Package ‘ATEHonest’

June 21, 2020

Title Honest inference for treatment effects under unconfoundedness

Version 0.1.1

Description Construct matching estimators, and optimal linear estimators, along with confidence intervals for conditional and population average treatment effects under unconfoundedness that are valid in finite samples under the assumption that the regression function satisfies a Lipschitz constraint.

Depends R (>= 4.0.0)

License GPL-3

Imports Matrix,
igraph,
MASS,
stats,
methods

Suggests spelling,
testthat,
CVXR,
lpSolve,
knitr,
rmarkdown,
reshape2,
ggplot2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

VignetteBuilder knitr

Language en-US

BugReports https://github.com/kolesarm/ATEHonest/issues

URL https://github.com/kolesarm/ATEHonest

R topics documented:

`ATTEffBounds` *Efficiency bounds for confidence intervals*

Description

Computes the asymptotic efficiency of two-sided fixed-length confidence intervals at smooth functions, as well as the efficiency of one-sided confidence intervals that optimize a given `beta` quantile of excess length, using the formula described in Appendix A of Armstrong and Kolesár (2018)

Usage

```
ATTEffBounds(res, d, sigma2, C = 1, beta = 0.8, alpha = 0.05)
```

Arguments

`res`	The `res` element of the output of `ATTh`.
`d`	Vector of treatment indicators with length `n`
`sigma2`	Estimate of the conditional variance of the outcome, used to optimize the tuning parameter.
`C`	Lipschitz smoothness constant
`beta`	The quantile `beta` of excess length for determining performance of one-sided CIs.
`alpha`	Level of confidence interval, `1-alpha`.

References

Armstrong, T. B., and M. Kolesár (2018): Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness, Unpublished manuscript

ATTh *Homotopy for average treatment effect for the treated*

Description

Calculates optimal weights m and r for the control and treated observations as a function of δ, or equivalently μ, using the algorithm described in the appendix to Armstrong and Kolesár (2018)

Usage

```
ATTh(
  D0,
  s,
  maxiter = 50,
  check = FALSE,
  tol = .Machine$double.eps * ncol(D0) * nrow(D0)
)
```

Arguments

`D0`	matrix of distances with dimension `[n1 n0]` between untreated and treated units, where `n0` is the number of untreated units and `n1` is the number of treated units
`s`	Set of state variables at which to start the homotopy. If not provided, the homotopy is started at the beginning. The state variables are as follows: **m0** A vector of length `n0` of corresponding to m **r0** A vector of length `n1` of corresponding to r **mu** A scalar corresponding to μ **D** A matrix of effective distances with dimension `[n1 n0]` **Lam** A sparse matrix of Lagrange multipliers with dimension `[n1 n0]` **N0** A sparse matrix of nearest neighbors with dimension `[n1 n0]`
`maxiter`	maximum number of steps in the homotopy. If the homotopy has less steps than `maxiter`, returns the whole solution path.
`check`	check at each step that solution matches `CVXR` (generic convex optimizer) solution.
`tol`	numerical tolerance for rounding error when finding the nearest neighbor. All observations with effective distance within `tol` of the closest are considered to be active.

Value

A list with two elements:

res A matrix with rows corresponding to steps in the homotopy, so that the maximum number of rows is `maxiter`, and columns corresponding to δ, m, r, μ, and `drop`, an indicator if an observations has been dropped from an active set, or added

s List of state variables at the last step with the same structure as the input `s`

References

Armstrong, T. B., and M. Kolesár (2018): Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness, Unpublished manuscript

Examples

```
x0 <- c(0, 1, 2, 3)
x1 <- c(1, 4, 5)
d <- c(rep(FALSE, length(x0)), rep(TRUE, length(x1)))
D0 <- distMat(c(x0, x1), d)
r <- ATTh(D0, maxiter=3)
## Get last, fourth step
ATTh(D0, s=r$s, maxiter=3)

## Check against cvx solution
ATTh(D0, check=TRUE)$res
```

ATTMatchEstimate *Inference on the CATT using the matching estimator*

Description

Computes matching estimator and confidence intervals (CIs) for the CATT. If `ATTMatchPath` used a single `M`, the estimator and CIs are based on a matching estimator with this number of matches. Otherwise, optimize the number of matches from the set in `M` according to `opt.criterion`.

Usage

```
ATTMatchEstimate(
  mp,
  sigma2,
  C = 1,
  sigma2final = sigma2,
  alpha = 0.05,
  beta = 0.8,
  opt.criterion = "RMSE"
)
```

Arguments

`mp`	Output of `ATTMatchPath`
`sigma2`	Estimate of the conditional variance of the outcome, used to optimize the tuning parameter.
`C`	Lipschitz smoothness constant

`sigma2final`	vector of variance estimates with lengthn for determining standard error of the optimal estimators. In contrast, `sigma2` is used only for determining the optimal tuning parameter.
`alpha`	Level of confidence interval, `1-alpha`.
`beta`	The quantile `beta` of excess length for determining performance of one-sided CIs.
`opt.criterion`	One of `"RMSE"` (root mean squared error), `"OCI"` (one-sided confidence intervals), `"FLCI"` (fixed-length two-sided confidence intervals)

Examples

```
Ahalf <- diag(c(0.15, 0.6, 2.5, 2.5, 2.5, 0.5, 0.5, 0.1, 0.1))
D0 <- distMat(NSWexper[, 2:10], Ahalf, method="manhattan", NSWexper$treated)
mp <- ATTMatchPath(NSWexper$re78, NSWexper$treated, D0, M=c(1, 2), tol=1e-12)
## Distance matrix for variance estimation
DM <- distMat(NSWexper[, 2:10], Ahalf, method="manhattan")
sigma2 <- nnvar(DM, NSWexper$treated, NSWexper$re78, J=3)
## Estimator and CI based on a single match
ATTMatchEstimate(mp, mean(sigma2), C=1, sigma2final=sigma2)
```

`ATTMatchPath` *Compute the matching estimator for the CATT*

Description

Computes the matching estimating and matching weights for a range of matches `M`. The output of this function is used as an input for `ATTMatchEstimate` for inference on the CATT.

Usage

```
ATTMatchPath(y, d, D0, M = 1:25, tol = 1e-12)
```

Arguments

`y`	outcome vector with length `n`
`d`	Vector of treatment indicators with length `n`
`D0`	matrix of distances with dimension `[n1 n0]` between untreated and treated units, where `n0` is the number of untreated units and `n1` is the number of treated units
`M`	a vector determining the number of matches. If `Inf`, then use the simple difference in means estimator.
`tol`	numerical tolerance for determining nearest neighbors in constructing matches

Examples

```
Ahalf <- diag(c(0.15, 0.6, 2.5, 2.5, 2.5, 0.5, 0.5, 0.1, 0.1))
D0 <- distMat(NSWexper[, 2:10], Ahalf, method="manhattan", NSWexper$treated)
ATTMatchPath(NSWexper$re78, NSWexper$treated, D0, M=c(1, 2), tol=1e-12)
```

ATTOptEstimate *Optimal estimation and inference for the CATT*

Description

Computes the estimator and confidence intervals (CIs) for the CATT. The tuning parameter is chosen to optimize `opt.criterion` criterion.

Usage

```
ATTOptEstimate(
  op,
  sigma2,
  C = 1,
  sigma2final = sigma2,
  alpha = 0.05,
  beta = 0.8,
  opt.criterion = "RMSE"
)
```

Arguments

`op`	Output of `ATTOptPath`.
`sigma2`	Estimate of the conditional variance of the outcome, used to optimize the tuning parameter.
`C`	Lipschitz smoothness constant
`sigma2final`	vector of variance estimates with lengthn for determining standard error of the optimal estimators. In contrast, `sigma2` is used only for determining the optimal tuning parameter.
`alpha`	Level of confidence interval, `1-alpha`.
`beta`	The quantile `beta` of excess length for determining performance of one-sided CIs.
`opt.criterion`	One of `"RMSE"` (root mean squared error), `"OCI"` (one-sided confidence intervals), `"FLCI"` (fixed-length two-sided confidence intervals)

Examples

```
Ahalf <- diag(c(0.15, 0.6, 2.5, 2.5, 2.5, 0.5, 0.5, 0.1, 0.1))
D0 <- distMat(NSWexper[, 2:10], Ahalf, method="manhattan", NSWexper$treated)
## Distance matrix for variance estimation
DM <- distMat(NSWexper[, 2:10], Ahalf, method="manhattan")
sigma2 <- nnvar(DM, NSWexper$treated, NSWexper$re78, J=3)
## Compute homotopy/solution path, and the class of optimal estimators based
## on the solution path
op <- ATTOptPath(ATTh(D0, maxiter=200)$res, NSWexper$re78, NSWexper$treated)
ATTOptEstimate(op, mean(sigma2), C=1, sigma2final=sigma2, opt.criterion="RMSE")
ATTOptEstimate(op, mean(sigma2), C=1, sigma2final=sigma2, opt.criterion="FLCI")
```

ATTOptPath — *Compute the class of optimal linear estimators for the CATT*

Description

Computes the class of optimal linear estimators that minimize variance subject to a bound on bias, and the optimal weights. The output of this function is used by `ATTOptEstimate` for optimal estimation and inference on the CATT.

Usage

```
ATTOptPath(res, y, d)
```

Arguments

`res`	The `res` element of the output of `ATTh` on which to base the estimate
`y`	outcome vector with length `n`
`d`	Vector of treatment indicators with length `n`

cv — *Critical values for CIs based on a biased Gaussian estimator.*

Description

Computes the critical value $cv_{1-\alpha}(B)$ such that the confidence interval $X \pm cv_{1-\alpha}(B)$ will have coverage $1 - \alpha$, where X is Normally distributed with variance equal to 1 and maximum bias at most B.

Usage

```
cv(B, alpha = 0.05)
```

Arguments

`B`	Maximum bias, a non-negative vector.
`alpha`	Determines CI level, `1-alpha`. Needs to be between 0 and 1. Can be a vector of values.

Value

Critical value

Examples

```
# 90% critical value:
cv(B = 1, alpha = 0.1)
# 95% critical values
cv(B = c(0, 1, 3), alpha = 0.05)
# 96 and 90% critical values
cv(B = 1, alpha = c(0.05, 0.1))
```

---

distMat *Matrix of distances between observations*

Description

Compute and return the distance matrix using the distance measure in `method`.

Usage

```
distMat(X, Ahalf = diag(NCOL(X)), method = "euclidean", d = NULL, p = 2)
```

Arguments

`X`	Design matrix for control variables with dimension `[n p]` (if $p = 1$ and `X` is a vector, then it is converted to a matrix)
`Ahalf`	$A^{1/2}$ weight matrix with dimension `[p p]` so that the distances are computed between $Ax_{0,j}$ and $Ax_{1,i}$.
`method`	the distance measure to be used. This must be one of `"euclidean"`, `"maximum"`, `"manhattan"`, `"canberra"`, `"binary"` or `"minkowski"`. Any unambiguous substring can be given.
`d`	Vector of treatment indicators with length `n`. If supplied, return the `[n1 n0]` distance matrix between treated and untreated observations, otherwise return the full `[n n]` distance matrix
`p`	The power of the Minkowski distance.

Value

`[n1 n0]` or `[n n]` matrix of distances.

`nnvar` *Nearest-neighbor variance estimator*

Description

Calculates an `n`-vector sigma2 of estimates of the variance of `y` using a nearest-neighbor estimator among observations with the same treatment status `d`.

Usage

```
nnvar(DM, d, y, J = 3, tol = 0)
```

Arguments

`DM`	distance matrix with dimension `[n n]`.
`d`	Vector of treatment indicators with length `n`
`y`	outcome vector with length `n`
`J`	number of nearest neighbors to average over
`tol`	numerical tolerance for determining nearest neighbors in constructing matches

`NSW` *Dataset from Dehejia and Wahba (1999)*

Description

Subset of National Supported Work and PSID data from Dehejia and Wahba (1999).

Usage

```
NSW
```

Format

A data frame with 2,675 observations (2,490 controls from PSID, and 185 treated individuals from NSW) and 11 variables.

treated Treatment indicator

age Age in years

education Year of education

black Indicator for black

hispanic Indicator for Hispanic

married Indicator for married

re74 Earnings in 1974 (in thousands of dollars)

re75 Earnings in 1975 (in thousands of dollars)

re78 Earnings in 1978 (in thousands of dollars)

ue74 Indicator for zero earnings in 1974

ue75 Indicator for zero earnings in 1975

Source

Rajeev Dehejia's website, http://users.nber.org/~rdehejia/nswdata2.html

References

Dehejia, R., and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," Journal of the American Statistical Association, 94, 1053-1062.

`NSWexper` *Experimental dataset from Dehejia and Wahba (1999)*

Description

National Supported Work data from Dehejia and Wahba (1999).

Usage

```
NSWexper
```

Format

A data frame with 445 observations (185 treated and 260 controls) and 11 variables.

treated Treatment indicator

age Age in years

education Year of education

black Indicator for black

hispanic Indicator for Hispanic

married Indicator for married

re74 Earnings in 1974 (in thousands of dollars)

re75 Earnings in 1975 (in thousands of dollars)

re78 Earnings in 1978 (in thousands of dollars)

ue74 Indicator for zero earnings in 1974

ue75 Indicator for zero earnings in 1975

Source

Rajeev Dehejia's website, http://users.nber.org/~rdehejia/nswdata2.html

References

Dehejia, R., and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," Journal of the American Statistical Association, 94, 1053-1062.

Index