factors.

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001.

# Details

Under a pre-specified number of common factors, this function tests homogeneity of the regression coefficients.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**p-value** Thep-value of the homogeneity test.

# References

Ando, T. and Bai, J. (2015) A simple new test for slope homogeneity in panel data models with interactive effects. *Economics Letters*, 136, 112-117.

# Example

fit <- HOMTEST(data1X,data1Y,2)

HOMTESTGLM *Testing homogeneity of panel data generalized linear models with interactive effects*

# Description

This function tests homogeneity of the regression coefficients in heterogeneous generalized linear models of panel data with interactive effects.

# Usage

HOMTESTGLM (X, Y, FAMILY, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

FAMILY A description of the error distribution and link function to be used in the model just like in glm functions.

Nfactors A pre-specified number of common factors.

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001.

# Details

Under a pre-specified number of common factors, this function tests homogeneity of the regression coefficients.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**p-value** Thep-value of the homogeneity test.

# References

Ando, T. and Bai, J. (2015) A simple new test for slope homogeneity in panel data models with interactive effects. *Economics Letters*, 136, 112-117.

# Example

fit <- HOMTESTGLM(data2X,data2Y,binomial(link=logit),2)

HYPTEST *Hypothesis testing for coefficients of panel data models*

# Description

This function undergoes hypothesis testing for regression coefficients obtained from the various functions in the package.

# Usage

HYPTEST (B, B0, Se, test ="two", variables=seq(1,nrow(B)), individuals=seq(1,ncol(B)))

# Arguments

B A dataframe of Coefficients as obtained in the output of any function in the package.

B0 A dataframe of hypothetical coefficients to be evaluated in the test. (nrows should match number of variables and ncols should match number of individuals

Se A dataframe of Standard Errors as obtained in the output of any function in the package.

test A string to determine what kind of test to run (“two” for two-tailed, “right” for right-tailed and “left for left-tailed).

variables A list of variables whose coefficients are to be tested. Default is all variables in the B dataframe.

individuals A list of individuals whose coefficients are to be tested. Default is all individuals in the B dataframe.

# Details

Under a pre-specified hypothetical beta, this function tests whether the hypothesis is to be rejected or not based on the model’s estimate

# Value

**pVal** A dataframe of p-values resulting from each individual test.

# Example

fit <- PDMIFLOGIT(data2X,data2Y,2)

HYPTEST(fit$Coefficients,data.frame(c(0,1),c(-1,2)),fit$Se, "two",c(1,3),c(1,2))

PDMIFCLUST *Clustering individuals with panel data models with interactive effects*

# Description

Cluster individuals by heterogeneous panel data models with interactive effects.

# Usage

PDMIFCLUST(X, Y, NGfactors, NLfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

NGfactors A pre-specified number of common factors across groups (see example).

NLfactors A pre-specified number of factors in each groups (see example).

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

Under a pre-specified number of groups and the number of common factors, this function implements clustering for N individuals in the panels. Each of individuals in the group are subject to the group-specific unobserved common factors.

# Value

**Label** The estimated group membership for each of individuals.

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**GlobalFactors** The estimated common factors across groups.

**GroupLoadings** The estimated factor loadings for each group.

**GroupFactors** The estimated group-specific factors.

**GlobalLoadings** The estimated factor loadings for the common factors.

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T. and Bai, J. (2016) Panel data models with grouped factor structure under unknown group membership *Journal of Applied Econometrics*, 31, 163-191.

Ando, T. and Bai, J. (2017) Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112, 1182-1198.

# Examples

fit <- PDMIFCLUST(data5X,data5Y,2,c(2,2,2)

PDMIFCLUSTGLM *Cluster individual units by nonlinear heterogeneous panel data models with interactive effects.*

# Description

Cluster individual units by nonlinear heterogeneous panel data models with interactive effects when the group membership is unknown. Exponential family of distributions are used.

# Usage

PDMIFCLUSTGLM(X, Y, FAMILY, NLfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

FAMILY A description of the error distribution and link function to be used in the model just like in glm functions.

NLfactors A pre-specified number of factors in each groups (see example).

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

Under a pre-specified number of groups and the number of common factors, this function implements clustering for N individuals in the panels using generalized linear models. Each of individuals in the group are subject to the group-specific unobserved common factors.

# Value

**Label** The estimated group membership for each of individuals.

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**GroupLoadings** The estimated factor loadings for each group.

**GroupFactors** The estimated group-specific factors.

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T. and Bai, J. (2016) Panel data models with grouped factor structure under unknown group membership *Journal of Applied Econometrics*, 31, 163-191.

Ando, T. and Bai, J. (2017) Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112, 1182-1198.

# Examples

fit <- PDMIFCLUSTGLM(data6X,data6Y, binomial(link=logit),c(2,2,2)

PDMIFCOUNT *Heterogeneous panel data models with interactive effects for count data*

# Description

Estimate heterogeneous poisson panel data models with interactive effects.

# Usage

PDMIFCOUNT(X, Y, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

Nfactors A pre-specified number of common factors.

maxite A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

This function estimates heterogeneous poisson panel data models with interactive effects.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**Predict** The conditional expectation of response variable

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T., Bai, J. and Li, K. (2021) Bayesian and maximum likelihood analysis of large-scale panel choice models with unobserved heterogeneity, *Journal of Econometrics*.

# Examples

fit <- PDMIFCOUNT(data3X,data3Y,3)

PDMIFGLM *Heterogeneous panel data models with interactive effects for generalized linear models*

# Description

Estimate generalized linear heterogeneous panel data models with interactive effects.

# Usage

PDMIFGLM(X, Y, FAMILY, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

FAMILY A description of the error distribution and link function to be used in the model just like in glm functions.

Nfactors A pre-specified number of common factors.

maxite A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

This function estimates heterogeneous panel data models with interactive effects using generalized linear models.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**Predict** The conditional expectation of response variable**.**

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T., Bai, J. and Li, K. (2021) Bayesian and maximum likelihood analysis of large-scale panel choice models with unobserved heterogeneity, *Journal of Econometrics*.

# Examples

fit <- PDMIFGLM(data2X,data2Y,binomial(link=logit),2)

PDMIFLIN *Heterogeneous panel data models with interactive effects*

# Description

Estimate heterogeneous panel data models with interactive effects.

# Usage

PDMIFLIN(X, Y, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

Nfactors A pre-specified number of common factors.

maxite A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

This function estimates heterogeneous panel data models with interactive effects. This function is similar version of PDMIFLING which accommodates a group structure.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**Predict** The conditional expectation of response variable**.**

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T. and Bai, J. (2015) Asset Pricing with a General Multifactor Structure *Journal of Financial Econometrics*, 13, 556-604.

# Example

fit <- PDMIFLIN(data1X,data1Y,2)

PDMIFLING *Heterogeneous panel data models with interactive effects under known group membership*

# Description

Estimate heterogeneous panel data models with interactive effects under known group membership.

# Usage

PDMIFLING(X, Y, Membership, NGfactors ,NLfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

Membership A pre-specified group membership.

NGfactors A pre-specified number of common factors across groups (see example).

NLfactors A pre-specified number of factors in each groups (see example).

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

Under a known group membership, this function estimates heterogeneous panel data models with interactive effects. Together with the regression coefficients, this function estimates the unobserved common factor structures both for across/within groups.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**GlobalFactors** The estimated common factors across groups.

**GlobalLoadings** The estimated factor loadings for the common factors.

**GroupFactors** The estimated group-specific factors.

**GroupLoadings** The estimated factor loadings for each group.

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T. and Bai, J. (2015) Asset Pricing with a General Multifactor Structure *Journal of Financial Econometrics*, 13, 556-604.

# Example

fit <- PDMIFLING(data4X,data4Y,data4LAB,2,c(2,2,2))

PDMIFLOGIT *Heterogeneous logistic panel data models with interactive effects*

# Description

Estimate heterogeneous logistic panel data models with interactive effects.

# Usage

PDMIFLOGIT(X, Y, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

Nfactors A pre-specified number of common factors.

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

This function estimates heterogeneous loigistic panel data models with interactive effects.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**Predict** Conditional expectation of response variable**.**

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T., Bai, J. and Li, K. (2021) Bayesian and maximum likelihood analysis of large-scale panel choice models with unobserved heterogeneity, *Journal of Econometrics*.

# Example

fit <- PDMIFLOGIT(data2X,data2Y,2)

PDMIFPROBIT *Heterogeneous probit panel data models with interactive effects*

# Description

Estimate heterogeneous probit panel data models with interactive effects.

# Usage

PDMIFPROBIT(X, Y, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

Nfactors A pre-specified number of common factors.

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

This function estimates heterogeneous probit panel data models with interactive effects.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**Predict** Conditional expectation of response variable**.**

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T., Bai, J. and Li, K. (2021) Bayesian and maximum likelihood analysis of large-scale panel choice models with unobserved heterogeneity, *Journal of Econometrics*.

# Example

fit <- PDMIFPROBIT(data2X,data2Y,2)

PDMIFQUANTILE *Heterogeneous quantile panel data models with interactive effects*

# Description

Estimate heterogeneous quantile panel data models with interactive effects.

# Usage

PDMIFQUANTILE(X, Y, TAU, Nfactors, Maxit=100, tol=0.001)

# Arguments

X The (NT) times p design matrix, without an intercept where N=number of individuals, T=length of time series, p=number of explanatory variables.

Y The T times N panel of response where N=number of individuals, T=length of time series.

TAU A pre-specified quantile point.

Nfactors A pre-specified number of common factors.

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

This function estimates heterogeneous quantile panel data models with interactive effects.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**Predict** Estimated quantile point under a given tau.

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T. and Bai, J. (2020) Quantile co-movement in financial markets *Journal of the American Statistical Association*.

# Example

fit <- PDMIFQUANTILE(data7X,data7Y,0.95,5)

PDMIFQVAR *Heterogeneous quantile panel data VAR models with interactive effects*

# Description

Estimate heterogeneous quantile panel data VAR models with interactive effects.

# Usage

PDMIFQVAR(Y, LAG, TAU, Nfactors, Maxit=100, tol=0.001)

# Arguments

Y The T times N panel of response where N=number of individuals, T=length of time series.

LAG The number of lags from yt-1 to yt-LAG used in the VAR.

TAU A pre-specified quantile point.

Nfactors A pre-specified number of common factors.

Maxit A maximum number of iterations in optimization. Default is 100.

tol Tolerance level of convergence. Default is 0.001

# Details

This function estimates heterogeneous quantile panel data models with interactive effects.

# Value

**Coefficients** The estimated heterogeneous coefficients.

**Lower05** Lower end (5%) of the 90% confidence interval of the regression coefficients.

**Upper95** Upper end (95%) of the 90% confidence interval of the regression coefficients.

**Factors** The estimated common factors across groups.

**Loadings** The estimated factor loadings for the common factors.

**Predict** Estimated quantile point under a given tau.

**pval** p-value for testing hypothesis on heterogeneous coefficients.

**Se** Standard error of the estimated regression coefficients.

# References

Ando, T. and Bai, J. (2020) Quantile co-movement in financial markets *Journal of the American Statistical Association*.

# Example

fit <- PDMIFQVAR (data8Y,2,0.1,2)