Package 'InfoTrad'

February 11, 2017

Type Package

Title Calculates the Probability of Informed Trading (PIN)

Version 1.1

Date 2017-02-11

Author Duygu Celik and Murat Tinic

Maintainer Murat Tinic <tinic@bilkent.edu.tr>

Description Estimates the probability of informed trading (PIN) initially introduced by Easley et. al. (1996) <doi:10.1111/j.1540-6261.1996.tb04074.x> . Contribution of the package is that it uses likelihood factorizations of Easley et. al. (2010) <doi:10.1017/S0022109010000074> (EHO factorization) and Lin and Ke (2011) <doi:10.1016/j.finmar.2011.03.001> (LK factorization). Moreover, the package uses different estimation algorithms. Specifically, the grid-search algorithm proposed by Yan and Zhang (2012) <doi:10.1016/j.jbankfin.2011.08.003> , hierarchical agglomerative clustering approach proposed by Gan et. al. (2015) <doi:10.1080/14697688.2015.1023336> and later extended by Ersan and Alici (2016) <doi:10.1016/j.intfin.2016.04.001> .

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2017-02-11 15:17:54

R topics documented:

InfoTrad-package *Calculates the Probability of Informed Trading (PIN)*

Description

Estimates the probability of informed trading (PIN) initially introduced by Easley et. al. (1996) <doi:10.1111/j.1540-6261.1996.tb04074.x> . Contribution of the package is that it uses likelihood factorizations of Easley et. al. (2010) <doi:10.1017/S0022109010000074> (EHO factorization) and Lin and Ke (2011) <doi:10.1016/j.finmar.2011.03.001> (LK factorization). Moreover, the package uses different estimation algorithms. Specifically, the grid-search algorithm proposed by Yan and Zhang (2012) <doi:10.1016/j.jbankfin.2011.08.003> , hierarchical agglomerative clustering approach proposed by Gan et. al. (2015) <doi:10.1080/14697688.2015.1023336> and later extended by Ersan and Alici (2016) <doi:10.1016/j.intfin.2016.04.001> .

Author(s)

Duygu Celik and Murat Tinic

Maintainer: Murat Tinic <tinic@bilkent.edu.tr>

References

D. Easley, N. M. Kiefer, M. O'Hara, and J.B. Paperman. Liquidity, information, and infrequently traded stocks. The Journal of Finance, 51(4):1405-1436, 1996

D. Easley, S. Hvidkjaer, and M. O'Hara. Is information risk a determinant of asset returns? The Journal of Finance, 57(5): 2185-2221, 2002.

D. Easley, S. Hvidkjaer, and M. O'Hara. Factoring information into returns. Journal of Financial and Quantitative Analysis, 45(2): 293-309, 2010.

Ersan, O., and Alici, A. An unbiased computation methodology for estimating the probability of informed trading (PIN). Journal of International Financial Markets, Institutions and Money, 43(1):74-94, 2016.

Q. Gan, W.C. Wei, and D. Johnstone. A faster estimation method for the probability of informed trading using hierarhical agglomerative clustering. Quantitative Finance, 15(11):1805-1821, 2015

H.W. W. Lin and W. C. Ke. A computing bias in estimating the probability of informed trading. Journal of Financial Markets, 14(4):625-640, 2011.

D. Mullner. fastcluster: Fast hierarchical, agglomerative clustering routines for r and python. Journal of Statistical Software, 53(9):1-18, 2013.

Y. Yan and S. Zhang. An improved estimation method and empirical properties of the probability of informed trading. Journal of Banking & Finance, 36(2):454-467, 2012.

EA *Ersan and Alici (2016) Cluster analysis with the altered steps.*

Description

It estimates PIN using Ersan and Alici (2016) modified clustering algorithm.

Usage

```
EA(data, likelihood = c("LK", "EHO"))
```

Arguments

data	Data frame with 2 variables
likelihood	Character strings for likelihood algorithm. Default is "LK".

Details

Argument for data must be a data frame with 2 columns that only contain numbers. Not any other type. You do not have to give names to the columns. We will assign first one as "Buy" and second as "Sell", therefore you should put order numbers with respect to this order.

Value

Returns a list of parameter estimates (output)

alpha	A Number
delta	A Number
mu	A Number
eb	A Number
es	A Number
LikVal	A Number
PIN	A Number

Warning

This function does not handle NA values. Therefore the datasets should not contain any missing value. This function do not conduct the iterative estimation procedure proposed in the same paper.

Author(s)

Duygu Celik and Murat Tinic

References

Ersan, Oguz, and Asli Alici . "An unbiased computation methodology for estimating the probability of informed trading (PIN)." Journal of International Financial Markets, Institutions and Money 43 (2016): 74-94.

Examples

```
# Sample Data
#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy=c(350,250,500,552,163,345,847,923,123,349)
Sell=c(382,500,463,550,200,323,456,342,578,455)
data=cbind(Buy,Sell)

# Parameter estimates using the LK factorization of Lin and Ke (2011)
# with the modified clustering algorithm of Ersan and Alici (2016).
# Default factorization is set to be "LK"

EA(data)
#           alpha     delta       mu epsilon_b epsilon_s   LikVal        PIN
#Output 0.1297379 0.9982431 509.7636   413.941  401.0135 44240.17 0.07506118

# Parameter estimates using the EHO factorization of Easley et. al. (2010)
# with the modified clustering algorithm of Ersan and Alici (2016).

EA(data,likelihood="EHO")
#           alpha     delta       mu epsilon_b epsilon_s   LikVal       PIN
#Output 0.7576503 0.1047257 383.1325   308.831  549.5443 52430.75 0.2527132
```

EHO *Likelihood factorization of Easley et. al. (2010) - EHO Factorization*

Description

The function calculates the likelihood factorization of Easley et. al. (2010) and computes paramaters for estimation of PIN value.

Usage

```
EHO(data, fixed = c(FALSE, FALSE, FALSE, FALSE, FALSE))
```

Arguments

data	Data frame with 2 variables
fixed	Initial values for parameters in the following order: alpha, delta, mu, epsilon_b, epsilon_s

Details

In order to use EHO's return in optimization functions, please **omit** second argument. With this way, EHO will return a function instead of a value. Moreover, argument for data must be a data frame with 2 columns that contain numbers. Not any other type.

Value

LK_out	Returns an optim() object including parameter estimates for the likelihood factorization of Easley et. al. (2010)

Warning

This function does not handle NA values. Therefore the datasets should not contain any missing values.

Author(s)

Duygu Celik and Murat Tinic

References

Easley, D., Hvidkjaer, S., & O'Hara, M. Factoring information into returns. Journal of Financial and Quantitative Analysis, 45(2):293-309,2010.

Examples

```
# Sample Data
#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy<-c(350,250,500,552,163,345,847,923,123,349)
```

```
Sell<-c(382,500,463,550,200,323,456,342,578,455)
data=cbind(Buy,Sell)

# Initial parameter values
# par0 = (alpha, delta, mu, epsilon_b, epsilon_s)
par0 = c(0.5,0.5,300,400,500)

# Call EHO function
LK_out = EHO(data)
model = optim(par0, LK_out, gr = NULL, method = c("Nelder-Mead"), hessian = FALSE)

# The structure of the model output
model

# $par
# [1] 9.111102e-01 1.231429e-04 4.171497e+02 3.360750e+02 4.662539e+02

# $value
# [1] -52831.29

# $counts
# function gradient
#     502       NA

# $convergence
# [1] 1

# $message
# NULL
##

## Parameter Estimates
model$par[1] # Estimate for alpha
# [1] 0.9111102
model$par[2] # Estimate for delta
#[1] 0.0001231429
model$par[3] # Estimate for mu
# [1] 417.1497
model$par[4] # Estimate for eb
# [1] 336.075
model$par[5] # Estimate for es
# [1] 466.2539

## Estimate for PIN
(model$par[1]*model$par[3])/((model$par[1]*model$par[3])+model$par[4]+model$par[5])
# [1] 0.3214394
####
```

Description

It estimates PIN using hierarchical agglomertaive clustering.

Usage

```
GAN(data, likelihood = c("LK", "EHO"))
```

Arguments

data	Data frame with 2 variables
likelihood	Character strings for likelihood algorithm. Default is "LK".

Details

Argument for data must be a data frame with 2 columns that only contain numbers. Not any other type. You do not have to give names to the columns. We will assign first one as "Buy" and second as "Sell", therefore you should put order numbers with respect to this order. This package uses the hclust() function of Mullner (2013) to cluster the data at default settings.

Value

Returns a list of parameter estimates (output)

alpha	A Number
delta	A Number
mu	A Number
eb	A Number
es	A Number
LikVal	A Number
PIN	A Number

Warning

This function does not handle NA values. Therefore, the dataset should not contain any missing values.

Author(s)

Duygu Celik and Murat Tinic

References

D. Mullner. fastcluster: Fast hierarchical, agglomerative clustering routines for r and python. Journal of Statistical Software, 53(9):1-18, 2013.

Gan, Q., Wei, W. C., & Johnstone, D. A faster estimation method for the probability of informed trading using hierarchical agglomerative clustering. Quantitative Finance, 15(11), 1805-1821, 2015.

Examples

```
  # Sample Data
  #   Buy Sell
  #1  350  382
  #2  250  500
  #3  500  463
  #4  552  550
  #5  163  200
  #6  345  323
  #7  847  456
  #8  923  342
  #9  123  578
  #10 349  455

  Buy<-c(350,250,500,552,163,345,847,923,123,349)
  Sell<-c(382,500,463,550,200,323,456,342,578,455)
  data<-cbind(Buy,Sell)

  # Parameter estimates using the LK factorization of Lin and Ke (2011)
  # with the algorithm of Gan et. al. (2015).
  # Default factorization is set to be "LK"

  GAN(data)
#            alpha     delta       mu epsilon_b epsilon_s  LikVal       PIN
# Output 0.3978912 0.9998532 435.8395  259.0336   420.786 44371.1 0.2032458

  # Parameter estimates using the EHO factorization of Easley et. al. (2010)
  # with the algorithm of Gan et. al. (2015)

  GAN(data, likelihood="EHO")
#             alpha    delta       mu epsilon_b epsilon_s   LikVal       PIN
# Output 0.6659866 0.4193557 319.0172  257.0148  359.9236 52826.15 0.2561626
```

LK *Likelihood factorization of Lin and Ke (2011) - LK Factorization*

Description

The function calculates the likelihood factorization of Lin and Ke (2011) and computes paramaters for estimation of PIN value.

Usage

```
LK(data, fixed = c(FALSE, FALSE, FALSE, FALSE, FALSE))
```

Arguments

data	Data frame with 2 variables
fixed	Initial values for parameters in the following order: alpha, delta, mu, epsilon_b, epsilon_s

Details

In order to use LK's return in optimization functions, please **omit** second argument. With this way, LK will return a function instead of a value. Moreover, argument for data must be a data frame with 2 columns that contain numbers. Not any other type.

Value

LK_out	Returns an optim() object including parameter estimates for the likelihood factorization of Lin and Ke (2011)

Warning

This function does not handle NA values. Therefore the datasets should not contain any missing value

Author(s)

Duygu Celik and Murat Tinic

References

Lin, H.W.W. and Ke, W.C. A computing bias in estimating the probability of informed trading. Journal of Financial Markets, 14(4), pp.625-640, 2011.

Examples

```
# Sample Data
#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy<-c(350,250,500,552,163,345,847,923,123,349)
Sell<-c(382,500,463,550,200,323,456,342,578,455)
data=cbind(Buy,Sell)

# Initial parameter values
```

```
# par0 = (alpha, delta, mu, epsilon_b, epsilon_s)
par0 = c(0.5,0.5,300,400,500)

# Call LK function
LK_out = LK(data)
model = optim(par0, LK_out, gr = NULL, method = c("Nelder-Mead"), hessian = FALSE)

## The structure of the model output ##
model

#$par
#[1]   0.480277   0.830850 315.259805 296.862318 400.490830

#$value
#[1] -44343.21

#$counts
#function gradient
#    502       NA

#$convergence
#[1] 1

#$message
#NULL

## Parameter Estimates
model$par[1] # Estimate for alpha
# [1] 0.480277
model$par[2] # Estimate for delta
# [1] 0.830850
model$par[3] # Estimate for mu
# [1] 315.259805
model$par[4] # Estimate for eb
# [1] 296.862318
model$par[5] # Estimate for es
# [1] 434.3046

## Estimate for PIN
(model$par[1]*model$par[3])/((model$par[1]*model$par[3])+model$par[4]+model$par[5])
# [1] 0.178391
####
```

YZ *Yan and Zhang (2012) Grid-Search based PIN Estimates*

Description

It estimates PIN using Yan and Zhang (2012) algorithm.

Usage

```
YZ(data, likelihood = c("LK", "EHO"))
```

Arguments

data	Data frame with 2 variables
likelihood	Character strings for likelihood algorithm. Default is "LK".

Details

Argument for data must be a data frame with 2 columns that only contain numbers. Not any other type. You do not have to give names to the columns. We will assign first one as "Buy" and second as "Sell", therefore you should put order numbers with respect to this order.

Value

Returns a list of parameter estimates (output)

alpha	A Number
delta	A Number
mu	A Number
eb	A Number
es	A Number
LikVal	A Number
PIN	A Number

Warning

This function does not handle NA values. Therefore the datasets should not contain any missing value

Author(s)

Duygu Celik and Murat Tinic

References

Y. Yan and S. Zhang. An improved estimation method and empirical properties of the probability of informed trading. Journal of Banking & Finance, 36(2):454-467, 2012.

Examples

```
# Sample Data
#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
```

```
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy<-c(350,250,500,552,163,345,847,923,123,349)
Sell<-c(382,500,463,550,200,323,456,342,578,455)
data<-cbind(Buy,Sell)

# Parameter estimates using the LK factorization of Lin and Ke (2011)
# with the algorithm of Yan and Zhang (2012).
# Default factorization is set to be "LK"

YZ(data)
#            alpha     delta       mu epsilon_b epsilon_s   Likval      PIN
# Output 0.3526071 0.8937947 438.1156   265.124  425.0304 44371.29 0.180108

# Parameter estimates using the EHO factorization of Easley et. al. (2010)
# with the algorithm of Yan and Zhang (2012).

YZ(data,likelihood="EHO")
#           alpha        delta       mu epsilon_b epsilon_s   Likval      PIN
# Output 0.3605192 5.086071e-05 499.3587   402.306   402.306 53275.94 0.166432
```

Index